Michael Hiebel

Fundamentals of Vector Network Analysis

© Rohde&Schwarz GmbH & Co. KG, 2005
Mühldorfstrasse 15
81671 München
Germany
www.rohde-schwarz.com

Sixth edition 2014
Printed in Germany

ISBN 978-3-939837-06-0
PW 0002.6729.00

Preface

Who should read this book?

This book is ideal for beginners in the field of network analysis and supplements the highly specific information provided in operating manuals, application notes, or academic papers. Yet, it can also serve as a reference book for experts. With this purpose in mind, important terminology has been highlighted in bold print and information has been listed in tabular form where possible.

How can I apply this book?

Depending on your background knowledge and fields of work, you can apply the information in this book in several ways. A few examples are provided below. Beginners are particularly encouraged to read chapter 1 as an introduction.

Application 1	If you want information on how to perform a specific measurement task, read chapters 4, 6, 7, 8, and 9. Before you read chapter 6, study chapter 5 to get a solid understanding of the fundamental relations.
Application 2	If you are interested in how the instruments are designed and function, read chapter 2. You may also want to select the right calibration method for your task or optimize the accuracy of your test setup. Therefore, read chapter 3.
Application 3	The available technical dictionaries only cover part of the special vocabulary encountered in network analysis. You can look up the context or definition of terms at any time by using the index's page references.

To help explain concepts, the book includes approximately 400 largely color figures. Moreover, it has been structured to require as little special knowledge from other fields as possible. To make formulas easier to interpret, the appendix provides some useful mathematical principles.

Table of contents

1	**Introduction**	12
1.1	**What is a network analyzer?**	12
1.2	**Wave quantities and S-parameters**	13
1.3	**Why vector network analysis?**	17
1.4	**A circuit example**	18
2	**Design of a heterodyne N-port network analyzer**	22
2.1	**Block diagram**	22
2.2	**Design of the test set**	23
2.2.1	Constancy of the a wave	24
2.2.2	Reflection tracking	25
2.2.3	Directivity	26
2.2.4	Test port match and multiple reflections	29
2.2.5	Summary	30
2.2.6	Outlook	32
2.3	**Implementation of the directional element**	33
2.3.1	VSWR bridge	34
2.3.2	Directional coupler	36
2.3.3	Other implementations	42
2.4	**Other components of the test set**	42
2.4.1	Receiver step attenuators	42
2.4.2	Generator attenuators	46
2.4.3	Active and passive test sets	48
2.5	**Generator**	49
2.6	**Reference and measurement receiver**	50
2.7	**Measurement procedure**	54
2.7.1	S-parameter measurement procedure	54

2.7.2	Measurement data processing chain	55
2.7.3	Trace generation	57
2.8	**Main setting parameters**	57
2.8.1	User interface	57
2.8.2	Channel settings	60
2.8.3	Trace settings	63
2.9	**Remote control of the instrument**	70
2.9.1	Usage of simple digital signals	70
2.9.2	Protocol-based interfaces	70
2.9.3	Automation	73
2.10	**Simplified implementations**	76
2.10.1	N+1 receiver analyzer	77
2.10.2	Network analyzer with an N-port switching matrix	78
3	**Measurement accuracy and calibration**	80
3.1	**Reduction of random measurement errors**	81
3.1.1	Thermal drift	81
3.1.2	Repeatability	81
3.1.3	Noise	84
3.2	**Correction of systematic measurement errors**	86
3.2.1	Nonlinear influences	86
3.2.2	Linear influences	87
3.3	**Calibration standards**	89
3.3.1	Coaxial calibration standards	91
3.3.2	Waveguide calibration standards	99
3.3.3	Microstrip calibration standards	102
3.3.4	Coplanar calibration standards	105
3.3.5	The uniform model of the calibration standards	109
3.4	**Linear error models and calibration techniques**	110
3.4.1	3-term error model (OSM technique)	111
3.4.2	7-term error model (TOM, TRM, TRL, TNA, UOSM techniques)	113
3.4.3	10-term and 12-term error models (TOSM technique)	119

3.4.4	15-term error model (TOM-X technique)	123
3.4.5	Adapters and noninsertable DUTs	124
3.4.6	Incomplete calibration techniques	127
3.4.7	Practical hints for calibration	129
3.5	**Verification**	132
3.5.1	T-check and Beatty standard	133
3.5.2	Measurement of the effective system data	136
3.5.3	A primer to statistics	143
3.5.4	Analysis of the measurement uncertainty	146
3.6	**Traceability**	151
3.6.1	The International System of Units	151
3.6.2	The pseudo units dB and dBm	153
3.6.3	Some important non-SI units	154
3.6.4	An organization for traceability	155
3.6.5	Traceability of a network analyzer	156
4	**Linear measurements**	157
4.1	**Performing a TOM calibration**	157
4.2	**Performing a TNA calibration**	160
4.3	**Measurement of the reflection coefficient and the SWR**	162
4.4	**Measurement of the transmission coefficient**	167
4.5	**Measurement of the group delay**	169
4.6	**Measurement of the phase delay, auto length**	173
4.7	**Measurement of the stability**	175
4.8	**Measurement with embedding**	178
4.9	**Measurement with deembedding**	183
4.10	**Measurement of balanced lines**	188

4.11 Measurement of the far-end and near-end crosstalk 192

4.12 Filter with balanced and unbalanced port, imbalance and common-mode rejection 196

4.13 Measurement of switching times and drift effects 201

4.14 Measurements on amplifiers in pulsed operating mode 208

4.15 Measurement of the efficiency 210

5 Time-domain measurements 214

5.1 Time-domain analysis 214
5.1.1 Impulse and step response 215
5.1.2 Time-domain analysis of linear RF networks 217
5.1.3 Time domain reflectometry using an oscilloscope 217
5.1.4 Fourier transform 224

5.2 Numerical inverse Fourier transform 226
5.2.1 Inverse discrete Fourier transform 227
5.2.2 Windowing 234
5.2.3 Bandpass mode 240
5.2.4 Transformations optimized for computing time 244

5.3 Using the time-domain option 244
5.3.1 Operation in lowpass mode 245
5.3.2 Operation in bandpass mode 249
5.3.3 Benefits of extrapolation 250
5.3.4 Processing sequence 252

5.4 Time gate 253

5.5 Tables and diagrams 257
5.5.1 Impulse and step responses for important reflection coefficients 257
5.5.2 Comparison of important window functions 264
5.5.3 Comparison of important time gates 265
5.5.4 Diagram for determination of the ambiguity range 267

6	**Examples of time-domain measurements**	268
6.1	**Distance-to-fault measurement and gating**	268
6.2	**Measurements on a SAW filter in the time domain**	274
6.3	**RF imaging for nondestructive evaluation**	280
6.4	**Measurement of the complex effective system data and OSML calibration**	283
7	**Nonlinear measurements**	288
7.1	**Features used for nonlinear measurements**	288
7.1.1	Automatic level control	288
7.1.2	Source power calibration	291
7.1.3	Receiver power calibration	293
7.1.4	Power sweep	294
7.1.5	Multiple source concept	295
7.1.6	Arbitrary mode	296
7.1.7	Direct generator and receiver access	298
7.1.8	Power sensors as receivers	298
7.1.9	External generator control	299
7.1.10	Additional equipment	300
7.2	**Measurement of the compression point**	301
7.3	**Measuring a detector characteristic**	302
7.4	**Harmonics**	304
7.4.1	Model of harmonic distortions	304
7.4.2	Measurement of the harmonics and their intercept point	308
7.5	**Intermodulation**	315
7.5.1	Model of intermodulation distortions	315
7.5.2	Measurement of the intermodulation products and their intercept point	318
7.6	**Boosted source with an external test set**	328

7.7	**Measuring hot S-parameters**	330
7.8	**Load-pull measurements**	332
7.9	**True-differential-measurements**	336
8	**Mixer measurements**	338
8.1	**Signals and parameters for a mixer**	338
8.1.1	Input and output signals of a mixer	338
8.1.2	Higher order mixing products	341
8.1.3	Important mixer parameters	342
8.2	**Features for mixer measurements**	346
8.2.1	Mixer measurement mode	346
8.2.2	Connection of a reference mixer	348
8.3	**Example 1: Mixer measurement**	348
8.4	**Example 2: Measurements on a down-converting module**	352
8.5	**Frequency extension**	354
9	**Antenna and radar cross section measurements**	360
9.1	**Antenna measurements**	361
9.2	**Important antenna measurement quantities**	366
9.3	**Radar cross section measurements**	369
10	**Conclusion and acknowledgement**	373

A Appendix 374

A1 Mathematical principles 374
A1.1 Complex numbers 374
A1.2 Matrix operations 377

A2 Important symbols and quantities 380

A3 Circuit symbols used 385

A4 Acronyms and abbreviations 386

A5 Sources of figures 390

A6 Bibliography 392
A6.1 Selected application notes 392
A6.2 Books, scientific publications and standards 393

A7 Index 398

A8 Current network analyzer product lines from Rohde&Schwarz 415

1 Introduction

1.1 What is a network analyzer?

One of the most common measuring task 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.

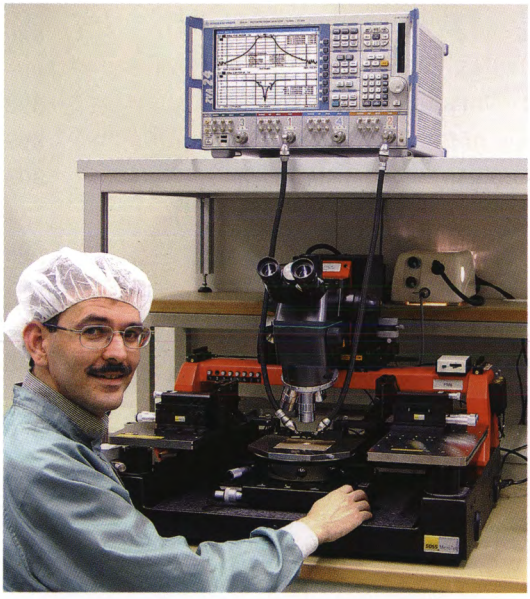


Fig. 1.1.1
The author at work at a test station for RF integrated circuits (network analyzer: R&S®ZVA24, wafer prober: SUSS MicroTec PM8).

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 & 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.

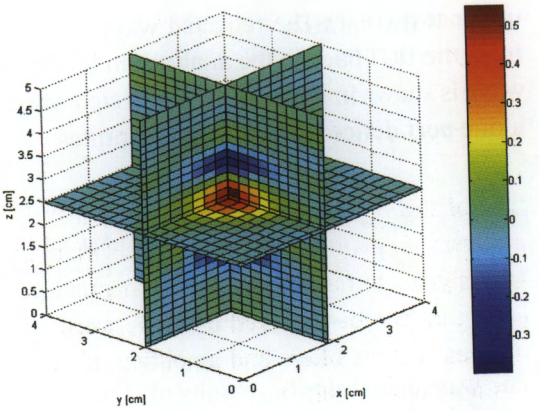


Fig. 1.1.2
Material defect.

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.

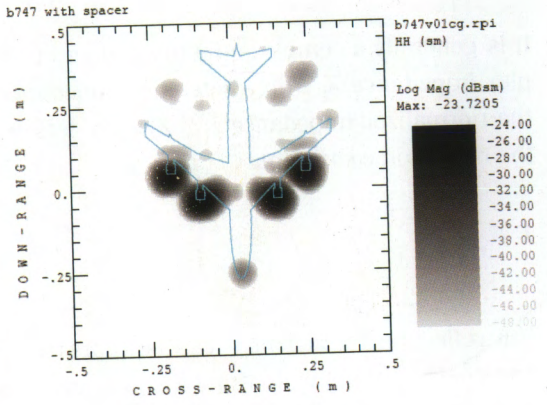


Fig. 1.1.3
ISAR image of the model of a Boeing 747.

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.

1.2 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.

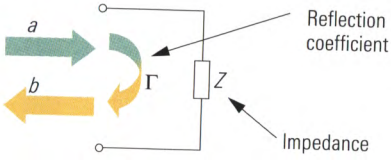


Fig. 1.2.1
One-port device with incident and reflected waves.

The wave quantities have units of $\sqrt{\mathrm{W}}$. 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.

$$\Gamma = \frac{b}{a} \tag{1.2-1}$$

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

$$\Gamma = \frac{z-1}{z+1} \tag{1.2-2}$$

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.

1) In RF engineering and RF measurement a reference impedance of 50 Ω is used. In broadcasting systems a reference impedance of 75 Ω 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.

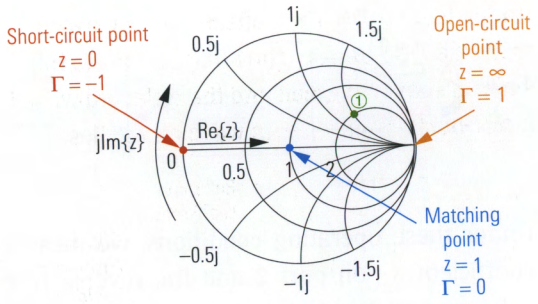


Fig. 1.2.2
Smith chart with sample points.

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.

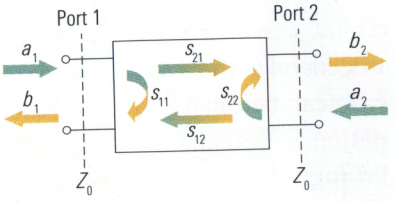


Fig. 1.2.3
Two-port device with its wave quantities.

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 reflection-free termination $\Gamma = 0$ (match) is used on port 2. This means that $a_2 = 0$. Port 1 is stimulated by the incident wave $a_1 \neq 0$.

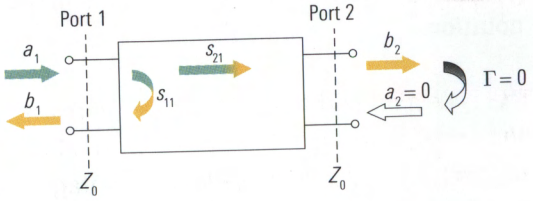


Fig. 1.2.4
Two-port device during forward measurement.

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.

$$s_{11} = \left.\frac{b_1}{a_1}\right|_{a_2 = 0} ; \; s_{21} = \left.\frac{b_2}{a_1}\right|_{a_2 = 0} \qquad \textbf{(1.2-3)}$$

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

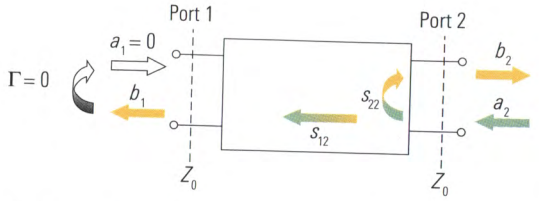


Fig. 1.2.5
Two-port device during reverse measurement.

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.

$$s_{12} = \left.\frac{b_1}{a_2}\right|_{a_1 = 0}; \quad s_{22} = \left.\frac{b_2}{a_2}\right|_{a_1 = 0} \tag{1.2-4}$$

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 $a_2 = 0$. This results in the following:

$$\begin{aligned} b_1 &= s_{11}a_1 + s_{12}a_2 \\ b_2 &= s_{21}a_1 + s_{22}a_2 \end{aligned} \tag{1.2-5}$$

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:

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

$$\mathbf{b} = \mathbf{S}\mathbf{a} \tag{1.2-7}$$

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:

$$\begin{aligned} b_1 &= s_{11}a_1 + s_{12}a_2 + s_{13}a_3 \\ b_2 &= s_{21}a_1 + s_{22}a_2 + s_{23}a_3 \\ b_3 &= s_{31}a_1 + s_{32}a_2 + s_{33}a_3 \end{aligned} \tag{1.2-8}$$

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 three × three 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 an similar manner for $N > 3$.

1.3 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.

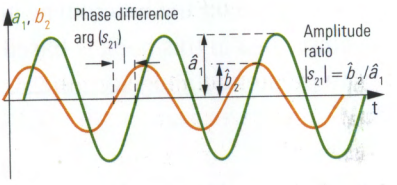


Fig. 1.3.1 Signals a_1 and b_2.

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. $\hat{b}_2$ 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 book 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.
- 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 book is based on this principle which is discussed in greater detail in section 2.6.

1.4 A circuit example

Fig. 1.4.1 below 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 ❶ at the start of the signal processing chain is used. The filtered signal is fed via a semi-rigid cable ❷ to a mixer ❸ which converts the signal from frequency f_{RF} to frequency f_{IF}. A switchable attenuator ❹ is used to set the IF level. The LO signal required by the mixer is prepared using several amplifiers ❺, a frequency doubler ❻ and a bandpass filter ❼. The module itself is controlled via a 48-pin interface and is commonly used in test receivers and spectrum analyzers.

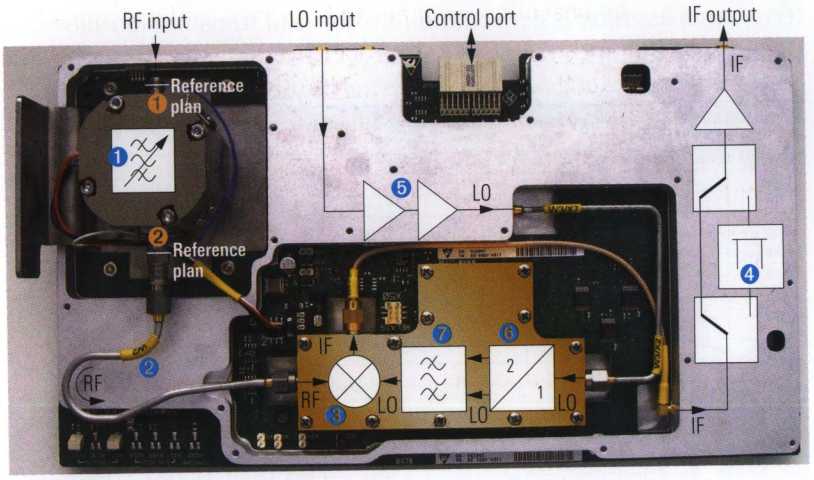


Fig. 1.4.1 *The RF EXTENSION module from Rohde & Schwarz.*

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

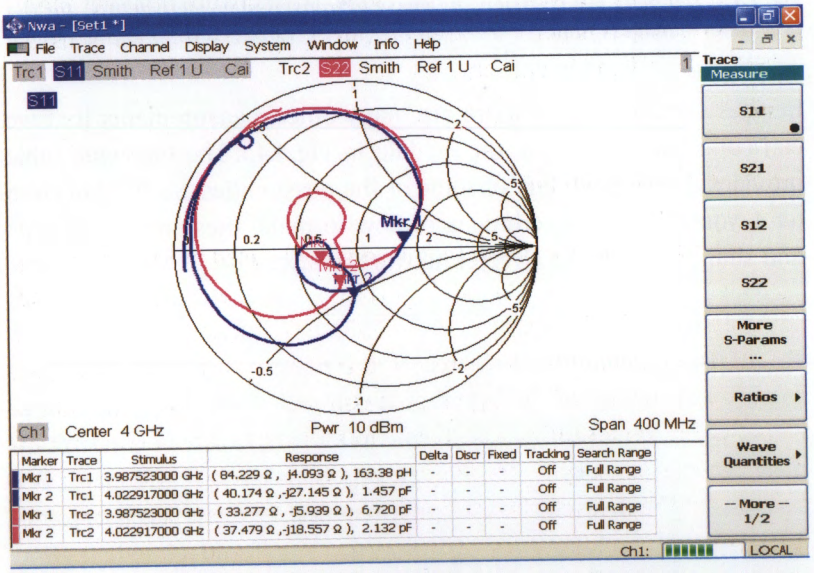


Fig. 1.4.2 *Input and output reflection coefficients of the tunable bandpass filter.*

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 the following figure.

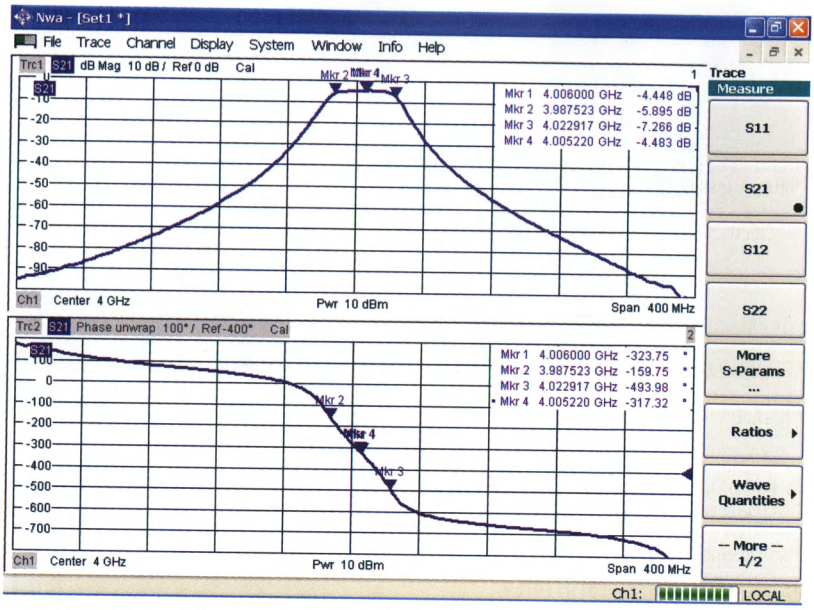


Fig. 1.4.3 *Forward transmission and phase of transmission coefficient s_{21} for the bandpass filter.*

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. Starting in chapter 4, you can read descriptions of how to make these measurements and similar ones using a vector network analyzer.

	Filters	Lines	Mixers	Switches, switchable attenuators	Amplifiers	Frequency multipliers	Balanced filters
Examples from Fig. 1.4.1 →	❶	❷	❸	❹	❺	❻	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.*

2 Design of a heterodyne N-port network analyzer

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

2.1 Block diagram

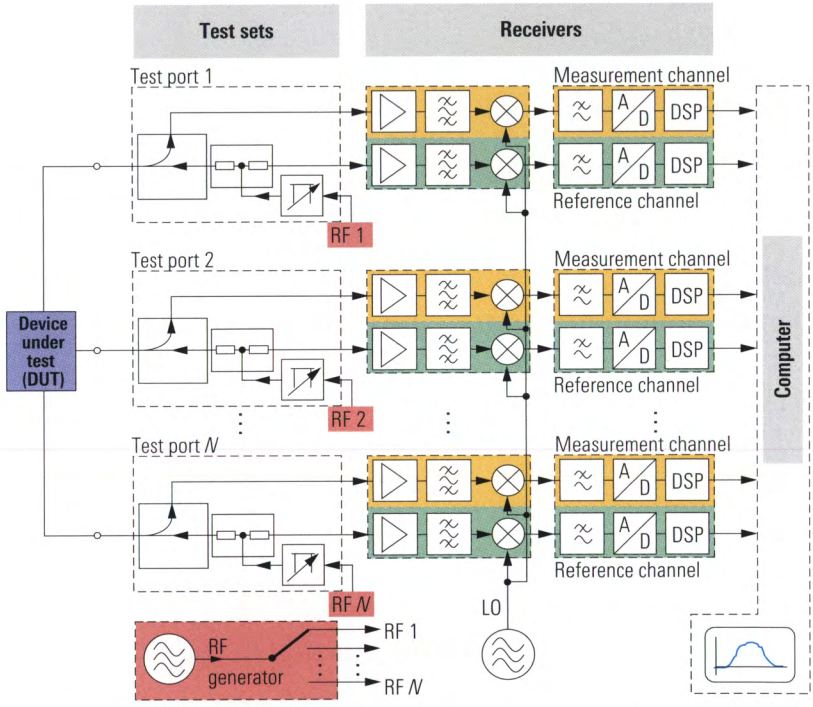


Fig. 2.1.1 Standard block diagram of a N-port vector network analyzer.

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 extends 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 seperate **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.
- 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.

2.2 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. The implementation of the **directional element** is considered in greater detail in section 2.3. In the following discussion, it is described as a three-port device.

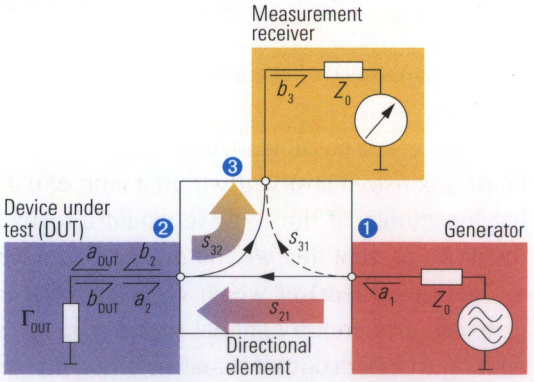


Fig. 2.2.1
Measurement circuit with directional element.

In the figure above, the two main signal directions of the directional element are shown in color. The wave a_1 produced by the generator is for-

1) In place of "measurement channel" the term "test channel" is also common. Further more the "measurement receiver" is also called "test receiver".

warded to port ❷ 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}.

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

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:

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

Finally wave a_2 reaches port ❸ 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 ❶ to port ❸ is undesired. Accordingly, we would like to obtain the best possible isolation in which $s_{31} = 0$. Reflection at port ❷ 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.

2.2.1 Constancy of the a wave

In practice, it is possible to maintain the generator wave a_1 at an approximately constant level, e.g. within a limit of 0 dBm ± 0.3 dB. The remaining inaccuracy of the a_1 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 possible mismatch on the output ports of the power splitter.[1] If

1) This characteristic of the power splitter is essential for its function. Unlike the power splitter a power divider formed of three $Z_0/3$ resistors is unsuitable for this use.

the DUT is connected to one branch, via a directional element, the wave quantity a_1' can be used instead of quantity a_1.[1)]

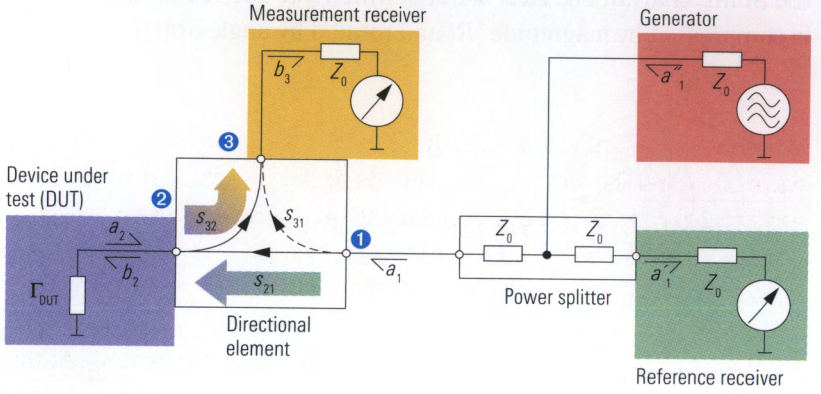


Fig. 2.2.2 Directional element with reference channel.

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

$$M = \frac{b_3}{a_1'} \qquad \textit{(2.2-3)}$$

2.2.2 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.

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

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

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

1) Along with the drawings of this book 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.

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 $|R|$ and rotated by angle $\arg(R)$.

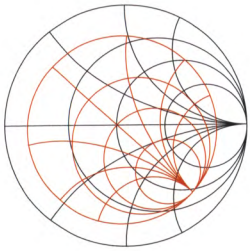


Fig. 2.2.3
Ideal and distorted Smith chart for formula (2.2-5).

2.2.3 Directivity

Crosstalk from port ❶ to port ❸ 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**:

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

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 $x = D/\Gamma_{DUT}$. Further more the product $R \cdot \Gamma_{DUT}$ is donated as W.

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

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 cir-

cle shown in Fig. 2.2.4. Any more accurate calculation of $(1+x)$ is not possible.

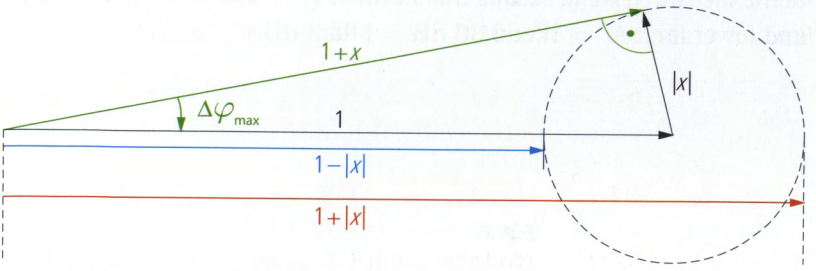


Fig. 2.2.4 Vectorial superposition of 1 and $|x|$.

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:[1)]

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

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

$$\Delta\varphi_{\max} = \arcsin(x) \qquad \textbf{(2.2-11)}$$

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 $W = \Gamma_{\mathrm{DUT}}$. To use table 2.2.1 we calculate the normalized value $|x| = |D|/|\Gamma_{\mathrm{DUT}}|$ meaning -40 dB $-$ $(-30$ dB$) = -10$ dB in dB scale. According to the

1) lg is used to designate the decimal logarithm (logarithm to the base 10, also called-Briggsian logarithm). Accordingly to the Standard ISO31-11: 1992 Item No. 11-8.6 log x should not be used in place of lg x. However log x is often used erroneously on calculators and in other books instead of lg x.

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.

$\lvert x\rvert$	$1+\lvert x\rvert$	$1-\lvert x\rvert$	$\Delta\varphi_{max}$
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.*

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 $b_3/a_1 = 2R \cdot D$ corresponding to values of $-\infty$ 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, 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 $R \cdot D$.

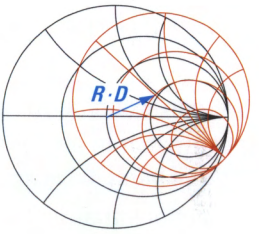


Fig. 2.2.5
Ideal and distorted Smith chart for formula (2.2-7).

2.2.4 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 $S \neq 0$. 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}.

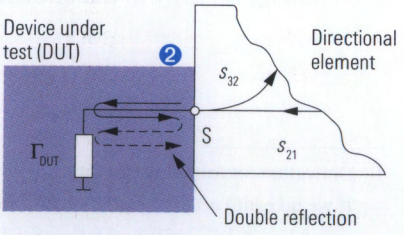


Fig. 2.2.6
Multiple reflection at the test port.

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, **mul-**

tiple 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 $b_2\Gamma_{DUT}$ 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:

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

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

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

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 $(1 + \Gamma_{DUT}S)$. If we want to take into consideration reflections that go beyond the double reflection, we can use the following formula.[1] It holds assuming $|S \cdot \Gamma_{DUT}| < 1$, which is generally the case since $S << 1$.

$$a_2 = \frac{\Gamma_{DUT}}{1 - S \cdot \Gamma_{DUT}} b_2 \qquad \textit{(2.2-14)}$$

2.2.5 Summary

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

1) Derivation for formula (2.2-14):
If we take into account higher-order reflections, instead of formula (2.2-12) we obtain
$a_2 = b_2\Gamma_{DUT} + b_2\Gamma_{DUT}S\Gamma_{DUT} + b_2\Gamma_{DUT}S^2\Gamma^2_{DUT} + b_2\Gamma_{DUT}S^3\Gamma^3_{DUT} + \ldots$
After we factor out $b_2\Gamma_{DUT}$, we obtain
$a_2 = b_2\Gamma_{DUT}(1 + S\Gamma_{DUT} + S^2\Gamma^2_{DUT} + S^3\Gamma^3_{DUT} + \ldots)$
Finally, the expression in parentheses can be approximated using the geometric series $1 + x + x^2 + x^3 \ldots = 1/(1-x)$ for $|x| < 1$ so that we obtain formula (2.2-14).

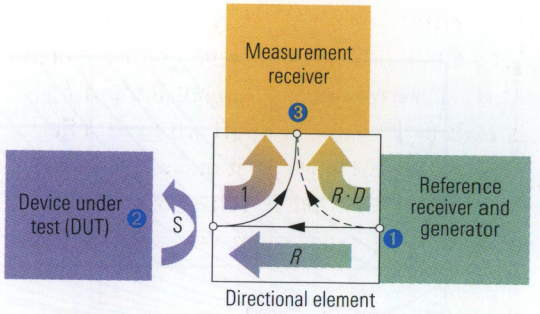


Fig. 2.2.7
Illustration of the quantities R, D, S

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

$$M = R\left(D + \frac{\Gamma_{DUT}}{1 - S \cdot \Gamma_{DUT}}\right) \qquad \textit{(2.2-15)}$$

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:

$$M = R\left(D + \Gamma_{DUT}\left(1 + S \cdot \Gamma_{DUT}\right)\right) \qquad \textit{(2.2-16)}$$

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 $|\Gamma_{DUT}|$. The following relationships hold:

- The directivity limits the measurement accuracy for small values of the reflection coefficients magnitude $|\Gamma_{DUT}|$.
- The test port match determines the limit for large magnitudes $|\Gamma_{DUT}|$.

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

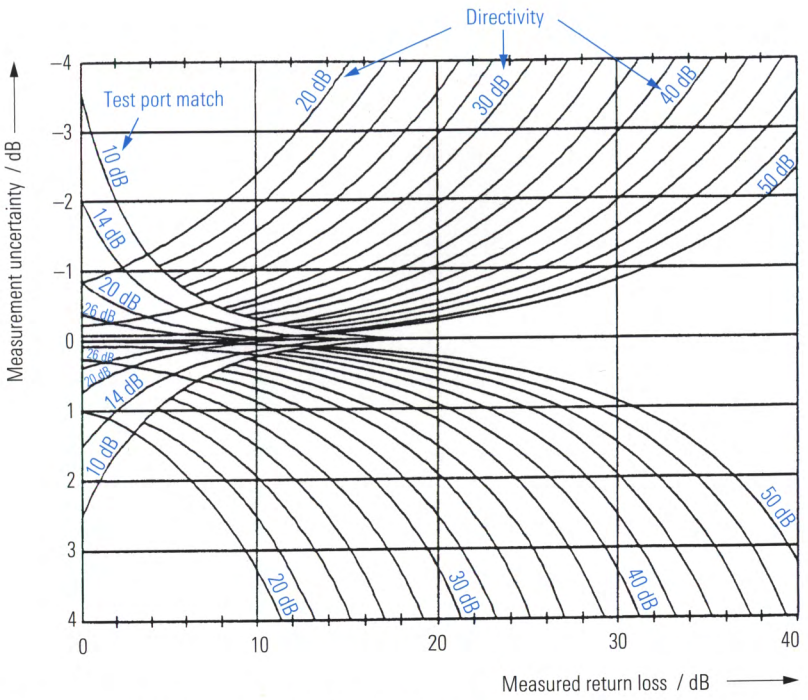


Fig. 2.2.8 *Measurement uncertainty as a function of the directivity, test port match and measured return loss.*

2.2.6 Outlook

The directional element in a network analyzer's test set can be described as a three-port device. Accordingly, it has nine S-parameters. In addition, the generator, measurement receiver and reference receiver have a reflection coefficient that is not equal to 0. The power splitter is only approximately ideal ($a_1 \approx a_1'$), and both receivers weighting the presented wave quantities b_3 or a_1' with factors unequal to one. There is also an influence of the test cables and the test fixture that are used to connect the DUT. However, it is possible to appropriately summarize the additional quantities and assign them depending on their influence to the reflection tracking R, the directivity D and the test port match S. To distinguish between the effects of the directional element and the over all error quantities, the **error terms** e_{00}, e_{01}, e_{10}, e_{11} are introduced. They form the **error two-port** shown in Fig. 2.2.9. The reference receiver, generator and measurement receiver form an ideal network analyzer. It has a wave source consisting of a generator and reference receiver to gen-

erate and receive the wave quantity a_1. The wave returning from the error two-port is measured by the measurement receiver. Wave quantities b_1 and a_3 will not be considered further since by definition they have a value of 0. The error two-port is connected at its port 2 directly to the DUT (no transmission line between them!).

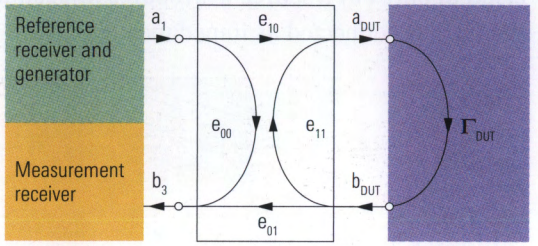


Fig. 2.2.9 Error two-port with e terms.

In the advanced sections like section 3.4 we will discuss transmission measurements and their related error model. This kind of measurement will involve two or more test ports. To be able to expand the error model easily, we distinguish the two quantities e_{10} and e_{01} instead of the quantity R. In technical literature by convention the quantity e_{00} is interpreted as the directivity of the network analyzer although this quantity corresponds functionally to the product $R \cdot D$. The quantity e_{11} represents the test port match of the overall system.

2.3 Implementation of the directional element

There are various possibilities of designing the directional element. VSWR bridges and directional couplers are often used because they have different frequency ranges that mutually complement one another. A VSWR bridge is suitable for a range from a few kHz up to about 4 GHz. Typical directional couplers require a minimum frequency of about 1 GHz for good coupling. If necessary, they can also be designed for frequencies above 100 GHz.

Section 2.3 is intended particularly for readers who would like to implement the test set themselves externally. This can be necessary, for example, in conjunction with a boosted source or other high power configurations (section 7.6). When you are first reading this book, you can initially skip the remaining parts of section 2.3 without any lack of continuity.

2.3.1 VSWR bridge

The **VSWR bridge** consists of resistors R_1, R_2 and impedance Z_0. When it is redrawn, it resembles a Wheatstone bridge circuit (cf. Fig. 2.3.1). However, no zero alignment of the diagonal voltage is performed. The bridge is fed at port 1 by a source with source impedance Z_0. At port 2 (test port), the DUT is connected as impedance Z_{DUT}.

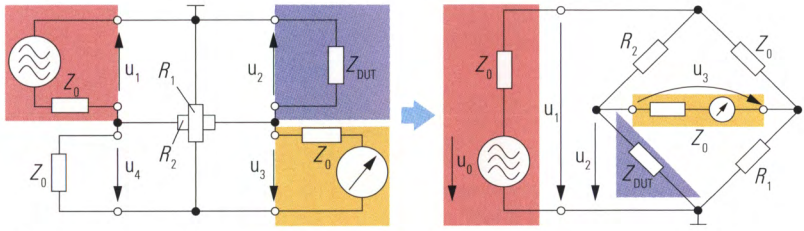


Fig. 2.3.1 Redrawing the VSWR bridge.

For a basic understanding of the related wave quantities, Fig. 2.3.1 can be compared to Fig. 2.2.1 using its background colors. For the test port match, we calculate as follows:

$$s_{22} = \frac{R_1 R_2 - Z_0^2}{2(R_1 + Z_0)(R_2 + Z_0)} \qquad \textbf{(2.3-1)}$$

To obtain an ideal test port match of $s_{22} = 0$, we need:

$$R_1 R_2 = Z_0^2 \qquad \textbf{(2.3-2)}$$

Given this condition, we obtain the S-matrix elements for a VSWR bridge as follows:

$$s_{21} = \frac{Z_0}{Z_0 + R_2} \quad \text{and} \quad s_{32} = \frac{R_2}{Z_0 + R_2} \quad \text{as well as} \quad s_{31} = 0 \qquad \textbf{(2.3-3/5)}$$

Assuming the usual choice with $R_1 = R_2 = Z_0$, we obtain from formulas (2.3-3) and (2.3-4) a theoretical reflection tracking $R = ½ \cdot ½ = ¼$

which corresponds to 12 dB.[1] This value usually coincides relatively well with the technical specifications of the device. Due to the unavoidable mechanical and electrical tolerances, a value of $s_{31} = 0$ cannot be exactly achieved. Thus, only a directivity $|D| > 0$ can be obtained.

One disadvantage of the circuit in Fig. 2.3.1 is that the diagonal voltage u_3 must be obtained in a balanced manner since the bridge will be partially short-circuited otherwise. Using a transformer, the diagonal voltage can be converted to a potential with a ground reference (cf. Fig. 2.3.2). However, the transformer cannot have an arbitrary broadband implementation and it is not suitable for frequencies above approx. 4 GHz. On the other hand, at low frequencies node (❷) of the bridge is subject to greater loading than node (❶). This asymmetry would impair the directivity of the bridge. By installing an auxiliary network (shown with a blue background in Fig. 2.3.2), it is possible to compensate the asymmetry of the bridge. Note that this network also changes the input impedance on the test port (port 2) since the auxiliary network is in parallel with the impedance Z_{DUT}.

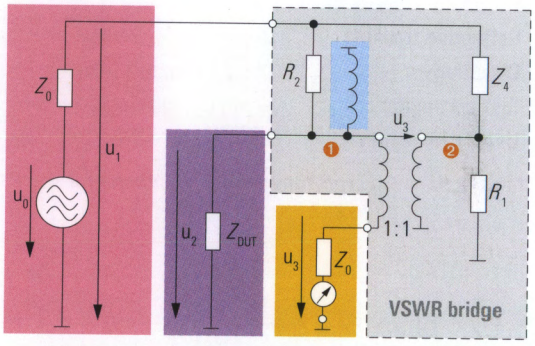


Fig. 2.3.2
VSWR bridge with a tap via a transformer.

To ensure good test port match nevertheless, the top two impedances Z_4 and R_2 of the bridge are configured to have a somewhat higher impedance. The resistor R_1 which is opposite the test port also influences the test port match. Accordingly, it must exhibit the real system impedance

1) When calculating the equivalent dB value from a reflection tracking $\mathrm{R} = ¼$ one would normally yield a value of $20\lg(¼) \approx -12$ dB. However in data sheets it is common to refer to the reflection tracking error when using a dB scale. In our example this yields to $20\lg(4) \approx 12$ dB. It is the same with the quantities directivity, test port match, transmission tracking and all the error terms. You could take table 2.3.1 (next page) as an example. For the sake of continuous notation we use the previous described dB scaling for excerpted data sheet values as well as theoretically derived dB values.

$Z_0 = 50\ \Omega$ over a wide frequency range. The following figure shows a VSWR bridge that has been opened. The transformer and the auxiliary network are clearly recognizable. To obtain the widest possible bandwidth, they have several graduated ferrite cores.

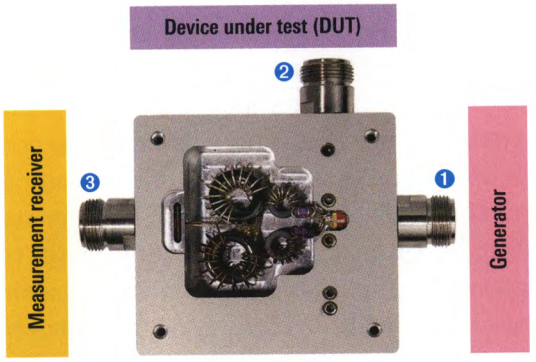


Fig. 2.3.3
The bridge R&S®ZRC with the cover removed.

The specifications for the bridge R&S®ZRC provide insight into the characteristics of a high-quality VSWR bridge.

Reflection tracking	$\lvert R\rvert = 0.224$ (@ 5 MHz)	Corresponds to 13 dB
Directivity	$\lvert D\rvert \leq 0.01$ (up to 3 GHz)	Corresponds to $\geq$40 dB
Test port match	$\lvert S\rvert \leq 0.07$ (400 kHz to 2.5 GHz)	Corresponds to $\geq$23 dB
Usable frequency	$f =$ 40 kHz to 4 GHz	

Table 2.3.1 *Excerpt from the specifications for the R&S®ZRC from [RS04] page 222.*

2.3.2 Directional coupler

A **directional coupler** basically consists of two closely spaced and coupled lines with a length of $\lambda/4$. We distinguish between forward and reverse couplers. In Figs 2.3.4 a / b, by way of example, a signal is fed in at port ❶ (by the generator) or port ❷ (by the wave reflected from the DUT). The stimulated transmission paths at port ❶ or port ❷ are marked with green and orange arrows, respectively. The name forward or reverse coupler is derived from the direction of the coupling path (curved arrow). In network analysis, forward couplers are used only in a few exceptional cases (e.g. when using waveguide couplers) and will not be considered further here.

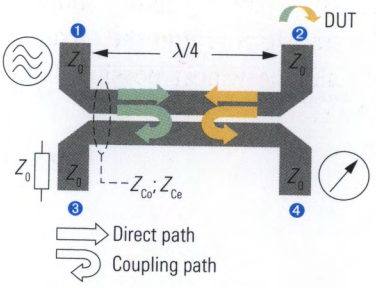


Fig. 2.3.4a Reverse coupler

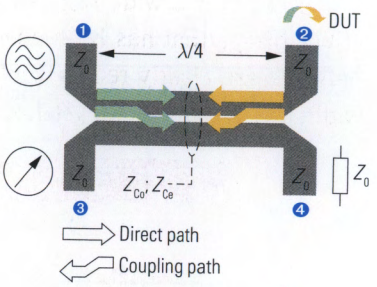


Fig. 2.3.4b Forward coupler

We will consider a reverse coupler under the following assumptions:

- The configuration is exactly symmetrical
- The line cross-section is homogeneous (i.e. TEM line)
- Losses are neglected

On the coupled lines, two different wave modes can propagate. The odd- and even-mode characteristic impedances Z_{Co}, Z_{Ce} are assigned to these different electromagnetic field configurations. The wavelength λ is the same for both wave modes. For use as a directional element, port ❸ is not necessary and is terminated in a reflection-free manner ($a_3 = 0$). The resulting three-port network is known as a **three-port directional coupler**. For the reflection measurement, we feed in a signal at port ❶. The DUT is located at port ❷. The reflection coefficient of the DUT is determined using the wave quantity b_4 (cf. Fig. 2.3.5).

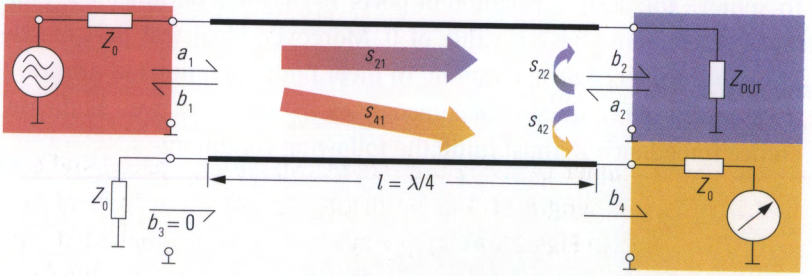


Fig. 2.3.5 A reverse coupler as a directional element.

For reasons of symmetry, it is sufficient to consider only the following S-parameters: s_{22} (plays a part in determining the test port match S), s_{21}, s_{42} (the product of the two yields the reflection tracking R) and s_{41} (besides R, it determines the directivity D). The S-parameters for a reverse coupler can be found in relevant literature, e.g. [MG92] page C28, [ZB93] page 199, [Sc84] page 18 and other publications.

$$s_{22} = \frac{1}{2}\left[\frac{j\left(\frac{Z_{Ce}}{Z_0} - \frac{Z_0}{Z_{Ce}}\right)\sin(2\pi l/\lambda)}{2\cos(2\pi l/\lambda) + j\left(\frac{Z_{Ce}}{Z_0} + \frac{Z_0}{Z_{Ce}}\right)\sin(2\pi l/\lambda)} + \frac{j\left(\frac{Z_{Co}}{Z_0} - \frac{Z_0}{Z_{Co}}\right)\sin(2\pi l/\lambda)}{2\cos(2\pi l/\lambda) + j\left(\frac{Z_{Co}}{Z_0} + \frac{Z_0}{Z_{Co}}\right)\sin(2\pi l/\lambda)}\right] \qquad \textit{(2.3-6)}$$

$$s_{21} = \left[\frac{1}{2\cos(2\pi l/\lambda) + j\left(\frac{Z_{Ce}}{Z_0} + \frac{Z_0}{Z_{Ce}}\right)\sin(2\pi l/\lambda)} + \frac{1}{2\cos(2\pi l/\lambda) + j\left(\frac{Z_{Co}}{Z_0} + \frac{Z_0}{Z_{Co}}\right)\sin(2\pi l/\lambda)}\right] \qquad \textit{(2.3-7)}$$

$$s_{42} = \frac{1}{2}\left[\frac{j\left(\frac{Z_{Ce}}{Z_0} - \frac{Z_0}{Z_{Ce}}\right)\sin(2\pi l/\lambda)}{2\cos(2\pi l/\lambda) + j\left(\frac{Z_{Ce}}{Z_0} + \frac{Z_0}{Z_{Ce}}\right)\sin(2\pi l/\lambda)} - \frac{j\left(\frac{Z_{Co}}{Z_0} - \frac{Z_0}{Z_{Co}}\right)\sin(2\pi l/\lambda)}{2\cos(2\pi l/\lambda) + j\left(\frac{Z_{Co}}{Z_0} + \frac{Z_0}{Z_{Co}}\right)\sin(2\pi l/\lambda)}\right] \qquad \textit{(2.3-8)}$$

$$s_{41} = \left[\frac{1}{2\cos(2\pi l/\lambda) + j\left(\frac{Z_{Ce}}{Z_0} + \frac{Z_0}{Z_{Ce}}\right)\sin(2\pi l/\lambda)} - \frac{1}{2\cos(2\pi l/\lambda) + j\left(\frac{Z_{Co}}{Z_0} + \frac{Z_0}{Z_{Co}}\right)\sin(2\pi l/\lambda)}\right] \qquad \textit{(2.3-9)}$$

To achieve the desired isolation of ports 1→4, the S-parameter s_{41}, formula (2.3-9), must have a value of 0. Moreover, ideal test port match $s_{22} = 0$, formula (2.3-6), is required. To meet both conditions at arbitrary frequencies, the characteristic impedances Z_{co} and Z_{ce} as well as the reference impedance Z_0 must fulfill the following condition:

$$Z_{Ce}Z_{Co} = Z_0^2 \qquad \textit{(2.3-10)}$$

Thus, we can simplify formulas (2.3-7) and (2.3-8) to obtain the following equations:

$$s_{21} = \frac{2}{2\cos(2\pi l/\lambda) + j\left(\frac{Z_{Ce}}{Z_0} + \frac{Z_0}{Z_{Ce}}\right)\sin(2\pi l/\lambda)} \qquad \textit{(2.3-11)}$$

$$s_{42} = \frac{j\left(\frac{Z_{Ce}}{Z_0} - \frac{Z_0}{Z_{Ce}}\right)\sin(2\pi l/\lambda)}{2\cos(2\pi l/\lambda) + j\left(\frac{Z_{Ce}}{Z_0} + \frac{Z_0}{Z_{Ce}}\right)\sin(2\pi l/\lambda)} \qquad \textit{(2.3-12)}$$

The wavelength λ is dependent on the frequency. The ideal choice of the coupler length where $l = \lambda/4$ is thus possible only for a specific center frequency f_0. At this frequency, formulas (2.3-11) and (2.3-12) can be simplified as follows:

$$s_{21} = -j\frac{2Z_{Ce}Z_0}{Z_{Ce}^2 + Z_0^2} = -j\sqrt{1 - s_{42}^2} \qquad \textit{(2.3-13)}$$

$$s_{42} = \frac{1 - Z_{Co}/Z_{Ce}}{1 + Z_{Co}/Z_{Ce}} \qquad \textit{(2.3-14)}$$

At the frequency f_0, the coupling coefficient s_{42} is real and is determined only by the ratio of the two characteristic impedances Z_{Ce} and Z_{Co}. Based on formulas (2.3-13) and (2.3-14), it is possible to derive a guideline for designing the odd and even characteristic impedances.

$$Z_{Ce} = Z_0\sqrt{\frac{1 + s_{42}}{1 - s_{42}}} \qquad \textit{(2.3-15)}$$

$$Z_{Co} = Z_0\sqrt{\frac{1 - s_{42}}{1 + s_{42}}} \qquad \textit{(2.3-16)}$$

For $s_{42} = 1/\sqrt{2}$, we read from formulas (2.3-13) and (2.2-4) the reflection tracking as $|R| = (1/\sqrt{2})^2 = ½$ or 6 dB. In a configuration with $s_{42} = 1/\sqrt{2}$, we obtain the lowest attenuation along the signal path $1 \rightarrow 2 \rightarrow 4$.

If we use the reverse coupler in a network analyzer, we will also want to use frequencies that are different than the center frequency f_0. To illustrate the frequency dependence of formulas (2.3-11) and (2.3-12), we evaluated them in the vicinity of f_0. The transmission coefficient and the coupling coefficient are shown in Fig. 2.3.6 using a decibel scale as $a_{21} = -20\lg(s_{21})$ and $a_{42} = -20\lg(s_{42})$ vs. a normalized frequency axis. As two examples the coupling coefficient s_{42} at the center frequency was

selected as $s_{42} = 1/\sqrt{2}$ (also called a 3 dB coupler) and $s_{42} = 0.5$ (also called a 6 dB coupler). Usage of the couplers makes sense in roughly the range $^2/_3 f_0$ to $^4/_3 f_0$, i.e. within an octave.

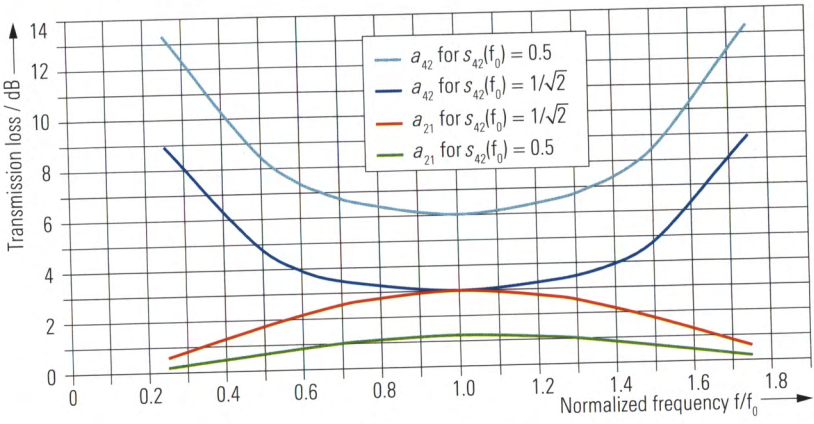


Fig. 2.3.6 *Frequency dependence of the reverse coupler.*

For usage in a broadband RF test instrument, the directional coupler from Fig. 2.3.6 would definitely be too narrowband. The bandwidth can be increased, for example, using multiple graduated line sections (cf. Fig. 2.3.7a). This simple variant alone increases the bandwidth to 2 to 3 octaves. If we use more line sections, we finally obtain the directional coupler shown in Fig. 2.3.7b.



Fig. 2.3.7a *Graduated directional coupler*



Fig. 2.3.7b *Directional coupler with sliding, position-dependent coupling*

However, the techniques used to increase the bandwidth cause increased coupling loss s_{42}. To be used in the high frequency range, couplers must have a line length of less than a few centimeters (e.g. to avoid line losses). This means that a sizeable coupling effect requires frequencies above 1 GHz. Due to the unavoidable tolerances and non-ideal conditions, the crosstalk s_{41} will not disappear entirely in the real world. Thus the directivity D of the directional coupler will not have a value of 0. There is a similar situation with the test port match s_{22}. The following table summarizes typical specifications for a broadband directional coupler:

Reflection tracking	$\|R\| \geq 0.2$	Corresponds to $\leq$14 dB
Directivity	$\|D\| \leq 0.158$	Corresponds to $\geq$16 dB
Test port match	$\|S\| \leq 0.178$	Corresponds to $\geq$15 dB
Usable frequency	f = 1 GHz to 20 GHz	

Table 2.3.2 Typical specifications for a broadband directional coupler.

For use at frequencies starting at approx. 100 GHz and beyond, directional couplers are typically built using waveguide sections. Although an integrated directional coupler with a center frequency of $f_0 = 180$ GHz and a bandwidth of 120 GHz using a pair of coupled strip lines has been described in [WK99].

In section 2.2.1, the reference channel signal was obtained using a power splitter. This results in a power loss of 6 dB. Accordingly, only 25 % of the output power of the generator can be used to power the directional element. This is a problem particularly when we would like to obtain the highest possible test port output power. If the reference channel signal is obtained using a directional coupler, we can choose, for example, a configuration with $s_{42} = 0.5$. Fig. 2.3.8 shows the relevant test setup. Coupler 1, which serves as a replacement for the power splitter, has a theoretical transmission loss of 1.25 dB. This means that 75 % of the generator output power will reach the directional element (coupler 2). This combination of the two three-port couplers is also known as a **dual directional coupler**. In most practical cases, we will have to accept a coupling loss of approx. 12 dB (corresponding to $s_{42} = 0.25$) to realize a broadband dual directional coupler with a total transmission loss of $2 \cdot 1.25$ dB = 2.5 dB.

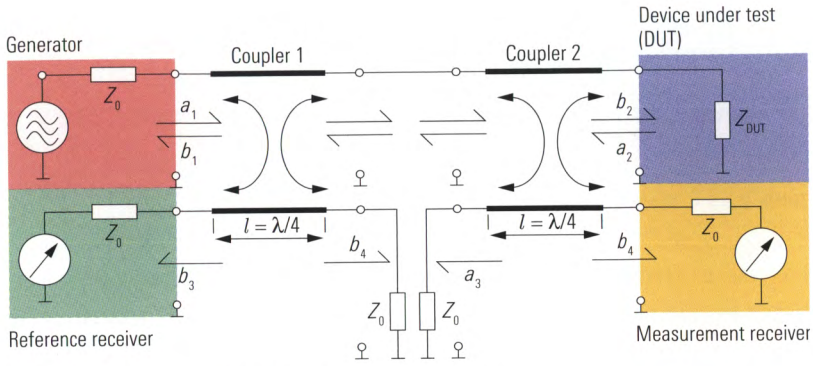


Fig. 2.3.8 Test set with a dual directional coupler.

2.3.3 Other implementations

The characteristic function of the directional element from Fig. 2.2.1 is the isolation of ports 1 and 3 through a corresponding $s_{31} << 1$. This behavior is necessary to achieve the required signal separation of the incident and reflected waves. Besides the implementations we have discussed earlier, a Wilkinson divider can also be used. However, it has worse directivity at the outer limits of its operating frequency range than a comparable directional coupler since it basically acts like a T-junction there. Circulators are also directional elements. However, they have a very small bandwidth but usually a high power capability.

2.4 Other components of the test set

The components described in this section are not always included in a network analyzer, but they do increase its functionality and are commonly available as options.

2.4.1 Receiver step attenuators

With DUTs such as a power amplifier, the output power of the DUT can exceed the linear input range of the measurement receiver. The relevant upper limit of the analyzer's linear input range is usually referred to as the maximum nominal test port input power.

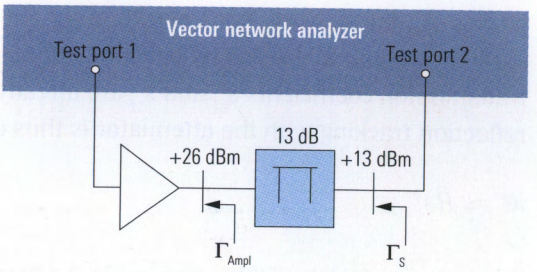


Fig. 2.4.1
Increase in the maximum test port input power at the expense of directivity.

By way of example, we assume a maximum nominal test port input power of 20 mW. The amplifier has an output power of 400 mW, then an attenuator with a value of 13 dB would just be sufficient between the amplifier output and test port 2 to prevent any overdrive from occurring. However, the attenuator shown in blue in Fig. 2.4.1 has significant disadvantages as described hereafter.

At its output, the amplifier has the reflection coefficient Γ_{Ampl}. The attenuator is characterized by its S-parameters $s_{\pi 11}$, $s_{\pi 12}$, $s_{\pi 21}$ and $s_{\pi 22}$. We know that $s_{\pi 12} = s_{\pi 21}$. The following figure shows the directional element and the attenuator. The reference plane for determining the reflection coefficient Γ_{Ampl} is prior to the attenuator so that the attenuator becomes part of the test set. The quantities R, D and S known from section 2.2 have to be redetermined for this arrangement. What we obtain are the quantities R', D' and S', which are referred to the reference plane that is drawn in below.

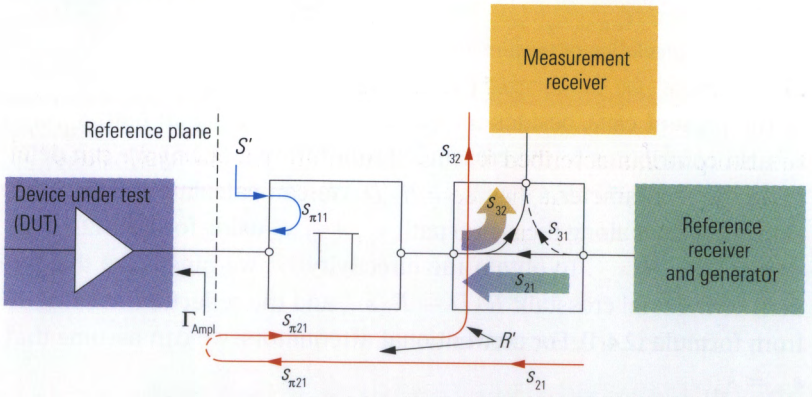


Fig. 2.4.2 *Calculation of the test port match S' and the reflection tracking R'.*

The stimulus signal passes through the directional element and the attenuator. Thus it is attenuated with transmission coefficients s_{21} and $s_{\pi 21}$ before being reflected at the DUT with Γ_{Ampl} (red line in Fig. 2.4.2). It

passes via the transmission coefficients $s_{\pi 21}$ and s_{32} finally to the measurement receiver. In the reflection tracking R from formula (2.2-4), transmission coefficients s_{21} and s_{32} are already taken into account. The reflection tracking with the attenuator is thus calculated as follows:

$$R' \approx R s_{\pi 21}^2 \qquad \textit{(2.4-1)}$$

The test port match S' (blue line in Fig. 2.4.2) is generally dominated by the reflection coefficient s_{11} of the attenuator, which justifies the following approximation:

$$S' \approx s_{11} \qquad \textit{(2.4-2)}$$

The directivity was defined in section 2.2.3 using formula (2.2-6).

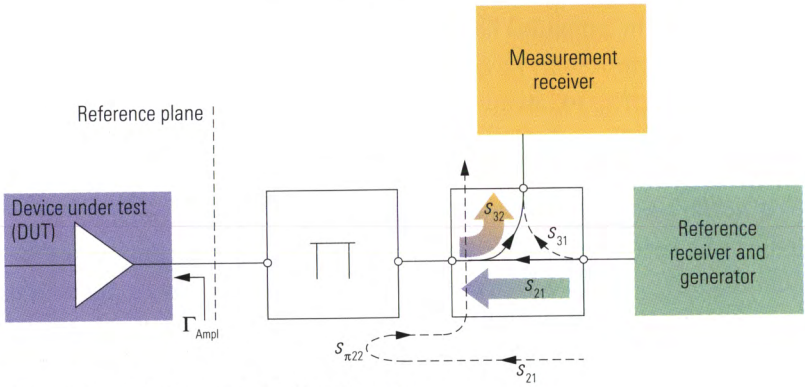


Fig. 2.4.3 Calculation of the directivity D'.

In the present case, besides the parameter s_{31}, the signal path s_{21}, $s_{\pi 22}$, s_{32} also contributes to the crosstalk. From formula (2.2-6), we can determine the S-parameter s_{31} as $s_{31} = R \cdot D$. We can calculate the transmission coefficient along the signal path s_{21}, $s_{\pi 22}$, s_{32} using formula (2.2-4) as $s_{21} \cdot s_{\pi 22} \cdot s_{32} = R \cdot s_{\pi 22}$. To obtain the directivity D', we must take the quotient of the total crosstalk $R \cdot D + R \cdot s_{\pi 22}$ and the reflection tracking R' from formula (2.4-1). For conventional attenuators, we can assume that $s_{\pi 22} = s_{\pi 11}$.

$$D' = \frac{R \cdot D + R \cdot s_{\pi 11}}{R \cdot s_{\pi 21}^2} = \frac{D + s_{\pi 11}}{s_{\pi 21}^2} \qquad \textit{(2.4-3)}$$

The characteristic quantities D and R of the directional element degrade due to usage of the attenuator to the values R', D'. In general, the test port match is determined by the match of the attenuator. The following table shows typical values for R, D and S. For the attenuator, a return loss of 20 dB was assumed along with an insertion loss of 13 dB.

Parameter	Without attenuator	With attenuator
Reflection tracking	$R = 13.5$ dB	$R' = 39.5$ dB ≈ 40 dB
Directivity	$D = 46$ dB	$D' = -6.4$ dB ≈ -6 dB
Test port match	$S = 26$ dB	$S' = 20$ dB

Table 2.4.1 Data degradation in a test set due to a 13 dB attenuator.

The degradation of the directivity from 46 dB to -6 dB is the most unwanted effect that occurs here. Basically, it is possible to perform a calibration which determines the values R', D' and S' and compensates them. However, an arbitrary improvement in accuracy cannot be achieved in this manner. For example, it is not possible to adequately cover the drift of the data R', D' and S' using system error correction. A reflection tracking of 40 dB as opposed to 13.5 dB results in a declined signal-to-noise ratio. If we are not willing to accept this sort of reduction in the measurement accuracy, we will need to use a different test configuration.

Some network analyzers provide access to the measurement channel, in which the 13 dB attenuator (shown in Fig. 2.4.4 in blue) could be inserted in between the directional element and the measurement receiver. The nominal upper power limit of the directional element is usually significantly greater than given by the receiver's linear input range. In our example, we assume that the directional element is designed for a maximum input power of 500 mW. The amplifier's 400 mW output power does not exceed this limit. That is the reason why we can move the attenuator from the test port 2 to the position it is drawn in Fig. 2.4.4.

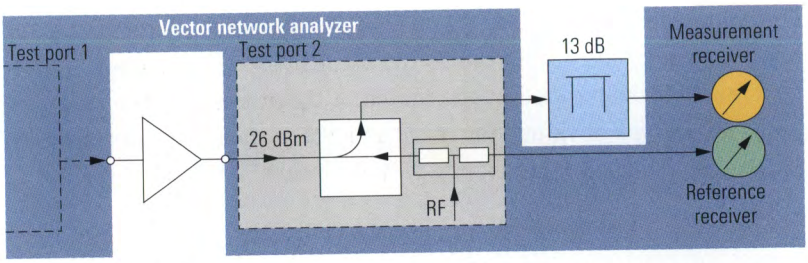


Fig. 2.4.4 Increasing the maximum input power while maintaining the directivity.

Since the directional element is now located directly at the amplifier output, its directivity D is now exploited to the full extent.

If the requirements for the input power change frequently, an internal **receiver step attenuator** should be used. Mechanical attenuators have relatively slow switching times and are subject to wear. To avoid frequent changes in the attenuator setting, emphasis is placed on achieving the largest possible distortion-free input level range in the receivers, when designing modern network analyzers. When a receiver step attenuator is switched, the electrical length and level between the directional element and the measurement receiver are changed. The change in level is compensated computationally as a first approximation. However, recalibration is needed to handle the change in phase and to provide exact compensation of the change in level. If frequent changes in the attenuator settings are required, multiple calibration data sets can be used depending on the desired attenuator position to simplify the overall process.

2.4.2 Generator step attenuators

Generator step attenuators are used to attenuate the stimulus signal. They extend the range of the stimulus signal toward lower power without causing any loss in directivity. This makes it possible to test DUTs with high gain, e.g. a 70 dB amplifier chain. There are basically two approaches to implementing this decrease in power which are shown in Fig. 2.4.5 and Fig. 2.4.6.

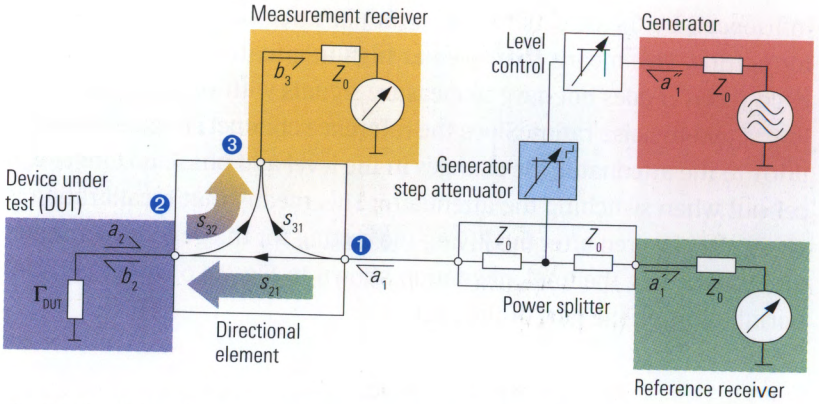


Fig. 2.4.5 Generator step attenuator prior to the power splitter.

In Fig. 2.4.5, the generator step attenuator is shown in blue. It is arranged at the input to the power splitter. Both the stimulus signal a_1 and the reference channel a_1' are obtained after the attenuator. The changes in the level and phase that occur due to different attenuator settings thus affect both signals. The S-parameters are ratios of the b_n wave to the a_k wave. When determining the S-parameters from a_1' and b_3, the changes in the level and phase that are caused by the generator step attenuator are thus cancelled out. Due to the lower level at the reference receiver, the signal-to-noise ratio there is reduced accordingly. However, it is possible to avoid this problem using automatic gain control or a suitable IF bandwidth.

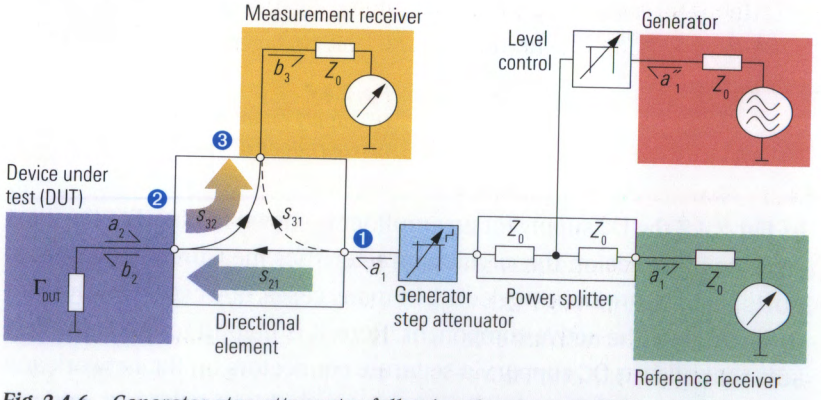


Fig. 2.4.6 Generator step attenuator following the power splitter.

In the implementation shown in Fig. 2.4.6, the attenuator is located after the power splitter so that it only attenuates the signal a_1. Aside from any feedback, the level and phase of the reference channel a_1' are not

influenced by the attenuator, i.e. a_1' is not affected by a large attenuator setting. This means that, even at high attenuator settings, the reference receiver does not have to measure signals with very low levels and low signal-to-noise ratios. Since the reference channel signal is obtained prior to the attenuator, the changes in the level and phase no longer cancel out when switching the attenuator. This means that recalibration is generally required after modifying the setting for the generator attenuator. Accordingly, the implementation shown in Fig. 2.4.6 will entail additional effort on the part of the user.

2.4.3 Active and passive test sets

The **biasing** of active components requires a DC current which is superimposed on the RF signal. The DC current is normally injected by a **bias T**, which has two main components. To influence the RF signal as little as possible the first main component is an inductor between the RF signal and the **DC supply**. Depending on the frequency range a combination of several graduated air-core coils and inductors can be applied. The second main component is a **DC block**. It is necessary because the adjacent RF components are usually not designed to handle a DC voltage. A capacitor or a broadband capacitor can serve for this.

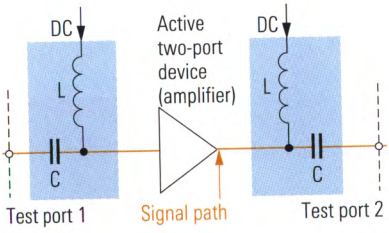


Fig. 2.4.7
DC supply on an active two-port device.

In Fig. 2.4.7, the DC supply for an amplifier is shown in blue. For the most part, its influence on the circuit characteristics measured between test ports 1 and 2 cannot be neglected. In many cases, however, we only want to investigate the active component. Here, it is helpful to use **active test sets** that allow a DC supply via separate connectors on the network analyzer (PORT BIAS). The circuit parts shown in blue in the test setup in Fig. 2.4.7 are then no longer necessary. Almost any DC biasing involves a (frequency dependent) attenuation on the RF path. To prevent degradation in the directivity, any unnecessary attenuation between the DUT and the directional element should be avoided. Accordingly, it is benefi-

cial to arrange the DC supply as shown in Fig. 2.4.8 between the power splitter and directional element (instead of between the test port and the DUT). The directional coupler in Fig. 2.3.5 is designed to handle a superimposed DC voltage. For the VSWR bridge shown in Fig. 2.3.2, suitable capacitors must be inserted at the ground nodes.

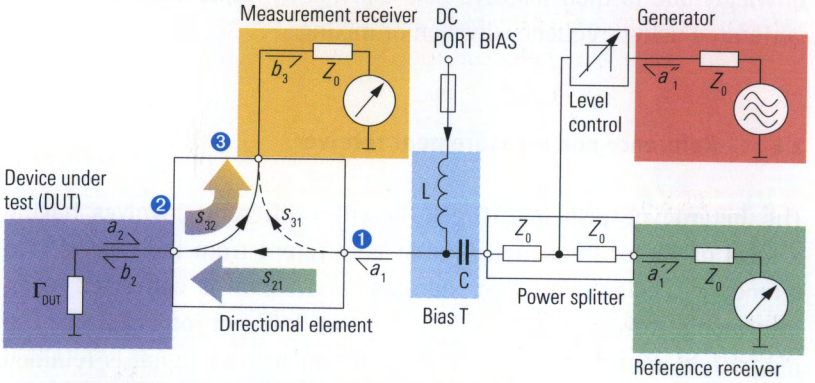


Fig. 2.4.8 Useful configuration for an active test set.

Passive test sets do not have a bias T and are generally not equipped with a DC block. Therefore they can be damaged by a superimposed DC voltage.

2.5 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.

2.6 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. 2.6.1). 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 antialiasing filter for the analog/digital converter.

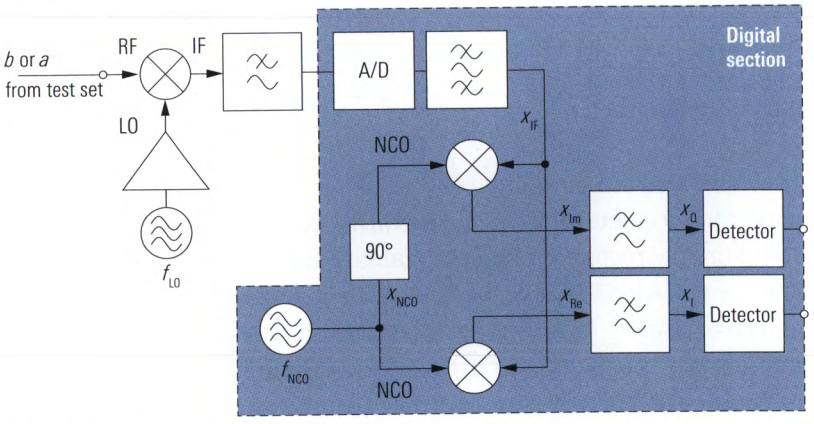


Fig. 2.6.1 Transmission measurement using a heterodyne receiver.

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 arbi-

trarily 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. 2.6.1. 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:

$$x_{IF}(t) = A_{IF} \cdot \cos(2\pi f_{IF} t + \varphi_{IF}) \qquad \textit{(2.6-1)}$$

$$x_{NCO}(t) = A_{NCO} \cdot \cos(2\pi f_{NCO} t) \qquad \textit{(2.6-2)}$$

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

$$x_{Re}(t) = \tfrac{1}{2} A_{IF} A_{NCO} \left[\cos(\varphi_{IF}) + \cos(4\pi f_{NCO} t + \varphi_{IF})\right] \qquad \textit{(2.6-3)}$$

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

The lowpass filtering suppresses the frequency components where $f \neq 0$. This produces the DC signals x_I, x_Q corresponding to the real and imaginary parts of the complex phasor $x_{IF}(f)$. In the field of communica-

1) $\cos\alpha\cos\beta = \frac{1}{2}[\cos(\alpha - \beta) + \cos(\alpha + \beta)]$
$\cos\alpha\sin{-\beta} = \frac{1}{2}[\sin(\alpha - \beta) - \sin(\alpha + \beta)]$

tions engineering, they are known as the **inphase** and **quadrature** components.

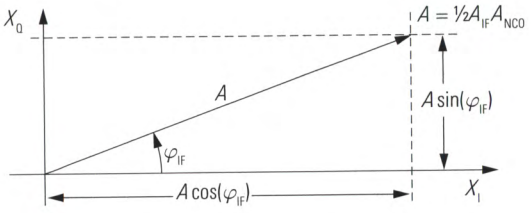


Fig. 2.6.2
Illustration of the quadrature and inphase components.

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. 2.6.3a, 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. 2.6.3b. 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 a_1 wave at the respective 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.

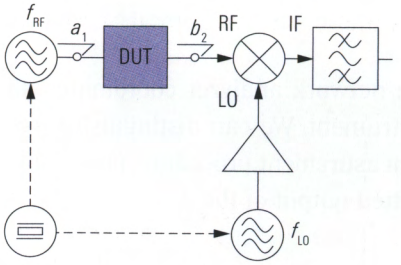


Fig. 2.6.3a Common reference.

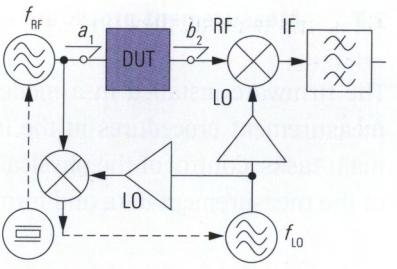


Fig. 2.6.3b Locking of the LO to the RF.

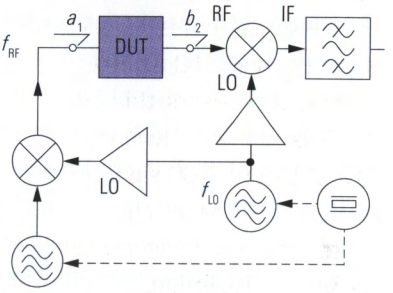


Fig. 2.6.3c Generation of the stimulus signal from the LO.

In Fig. 2.6.3c, the generator signal is produced from the LO signal using an auxiliary oscillator. To ensure that the frequencies f_{RF} and f_{LO} are stable, the auxiliary oscillator must be synchronized to the common reference oscillator. Compared to the implementation in Fig. 2.6.3b, 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 $f_{LO} = \frac{1}{3}(f_{RF} \pm f_{IF})$ 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.

2.7 Measurement procedure

The firmware installed in a modern network analyzer coordinates all measurement procedures in the instrument. We can distinguish three main tasks: Control of the physical measurement procedure, processing of the measurement data and formatted output of the data.

2.7.1 S-parameter measurement procedure

The complete S-matrix for an N-port network comprises N^2 S-parameters. As the primary measurement quantities, the network analyzer measures the (complex) wave quantities a and b. The measurements required to assess an N-port network are subdivided into N partial measurements. In each **partial measurement**, a different one of the N test ports is operated as the active port. All of the remaining ports are considered to be passive and having ideal test port match[1]. Each partial measurement covers the following waves:

- The outgoing wave at the active test port
- The wave reflected to the active test port
- The waves transmitted to the remaining $N-1$ test ports

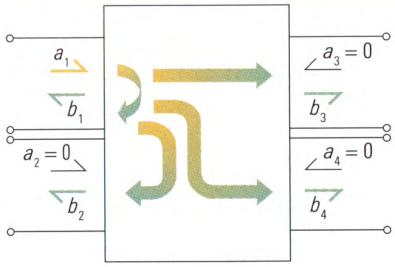


Fig. 2.7.1
Partial measurements with active test port 1.

Fig. 2.7.1 illustrates this situation for the partial measurement with active test port 1 using the example of a four-port network.

1) In the real world, the test ports have test port match $S \neq 0$ (see section 2.2.4). This means that the ports of the DUT are not ideally terminated during measurement; also to be considered are further effects. Accordingly, it is usually necessary to subject the S-matrix determined here to a system error correction procedure (see chapter 3).

This partial measurement corresponds to equation set (2.7-1). It can be used to determine the first column of the S-matrix. The other partial measurements yield the remaining columns in the S-matrix.

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = S \begin{bmatrix} a_1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{aligned} b_1 &= s_{11}a_1 \\ b_2 &= s_{21}a_1 \\ b_3 &= s_{31}a_1 \\ b_4 &= s_{41}a_1 \end{aligned} \qquad \textit{(2.7-1)}$$

In **alternating mode**, each partial measurement is performed across the entire stimulus range (e.g. frequency sweep) before the analyzer begins the next partial measurement. Some network analyzers use **chopper mode**. During a frequency sweep in this mode, the analyzer performs all of the partial measurements at a given frequency point before moving to the next frequency point. In this case, alternating mode is generally available as an alternative.

2.7.2 Measurement data processing chain

At the start of the **processing chain** there are **raw numerical measured values** from the reference and measurement receivers. In Fig. 2.7.2, the processing steps are represented using arrow-shaped symbols and the data objects are represented using barrel-shaped symbols. The processing steps with a gray background cannot be skipped.

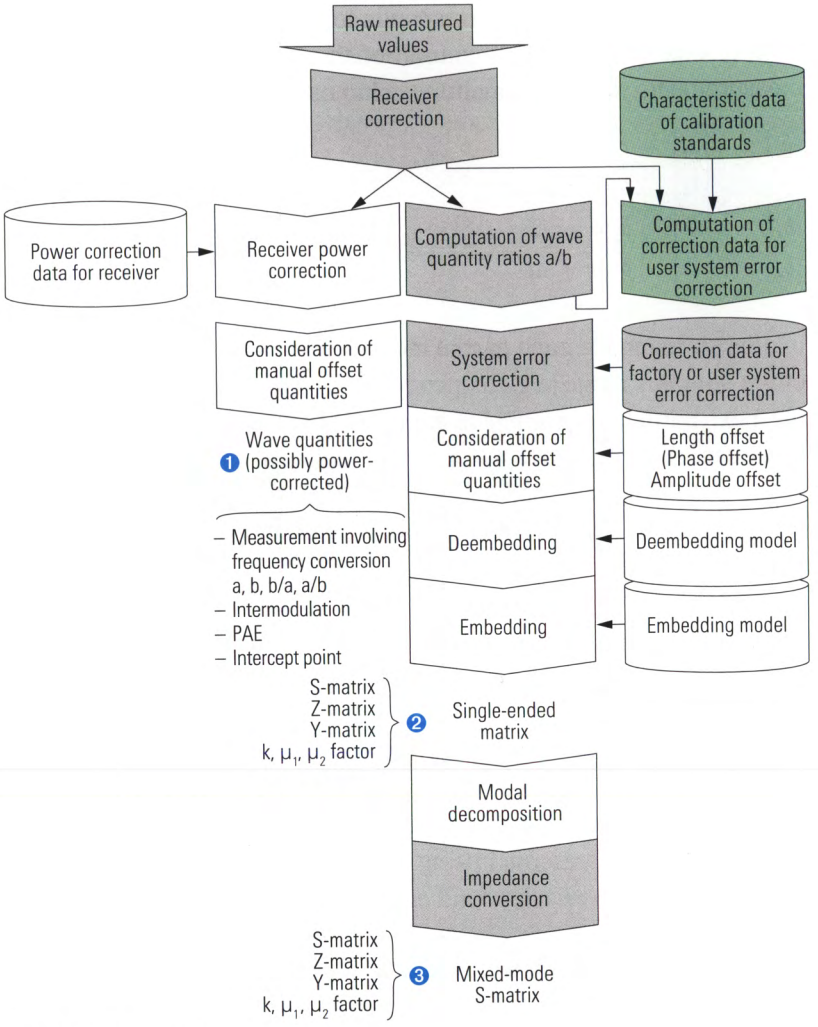


Fig. 2.7.2 Processing chain, part 1.

The processing chain is considered to be completed as soon as the required **measured values** have been determined. The associated endpoints in the processing chain are shown in Fig. 2.7.2 using the numbers ❶ to ❸. The procedure marked in green is required only for **user calibration** and ends with determination of the correction data. Section 4.1 gives an example of how this calibration is performed. If user calibration is not activated, the **factory system error correction** is used.

2.7.3 Trace generation

The second part of the processing chain (**trace generation**) is shown in Fig. 2.7.3. It handles all of the computations needed to display the results. The complex input quantities are taken from either the first part of processing chain, the main memory or a data file. All of the processing steps with a light blue background involve complex values.

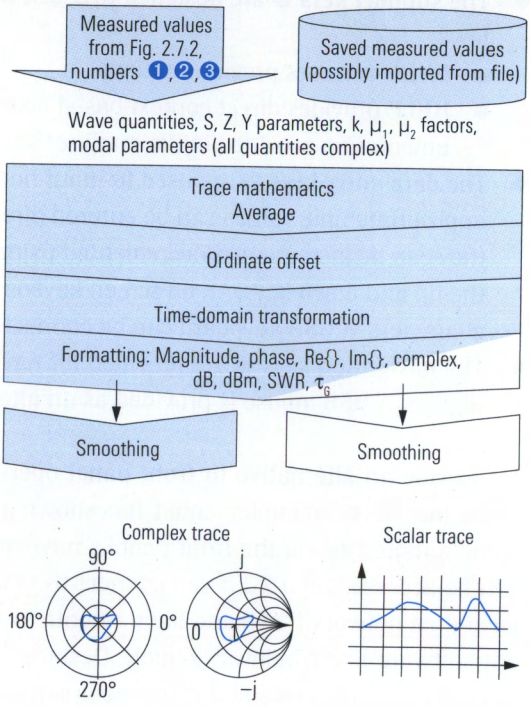


Fig. 2.7.3
Processing chain, part 2.

2.8 Main setting parameters

2.8.1 User interface

We will consider the **user interface** based on the example of a network analyzer family. Fig. 2.8.2 shows the front panel of the vector network analyzer R&S®ZVA8. The control keys are divided into five groups based on their functionality:

- Depending on the current operating context, the **softkeys** ❶ have different functions, which helps to reduce the number of front

panel keys required to a reasonable amount. At any time, functions that are not needed are hidden from view.

- Using the **setup keys** ❷, it is possible to directly access the basic operating functions. They are divided into four groups. A popular key is the **MEAS WIZARD**. It invokes a wizard that makes the necessary instrument settings based on an interactive dialog with the user.
- The **support keys** ❸ are designed to assist the user. They are as follows:
  - **UNDO** reverses previous actions
  - **HELP** provides direct context-based access to the extensive documentation provided in the analyzer's help system
- The **data entry keys** ❹ are used to input numerical values with the appropriate unit. Values can be entered directly on the keypad or they can be incremented/decremented using the rotary knob or the up and down keys. An on-screen keyboard can be used to enter characters. A USB keyboard can be connected instead.
- The **navigation keys** ❺ are provided for navigation in input fields or dialogs. A USB mouse is provided as an alternative.

To provide an alternative to front panel operation, a typical Windows® **menu bar** has been implemented. It is shown in 2.8.1. It can be operated using a mouse or via the front panel's navigation keys. Furthermore it can be accessed on a PC if you connect the PC to the instrument via a local area network (LAN) and login remote to the instrument with a client software like Windows® Remote Desktop. This makes it possible to have the user interface and instrument separated over great distances and controlling the instrument as a remote user (for details see page 72).

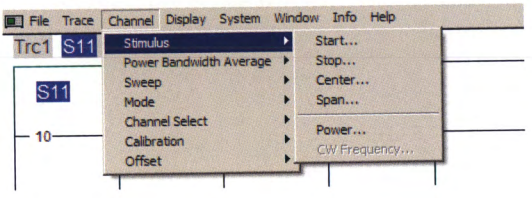


Fig. 2.8.1
Menu bar in the Windows application.

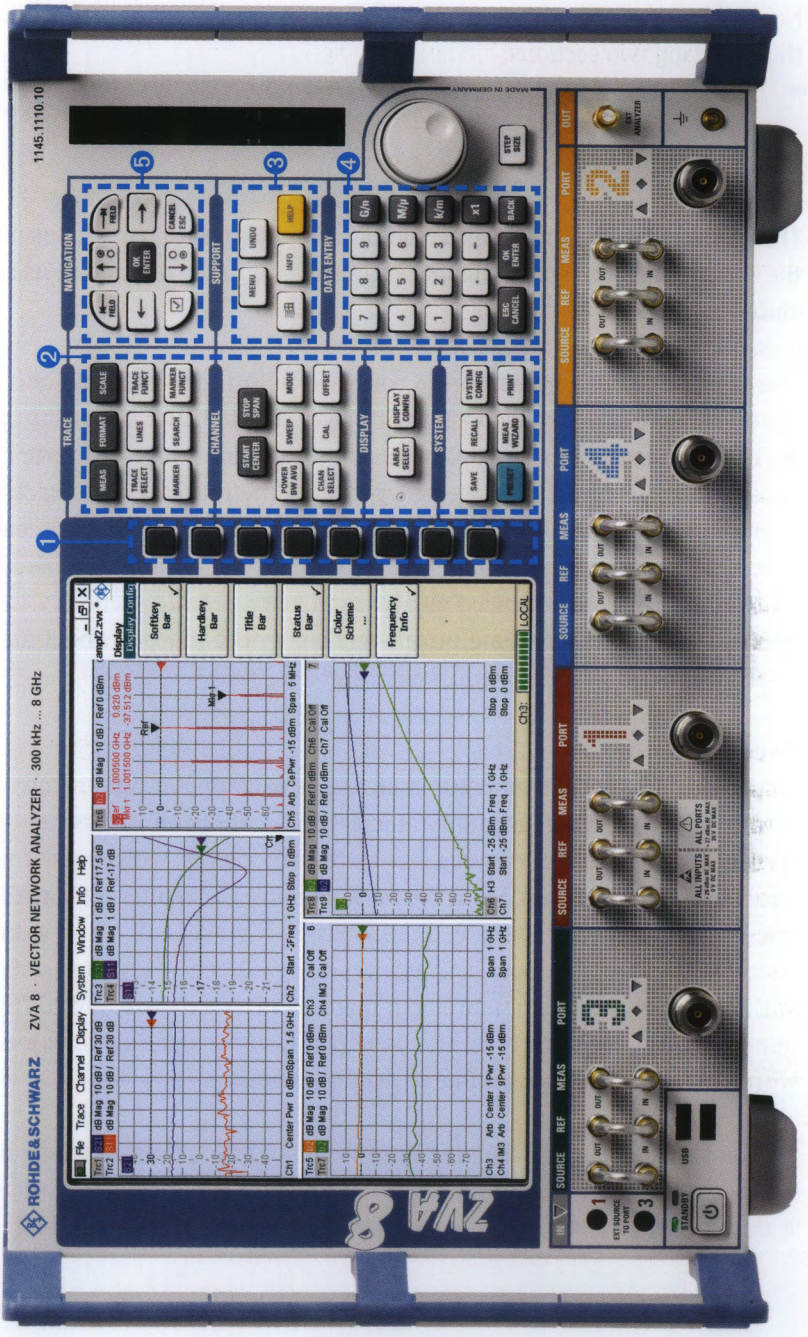


Fig. 2.8.2 *User interface of the Vector Network Analyzer R&S®ZVA8.*

Nearly all network analyzers allow the basic settings that are covered in the following two sections.

2.8.2 Channel settings

The **channel settings** have a direct impact on the physical conditions of the measurement. A set of channel settings is normally referred to as a **channel**.

The term "channel" should not be confused with the terms "reference channel" (wave quantity *a*) and "measurement channel" (wave quantity *b*). To be able to measure multiple traces with different channel setting parameters in parallel, it is necessary to map the measurement traces to different channels. The network analyzer then carries out the measurements in sequence and uses the relevant channel settings. This is relevant only for the advanced measurement examples in sections 7.4.2, 7.5.2 and 8.3. For the first time, beginners do not need to worry about the term "channel".

Sweep type

The **frequency sweep** is most common and is used to measure the characteristics of the DUT (e.g. S-parameters) as a function of frequency. With a **linear frequency sweep**, measurement points with equal frequency spacing are used. Many network analyzers also offer a **logarithmic frequency sweep**.

With a **time sweep**, the measurement quantity is observed at a fixed frequency as a function of time. This is useful for investigating networks with variable behavior over time or drift phenomena.

The **power sweep** is based on the time sweep but the power of the stimulus signal is varied during the sweep. Hence the measurement quantity is determined at a stationary frequency as a function of the input level.

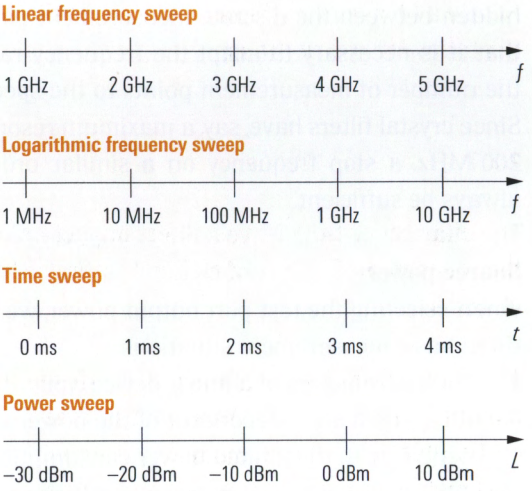


Fig. 2.8.3
Examples of stimulus axes (sweep types).

Sweep range

The sweep range to use for a measurement can be defined using start and stop values or the center value and the span. In case of a frequency sweep, this is either the **start** and **stop frequency** or the **center frequency** and the **span**.

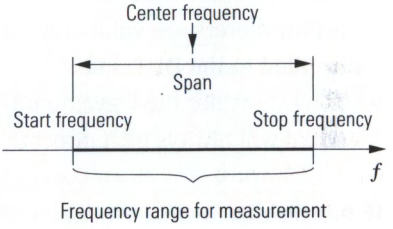


Fig. 2.8.4
Illustration of the stimulus range.

Number of points

Network analyzers normally perform a point by point sweep. This means that the behavior of the DUT is assessed only at discrete points in the sweep range. In case of a frequency sweep, the individual discrete measurement points $n = 1, 2, 3, \ldots N$ are calculated as follows from the start frequency f_{Start}, stop frequency f_{Stop} and number of points N:

$$f(n) = f_{Start} + (n-1)\frac{f_{Stop} - f_{Start}}{N-1} \qquad \textbf{(2.8-1)}$$

For a span from, e.g. 0 Hz to 24 GHz and a resolution of 101 points, the behavior of the DUT is assessed at intervals of 240 MHz. With this setting, the resonance curve for a narrowband crystal filter with, for example, a center frequency of 100 MHz and a bandwidth of 1 MHz will be

hidden between the discrete measurement points. The example shows that it is necessary to adapt the frequency range to be displayed and the number of measurement points to the specific behavior of the DUT. Since crystal filters have, say, a maximum resonance frequency of about 200 MHz, a stop frequency on a similar order of magnitude should always be sufficient.

Source power

When selecting the test port output power, we can distinguish between three basic measurement situations:

- The S-parameters of a linear device under test are pure ratio quantities which are independent of the power used for the measurement. Due to the limited power consumption (possible destruction limit), however, the source power must comply with the limits of the device under test. For a device under test such as a high power amplifier the maximum input power of the network analyzer must also be considered. For additional protective measures when testing components of this type, see section 4.7.
- An entirely different situation arises with nonlinear DUTs since their behavior is dependent on the power level. The measured S-parameters are valid only for the source power that is currently applied to the DUT.
- If we consider the wave quantities instead of the S-parameters, there will always be a direct dependence on the source power.

IF bandwidth (measurement bandwidth)

The IF bandwidth (also known as the **measurement bandwidth**) corresponds to the bandwidth of the IF filter. This filter is usually implemented using digital signal processing. Typical IF bandwidth values are, for example, 10 Hz, 100 Hz, 1 kHz, 10 kHz, 100 kHz, 1 MHz, other values are possible depending on the instrument model. As a first approximation, we can assume the power density of the broadband noise is uniformly distributed over the frequency axis (see Fig. 2.8.5). In other words, the smaller the IF bandwidth we choose, the less broadband noise the receiver will pick up along with the actual signal of interest. This means that a small IF bandwidth basically increases the dynamic range available with the measurement. On the other hand, the time that the IF filter needs to settle is dependent on its bandwidth. If we cut the IF bandwidth in half, for example, this leads to a doubling of the sweep time.

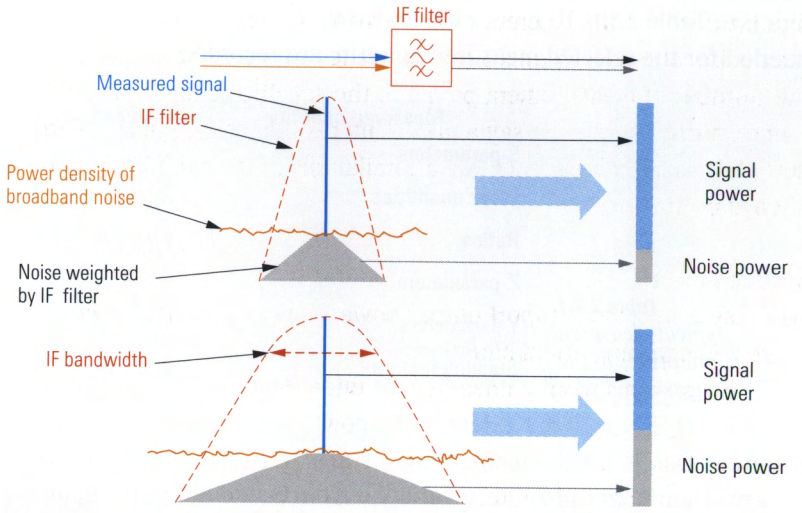


Fig. 2.8.5 Illustration of the changes in signal-to-noise ratio using different IF bandwidth settings.

If we choose an IF bandwidth that is too large, this can result in a loss of some of the detailed information. Accordingly, the IF bandwidth should be several orders of magnitude smaller than the frequency selectivity of the device under test.

Sweep time

The sweep time is the time required to perform a single sweep. A network analyzer normally attempts to minimize the sweep time based on the requirements for internal settling (IF filters and so on). When analyzing DUTs with unusual settling characteristics, it is necessary in some cases to manually select a longer sweep time.

2.8.3 Trace settings

The **trace settings** only influence how the measurement results are reported. Each **trace** has a set of individual trace settings.

Measured quantity

Besides measuring all the S-parameters of a DUT, a vector network analyzer offers several other measured quantities that are directly related to the wave quantities or that can be derived from a set of S-parame-

ters (see Table 2.8.1). To reduce the sweep time, only the wave quantities needed for the selected measured quantity are recorded.[1)]

Table 2.8.1 Typical measured quantities of the network analyzer.

Measured quantity	Examples
S-parameters	$s_{11}, s_{21}, \ldots$
Wave quantities	$a_1, b_1, \ldots$
Ratios	$b_1/a_1, b_2/b_3, \ldots$
Z-parameters	$z_{11}, z_{21}, \ldots$
Y-parameters	$y_{11}, y_{21}, \ldots$
Stability factors	k, μ_1, μ_2

Diagram

A network analyzer provides a variety of charts and diagrams adjusted to the different measurement tasks. In the **Cartesian diagram**, the magnitude or phase of the measured value is plotted vs. the stimulus axis. Examples of the stimulus axis can be seen in Fig. 2.8.3.

In contrast to this, a complex diagram assigns each measured value a point of the complex plane. The x and y coordinate of the point are derived from the real and imaginary part of the measured value. The collection of points is forming a continuous figure that is also known as a locus curve. The parameter along this curve is given by the sweep type (usually the frequency). The following complex diagrams are typically supported by a vector network analyzer:

- **Smith chart**
- **Inverted Smith chart**
- **Polar diagram**

The three types of complex diagrams mentioned above differ only in terms of the auxiliary coordinate system that is displayed! When changing between the different types of complex diagrams, the location of the measurement points remains the same (cf. following Fig. 2.8.6e-g). If a reflection factor such as $s_{11}, s_{22}, \ldots$ is displayed the real and imaginary parts of the normalized complex impendance $z = r + jx$ can be read off directly by the auxiliary coordinate sytem of the Smith chart. Doing the same the inverted Smith chart leads to the normalized admittance

1) This is not true if a type of system error correction is activated that requires additional wave quantities in its computations (e.g. for a one-port measurement following a TOSM calibration; see section 3.4).

$y = g + jb$. The polar diagram enables direct access to the magnitude and phase of the reflection factor. It is also used with other measured quantities.

Format

Formatting takes place prior to graphical display.

Format
dB magnitude
Linear magnitude
Phase
Unwrapped phase
Group delay
SWR, VSWR
Real part
Imaginary part
Complex

Table 2.8.3
Typical formats.

A detailed insight into the formats and their usages is given in section 4 by the way of examples. In the following figures, the same reflection measurement is presented using different diagrams and formats.

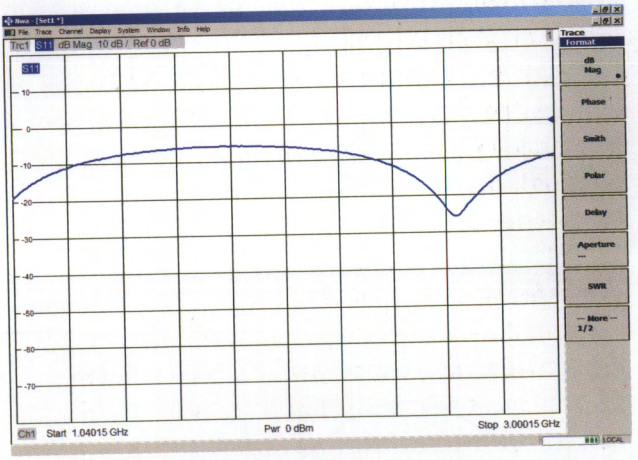


Fig. 2.8.6a
Cartesian diagram with magnitude of S_{11} in dB.

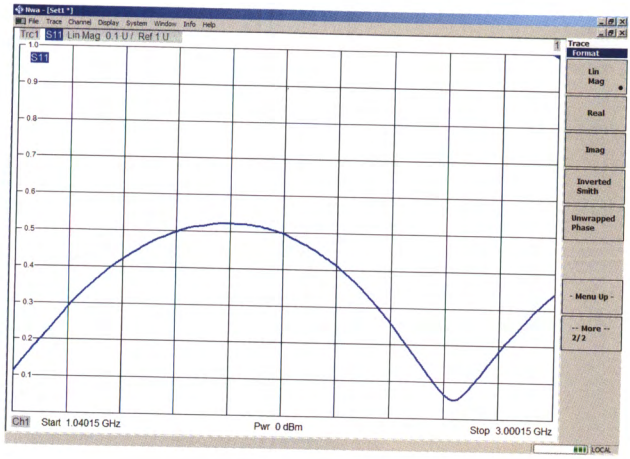


Fig. 2.8.6b Cartesian diagram with magnitude of S_{11} linear.

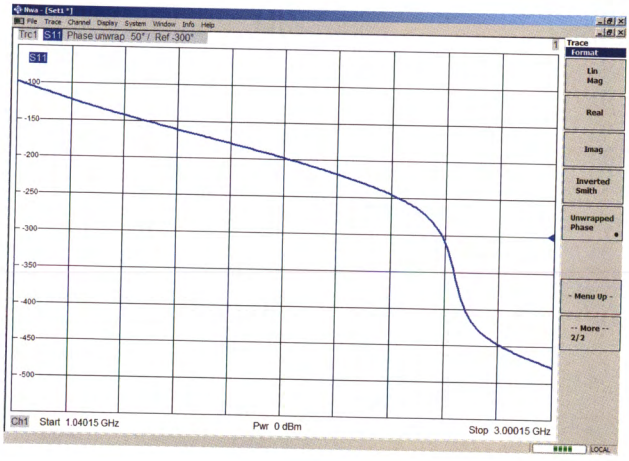


Fig. 2.8.6c Cartesian diagram with unwrapped phase of S_{11}

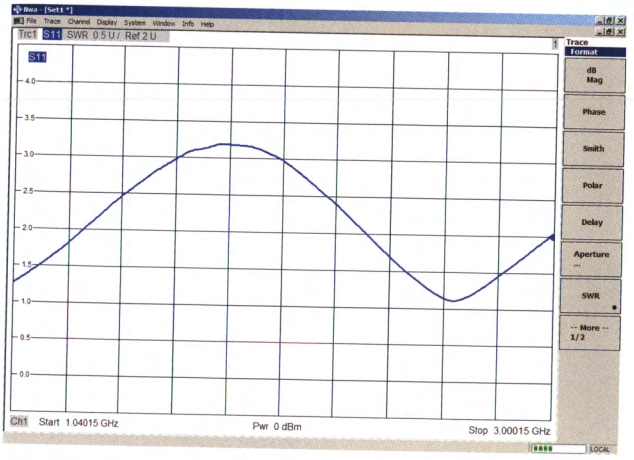


Fig. 2.8.6d Cartesian diagram with SWR

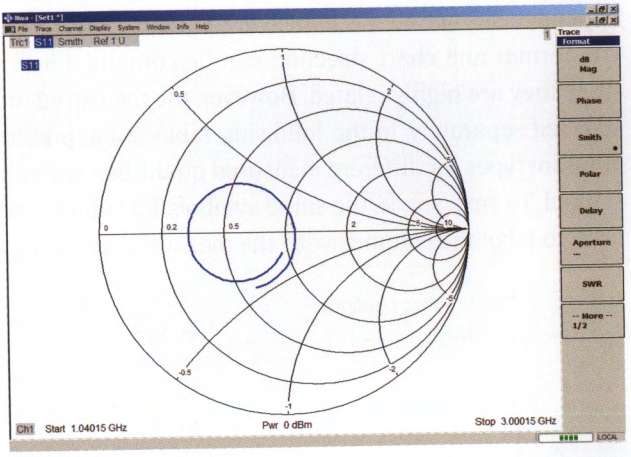


Fig. 2.8.6e
Smith chart

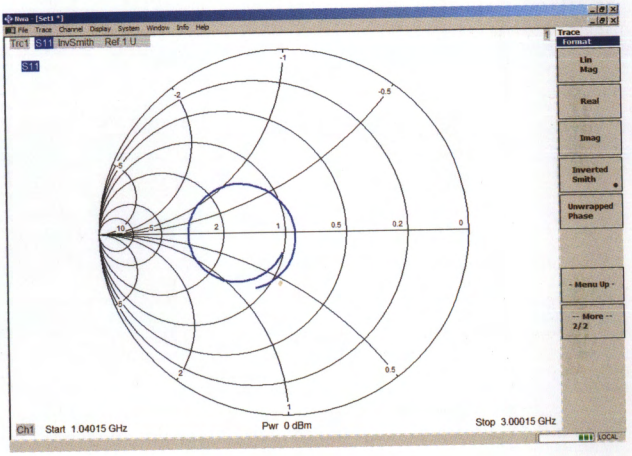


Fig. 2.8.6f
Inverted Smith chart

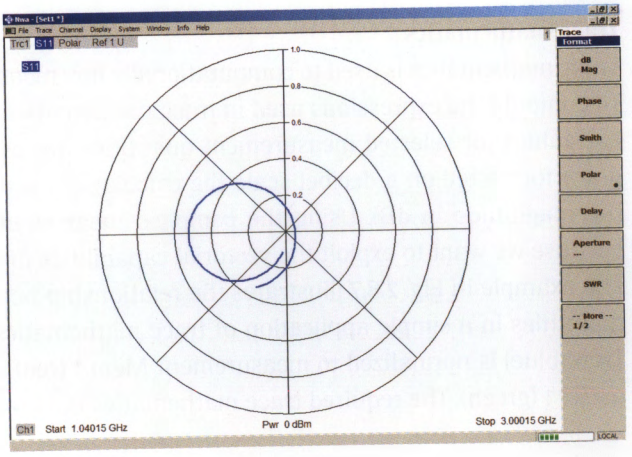


Fig. 2.8.6g
Polar diagram

Dependencies between the measured quantity, diagram and format

The format and chart selection can be combined in the user interface since they are highly related. However, the measured quantity is usually selected separately. In the following table, some preferred formats and diagram types for different measured quantities are indicated with the ✓ symbol. To save space, the same symbols introduced in Table 2.8.1 were used to label the columns with the measured quantities.

Measurement quantity								Format	Diagram
s_{ii}	s_{ik}	a, b	b_i/a_k	b_k/a_k	z	y	k, μ		
✓	✓	✓	✓	✓	✓	✓		dB magnitude	Cartesian diagram
✓	✓	✓	✓	✓	✓	✓	✓	Linear magnitude	
✓	✓		✓	✓	✓	✓		Phase	
✓	✓		✓	✓	✓	✓		Unwrapped phase	
	✓							Group delay	
✓								SWR, VSWR	
✓	✓		✓	✓	✓	✓	✓	Real part	
✓	✓		✓	✓	✓	✓		Imaginary part	
✓								Complex	Smith chart
✓									Inverted Smith chart
✓	✓		✓	✓	✓	✓			Polar diagram

Table 2.8.4 *Preferred combinations of the format, measured quantity and diagram.*

Trace mathematics

Trace mathematics is used to computationally link measurement results. To compute the expressions used in trace mathematics, the complex linear values for selected measurement quantities are used (e.g. in a diagram formatted on a decibel scale the calculations are not done using the magnitude in dB). Using the complex linear values makes sense because we want to exploit the vectorial capabilities of the instruments. The example in Fig. 2.8.7 illustrates the relationship between computed quantities in a simple application of trace mathematics. Measurement Trc6 (blue) is normalized to measurement Mem4 (red) and is displayed as Trc1 (green). The required trace mathematics is as follows:

$$Trc\,1 = \frac{Trc\,6}{Mem\,4} \qquad \textit{(2.8-2)}$$

If we choose to display the magnitude on a dB scale, then the trace associated with Trc 1 corresponds to the difference between traces Trc 6 and Mem4 (see the upper half of Fig. 2.8.7). When referred to the phases, formula (2.8-2) is equivalent to the subtraction arg(Trc 1) = arg(Trc 6) – arg(Mem 4) (see the lower half of Fig. 2.8.7).

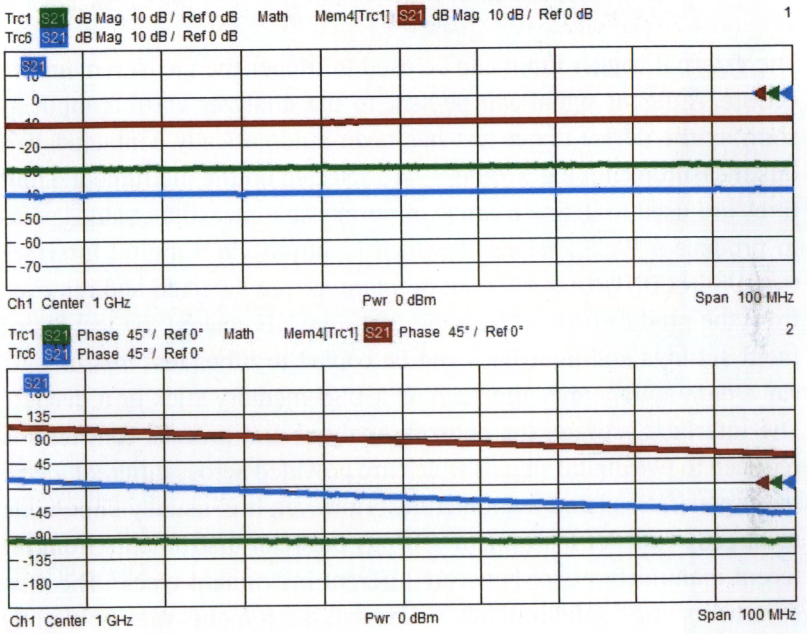


Fig. 2.8.7 Trace mathematics example.

Markers

Markers are useful for convenient readout of the measurement results. They can be freely placed on the trace or they can be used to determine the values of specific trace points (e.g. maximum, minimum, bandwidth). A numeric display field is associated to the markers which shows their position in the sweep range along with the measured value.

2.9 Remote control of the instrument

One of the main applications of network analyzers involves final testing of components and circuits. Accordingly, the instruments must be easily incorporated into fully automated production lines. Here, we can distinguish two main approaches.

2.9.1 Usage of simple digital signals

The **external trigger input** can be used to trigger the measurement procedure. A digital signal can be sent to the analyzer via this input as soon as the device under test has been automatically contacted. Various measurements, e.g. s_{21} and s_{11} are selected on the instrument. **Limit lines** are used to define a range of permissible measured values so as to produce a PASS/FAIL result which is output via a **digital interface** (e.g. USER CONTROL). DUTs out of tolerance are detected and removed from the production line based on this signal. If desired the test instrument settings and limit lines can be copied to other test instruments via a data storage medium such as a USB memory stick or a diskette. The interface between the network analyzer and process controller is reduced to two digital signals which are provided across different instrument generations and manufactures. However, it is usually impossible to directly transfer instrument settings between instruments from different manufacturers or between different instrument types.[1] The measurements and configurations of different instruments may be slightly different, too. Moreover, this approach only allows relatively simple measurements.

2.9.2 Protocol-based interfaces

Besides the simple digital interfaces described above, most network analyzers also have protocol-based interfaces. The figure below shows some of the most common connectors used with protocol-based interfaces.

1) Normally, the firmware in a network analyzer has a mechanism to automatically import instrument settings that were created with earlier firmware versions.

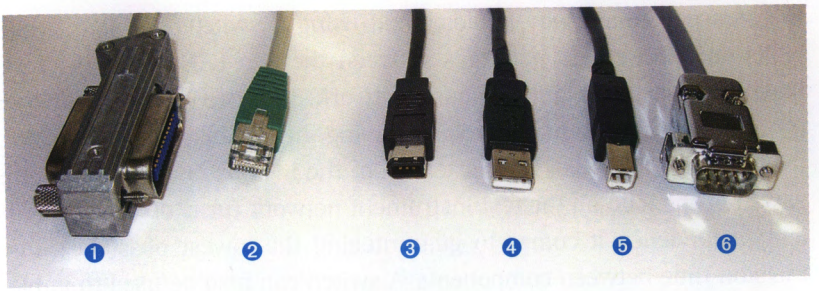


Fig. 2.9.1 Remote control cables with connectors:
❶ IEC bus (GPIB bus), ❷ LAN, ❸ FireWire™, ❹ USB-A, ❺ USB-B, ❻ RS-232-C.

The **IEC bus/IEEE bus** has a 24-pin interface that is defined in the standards IEC 625 and IEEE 488. This standard is often referred to as **GPIB** (General Purpose Interface Bus) and requires a computer with a GPIB controller. Several test instruments can be directly connected to this controller and distinguished using different addresses. The connecting cables that are used have connectors which allow interconnection of multiple cables (Fig. 2.9.1 ❶). Communication is normally initiated by the GPIB controller. A GPIB controller can be simply added to any PC through the addition of a commercially available plug-in board or an USB/GPIB adapter. Network analyzers can also be fitted with a **second IEC bus interface** with a controller functionality. The second interface is used to control external instruments such as power meters and RF generators which form a subsystem with the analyzer. In network analyzers with a USB interface, the second IEC bus interface can also be implemented using a USB/GPIB adapter available as an accessory for some instruments.

The **LAN interface** is used to connect the test instrument to an Ethernet network (IEEE 802 standard), and can also be used for automation purposes. Details can be found, for example, in the application note [1EF47]. The network analyzer is addressed via a unique IP address. The LAN connectivity of the network analyzer is normally implemented by an RJ45 female connector. The male connector on the network cable is shown in Fig. 2.9.1 ❷. By using a crossover cable, it is possible to directly connect the controller PC and the test instrument. However, it is more common to use an Ethernet hub or switch which allows interconnection of several computers and test instruments to form a local area network (LAN). The connection between this LAN segment and an in-house network or

the Internet is normally implemented using a switch. A switch transports only those data packages between the two LAN segments (test instrument network and in-house network) that are intended for the other network segment. This avoids having the test instrument network overloaded with irrelevant packets from the in-house network. Minimization of the load on the test instrument network turns out to be a critical factor when it comes to guaranteeing the lowest possible transmission time between components. A switch can also be used to implement a firewall. Its job is to prevent unauthorized traffic between the Internet and the test instrument network. Depending on the implementation of the interface and network, the nominal transmission speed is 10 Mbit/s for a 10Base-T LAN, 100 Mbit/s for a 100Base-T LAN and 1 Gbit/s for a Gigabit LAN.

Besides the potential of automation, direct access to the user interface of the network analyzer is also possible. Management of the user interface is handled by software such as Remote Desktop in Microsoft Windows® and other programs such as pcANYWHERE® from Symantec or VNC (Virtual Network Computing).[1] This enables remote maintenance (e.g. firmware updates) from anywhere in the world via the Internet.[2]

With **USB interfaces**, we must distinguish between the USB host and devices. The host handles management of the USB system and assigns addresses to the individual USB devices including time slots to be used for communication. At the host (a network analyzer or a computer), the USB-A connector shown in Fig. 2.9.1 ❹ is connected. The USB-B connector in Fig. 2.9.1 ❺ is used for the USB device. Standard USB connecting cables have a USB-A connector on one end and a USB-B connector on the other end. This makes it impossible to create a direct electrical connection between two hosts or two USB devices. A **USB hub** is used to connect multiple USB devices to a host or to another USB hub. A tree

1) Remote Desktop access to the instrument is usually disabled by default. Programs like pcANYWHERE® or VNC are usually not included and have to be installed on the instrument prior to their first usage. Depending on your solution cross platform sessions are possible (e.g. between Microsoft Windows® and Linux).

2) With connections involving public networks like the Internet, any direct access to/from the in-house network should be blocked by a firewall. In this case, usually a virtual private network (VPN) has to be configured. It should involve a so called cryptographic tunneling protocol and hide the involved instrument, its client and the interchanged data from the public network. In case of the slightest doubt regarding a secure network connection, please consult your local network administrator for advice.

structure is created in this manner. The host is located at the root of the tree. Modern network analyzers usually have several USB-A connectors. Some typical USB devices are as follows: Keyboard, mouse, memory stick, printer, power meter and automatic calibration equipment. The nominal transmission speeds are 1.5 Mbit/s or 12 Mbit/s for USB 1.1. With USB 2.0, a nominal transmission speed up to 480 Mbit/s is possible. Depending on the transmission mode, for example, instead of 12 Mbit/s, an effective data transfer rate of 500 kbit/s to 9.7 Mbit/s is achieved. The USB devices can be powered via the USB bus with up to 500 mA and a voltage of 5 V. Since multiple USB devices can be operated on a USB hub, it might require in some cases an external power supply to meet the power requirements of the devices that are connected. This is known as an active USB hub.

The RS-232-C interface is not used very frequently in modern instruments. The FireWire™ (or IEEE 1394) interface has not become popular in automation applications.

2.9.3 Automation

Introduced in 1990, the **SCPI standard** (Standard Commands for Programmable Instruments) is a set of commands based on ASCII characters that are used to control programmable test and measuring instruments. The standard was designed to enable compatibility of instruments made by different manufacturers or from different product series. Of course, this is possible only to the extent that the instrument functions are identical. This is known as "vertical compatibility". "Horizontal compatibility" is the ability to control common instrument functions in different instrument types such as the identification process in generators, spectrum analyzers, network analyzers and so on in a uniform manner. This is based on the unified instrument model defined in the standard (see Fig. 2.9.2). It includes a function group for the measurements which are implemented, for example, in spectrum analyzers, power meters, volt meters and so on. The signal generation function group is relevant, for example, with signal generators, voltage sources and current sources. A network analyzer has both function groups. In addition, all instruments have function groups related to formatting, triggering and storage. The commands in the standard are structured hierarchically. The prefix for a command can be seen in the following figure. User documentation

includes a list of the instrument commands and describes how they are used. A special notation for commands is customary for documentation purposes. For example, the notation "`TRIGger:SOURce EXTernal`" means that the command has the short form "`TRIG:SOUR EXT`" or the complete form "`TRIGGER:SOURCE EXTERNAL`". Both forms are used for communication with the test instrument and can be mixed as required (e.g. "`TRIG:SOURce EXT`"). SCPI commands are not case sensitive.

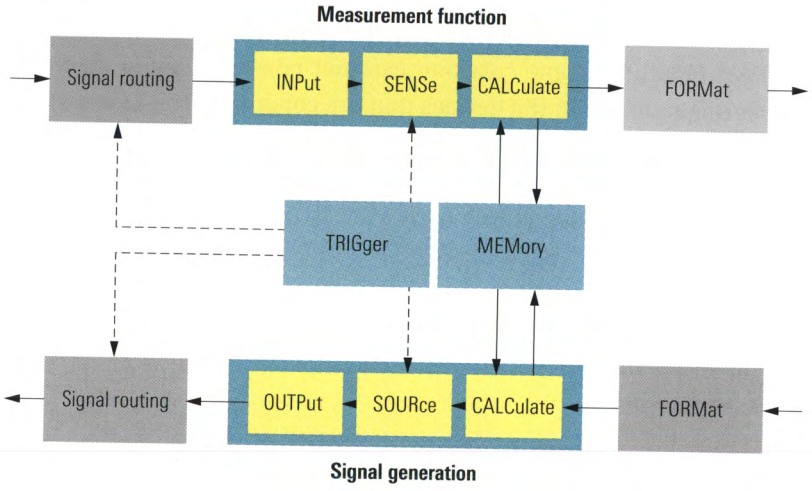


Fig. 2.9.2 The unified instrument model in the SCPI standard.

We can distinguish between **commands**, which are messages that are sent from the controller to the instrument, and **device responses**, which are messages in the opposite direction. In programming terms, we can distinguish the following procedures:

- **Setting commands** modify the instrument settings. For example, the instrument can be reset with "`*RST`" or a specific measurement bandwidth can be selected with "`SENS:BWID 10kHZ`".
- **Queries** cause data to be output, e.g. a marker value can be queried with "`CALC:MARK1:Y?`". The instrument response is captured by the control computer using a read sequence. As a basic rule, most setting commands can also be formulated as query commands. For example, the test instrument responds to "`SENS:BWID?`" with the current measurement bandwidth value.
- **Error handling** is implemented using query commands that affect special status registers.

A detailed treatment of SCPI and GPIB can be found in the relevant literature [Pi00]. When using a PC plug-in board or a USB/GPIB adapter, the available programming languages will depend on the drivers associated with this interface hardware. Additional information can be obtained from the manufacturer of the interface hardware. Most manufacturers will offer a datasheet with the relevant information. If the LAN interface is used for communication, a I/O-library (e.g. a DLL) supplied by the test instrument manufacturer must be installed on the control computer usually. More information about installing this DLL is found, for example, in application notes [1EF46] and [1EF47]. The homepage of the test instrument manufacturer will usually provide information about the programming languages that are supported. The most common programming languages are as follows: C, Microsoft Visual C++, Microsoft Visual Basic, MATLAB® and the test development environments LabWindows/CVI and LabVIEW. Section 4.13 includes an example of automation using MATLAB® and the LAN interface.

The ASCII sequences exchanged between the instrument and control computer are independent of the physical interface. However, the framework used to establish a connection, configure it, send a command and read in a device response is dependent on the I/O-library used on the control computer. The I/O-library is dependent on the physical interface, among other things. To create a unified interface in this area, a consortium of major manufacturers of test instruments created the **IVI standard** (Interchangeable Virtual Instruments). The **VISA I/O library** has been defined. It serves as a unified functional library that is largely independent of the physical interface and other criteria (see Fig. 2.9.3). This makes it easier to port programs from one communications interface to another.

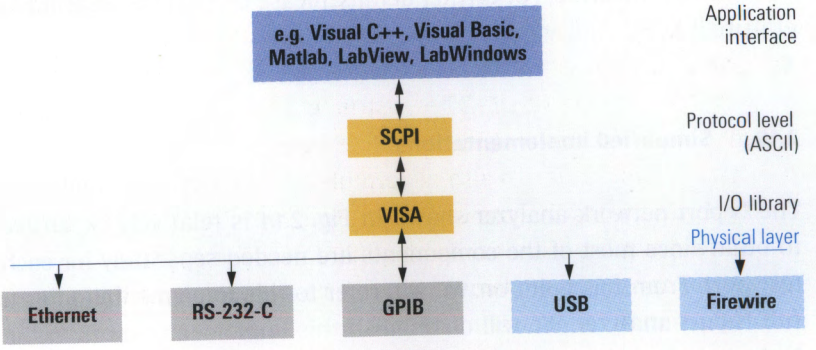


Fig. 2.9.3 The VISA I/O library and the SCPI standard.

In automation applications, LAN based interconnectivity is becoming increasingly common. Some of the benefits include modular design, good flexibility, low cost and high transmission speeds. The **LXI standard** (LAN eXtensions for instrumentation) is the successor of GPIB. It defines a standardized LAN and web browser interface for controlling test instruments. It also includes the ability to automatically detect LXI instruments in a LAN using the discovery protocol described in VXI-11. There is no need for a user interface on the instruments. The instrument can be configured by a web browser installed on the control computer. There are three compatibility classes: A, B and C. Except for class C, the instruments use the precision timing protocol (PTP) described in the standard IEEE 1588 for synchronizing a clock within the instrument via a LAN. Based on this time reference, timer events can be initiated in a synchronized manner and measured values can be provided with timestamps. Measured values can then be processed asynchronously or offline using the timestamps. In a 100Base-T LAN, it is possible to achieve accuracies of better than 10 ns for the synchronization of the clocks. However, the remaining timing offset is subject to random fluctuations and is not constant. The signal delay on the cables does not have a significant influence on the accuracy since it is compensated by the precision timing protocol. However, this is possible only if the architecture of the LAN involves unique connection paths for all participants in the communication. This is not guaranteed in case of communication via the Internet, for example. Test instruments in class A are equipped additionally with an eight-channel LVDS hardware trigger interface (LVDS = Low Voltage Differential Signaling). With triggering via this interface, the timing offset is determined by the cable length, so that it is nearly constant. With a trigger cable having the specified maximum length of 20 m, a constant timing offset of significantly greater than 10 ns is involved. For further details, please consult the relevant literature [LX05].

2.10 Simplified implementations

The N-port network analyzer shown in Fig. 2.1.1 is relatively expensive to build since most of the components are needed separately for each test port. From this point on, we will refer to this implementation as a **full N-port analyzer**. We will distinguish this implementation from the **N+1 receiver analyzer** and the **analyzer with an N-port switching matrix**.

Both of these latter examples have a reduced number of functional modules. Significant cost savings are realized particularly for $N > 3$ test ports. However, we must consider the deterioration of the specifications.

2.10.1 N+1 receiver analyzer

This receiver design involves a common reference receiver and N measurement receivers. As a general rule, we can reduce the number of receivers in an N-port analyzer to $N + 1$ with this configuration. This reduction is possible if the reference channel of the active test port is obtained in a central manner prior to the source switch regardless of the measurement direction (see Fig. 2.10.1).

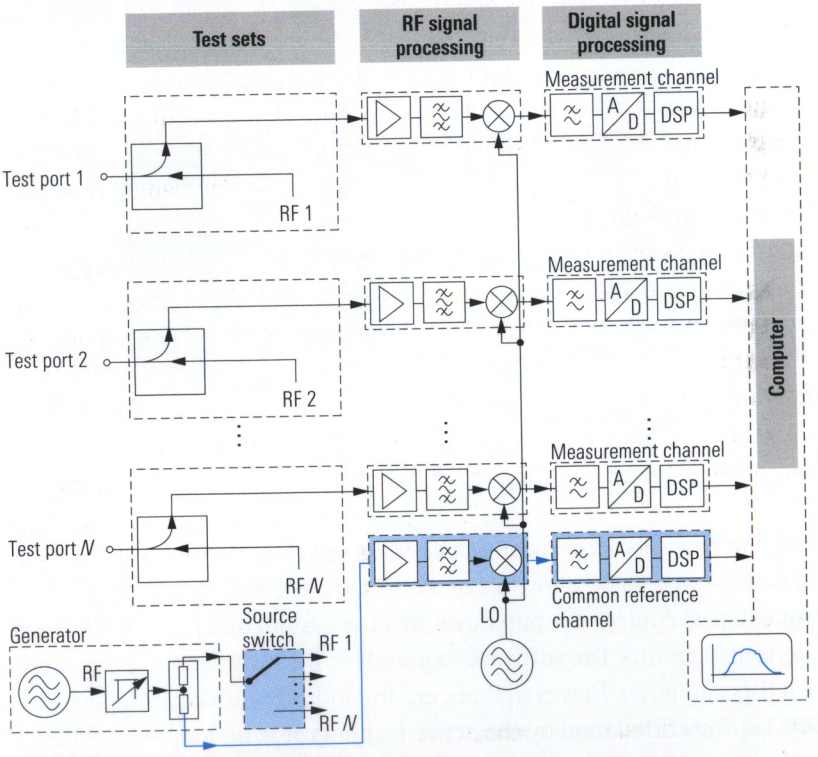


Fig. 2.10.1 N+1 receiver analyzer.

However, it is no longer possible to monitor the crosstalk of the source switch during the measurement since no information is available about the wave quantity a_i at the passive test ports. Accordingly, only certain calibration techniques can be used. They are presented in sections

3.4.1, 3.4.3 and 3.4.6 as part of the discussion of system error correction as OSM and TOSM techniques as well as incomplete techniques. The source switch is then part of the error model. Its switching behavior must be reproducible with great accuracy to ensure the error model and thus the calibration to be valid.

2.10.2 Network analyzer with an N-port switching matrix

Using an **N-port switching matrix** (shown in blue in Fig. 2.10.2), we can basically convert any two-port network analyzer to an N-port network analyzer. We do not have to modify the instrument itself.

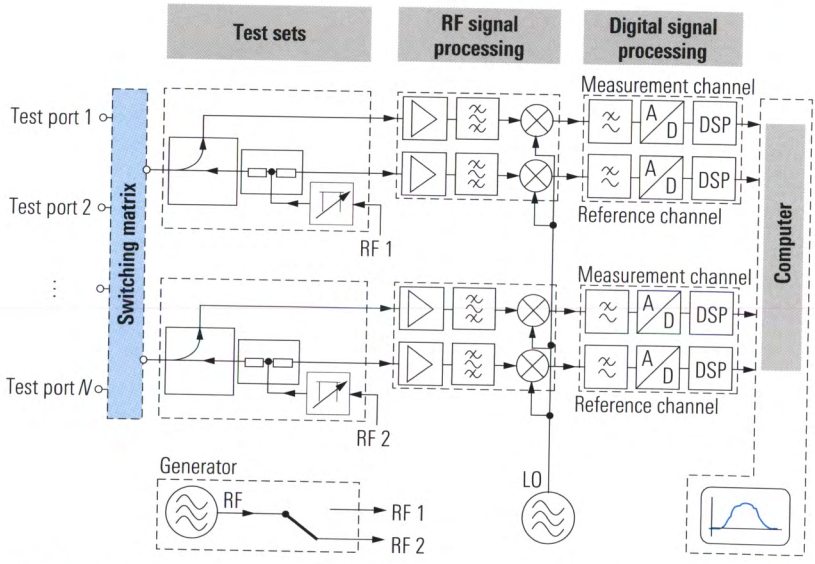


Fig. 2.10.2 A network analyzer with an N-port switching matrix.

An external control computer oversees the system and presents the measurement results. The software required for the switching matrices runs on this computer. However, there are the following disadvantages:

- During stimulation by the active test port, it is no longer possible to simultaneously measure all of the b waves. Thus in difference to Fig. 2.10.1 and Fig. 2.1.1, they must be measured sequentially. A two-port network analyzer is always limited to measuring only a single pair of ports. If we take in to account that double measurements are not necessary we count $\frac{1}{2}(N-1)N$ sweeps for a N-port DUT. The increased sweep times that this involves can be seen

in Fig. 2.10.3. Due to the reduced dynamic range obtained with a matrix, a smaller IF bandwidth will normally be selected by the user. This in turn results in a further increase in the sweep time.

- During a normal N-port measurement, a large number of switching procedures required in the N-port matrix. Mechanical switches cannot be used due to their mechanical wear and slow switching speed. Electronic switches usually have a return loss of only approx. 10 dB to 15 dB. This means they exhibit significantly worse matching than the test port of a good network analyzer. The typical insertion loss of a switch is 2 dB to 6 dB at 8 GHz. Since normally several switches are cascaded to obtain the necessary number of ports and meet the isolation requirements, switching matrices have an insertion loss of 6 dB to 16 dB at 8 GHz. The influence of this attenuation is simular to having attenuators between the test ports and the DUT (see section 2.4.1). In particular, the insertion loss causes a drastic deterioration in the directivity.
- The signals from the different ports of the DUT can be bypassed due to crosstalk inside the switching matrix. The isolation characteristics of the switches are generally worse than the dynamic range of a good network analyzer.

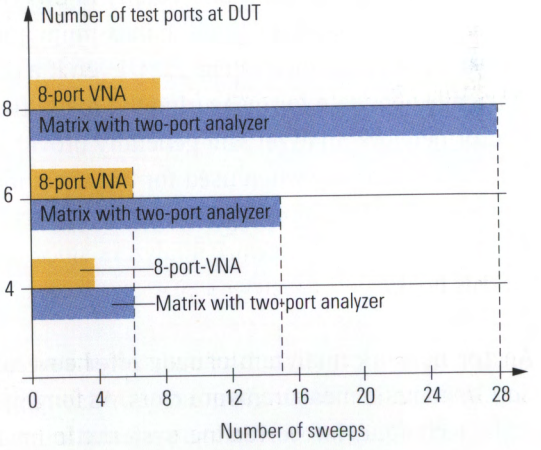


Fig. 2.10.3
Speed advantage using a full 8-port analyzer.

3 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 book.

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.

3.1 Reduction of random measurement errors

3.1.1 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 is 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). This also applies to the calibration and verification standards described in section 3.3.

3.1.2 Repeatability

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

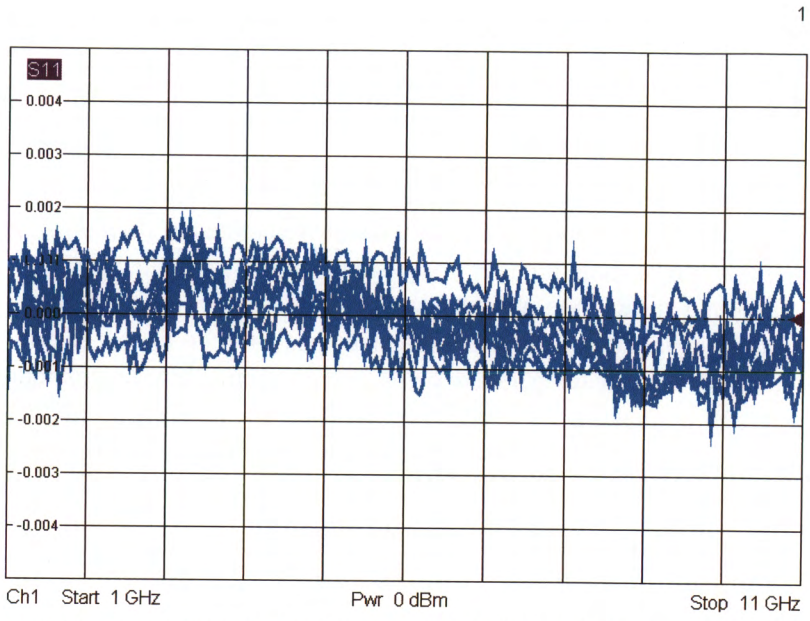


Fig. 3.1.1 Repeatability for a PC3.5 connection during a reflection measurement.

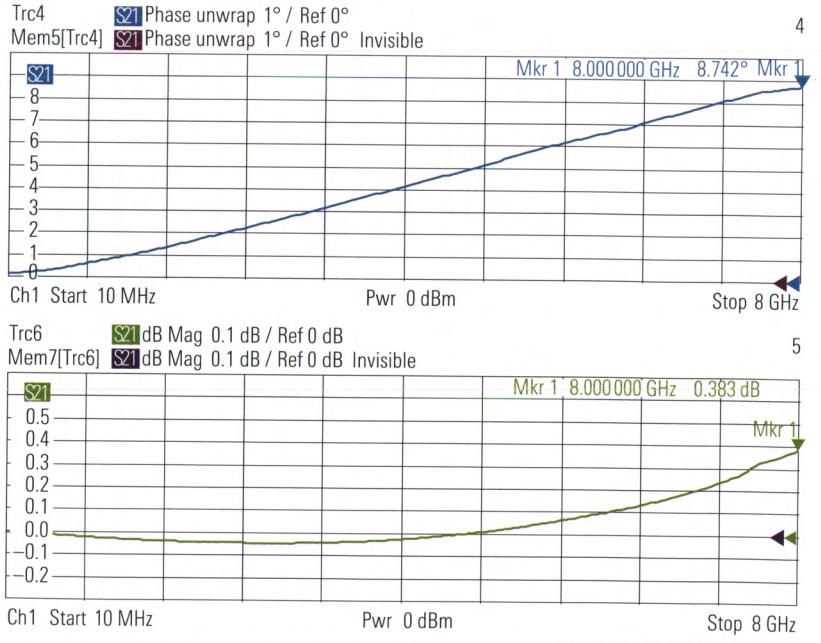


Fig. 3.1.2a Deviation due to bending on two RG400 cables with a length of 1 m each.

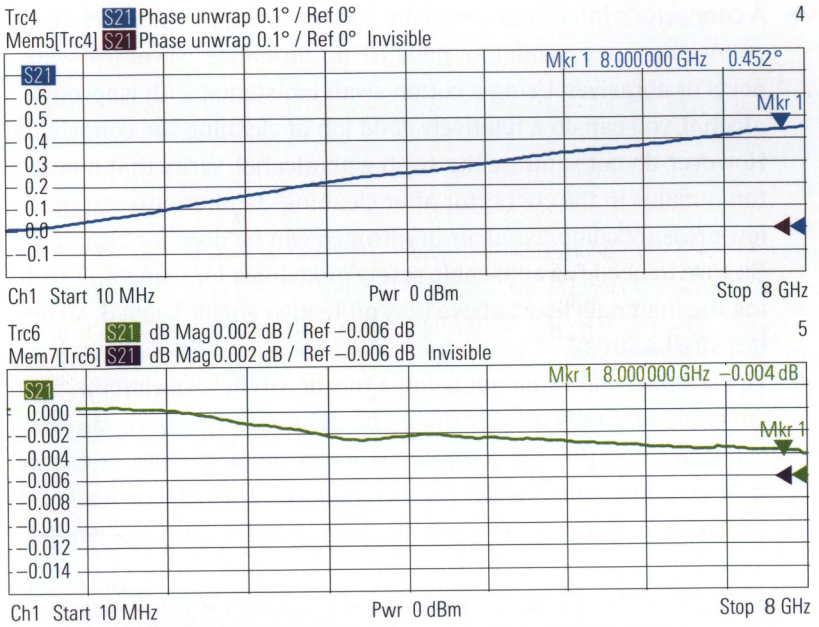


Fig. 3.1.2b *Deviation due to bending on a high-performance test port cable set with a length of 2 × 1 m.*

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. 3.1.2b 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** (see Fig. 3.3.8 ⑩), 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.

3.1.3 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 $4 \cdot 10^{-21}$ 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 (Fig. 7.1.9) 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:

$$L_N = -174\,\text{dBm} + NF + 10\lg(S_F)\,\text{dB} + 10\lg\left(\frac{B_{IF}}{\text{Hz}}\right)\text{dB} \qquad \textit{(3.1-1)}$$

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.

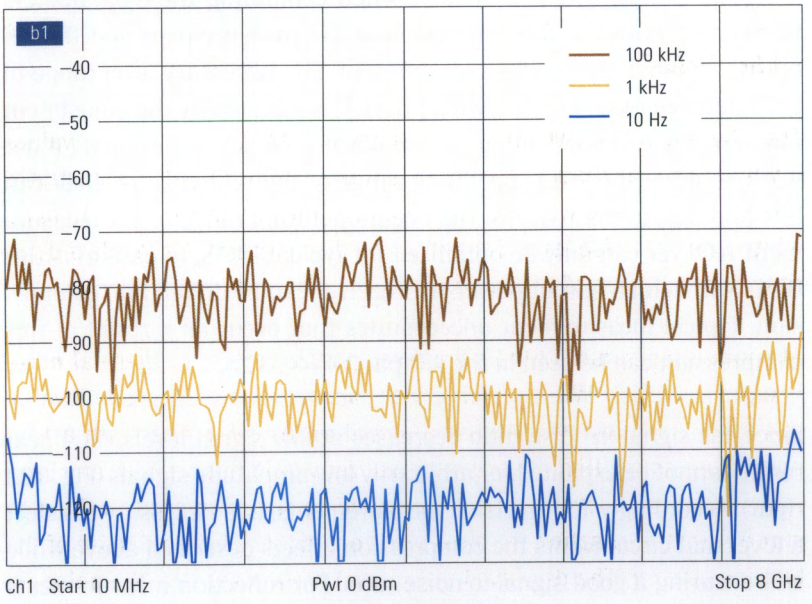


Fig. 3.1.3 Noise level for different IF bandwidths.

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.

3.2 Correction of systematic measurement errors

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

3.2.1 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 right). Accordingly, for precision measurements, it is necessary to select a level that 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.

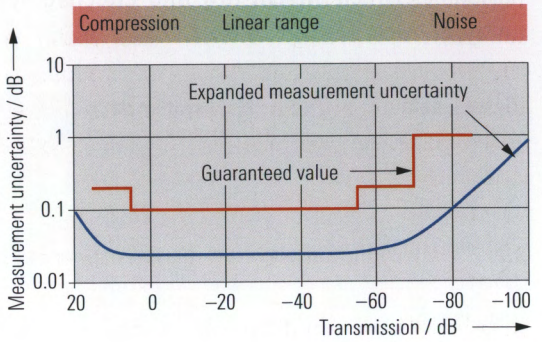


Fig. 3.2.1
Bath-tub curve

3.2.2 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.

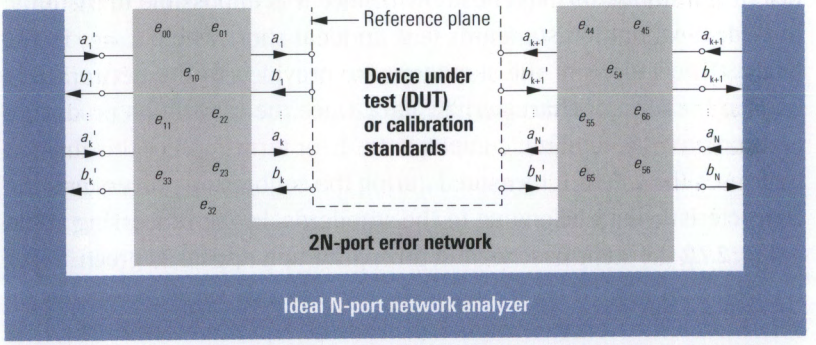


Fig. 3.2.2 Separation between an error network and an ideal network analyzer.

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

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

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

error correction is limited by random measurement errors caused by temperature drift, noise and so on. The following table 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** e_{ik}. 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 $\Gamma = -1$), 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 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).

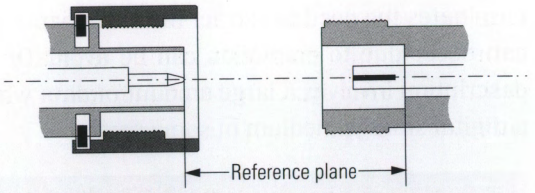


Fig. 3.2.3
Location of the reference plane in the N-type connector.

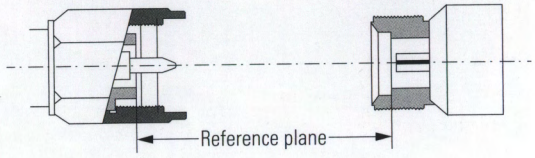


Fig. 3.2.4
Location of the reference plane in the connector types PC3.5, 2.4 mm and 1.85 mm.

3.3 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[1] 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 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

1) Be sure not to confuse the National Metrology Institutes (NMI) with meteorological organizations such as the National Weather Service in the United States.

eliminates the need to extract the coefficients. A loss of accuracy which can occur due to extraction can be avoid. 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 3.5 mm Calibration Kit R&S®ZV-Z32 with diskette.

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.



Fig. 3.3.2
Calibration Unit R&S®ZV-Z51 with four ports.

3.3.1 Coaxial calibration standards

Short (S)

Using a coaxial design, it is possible to build a short with nearly ideal characteristics (total reflection $|\Gamma| = 1$).

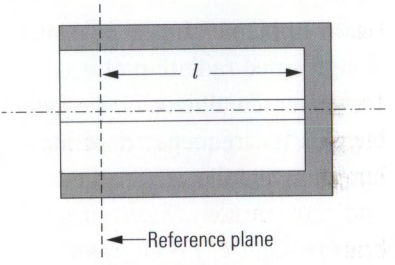


Fig. 3.3.3
Short

The reflection coefficient of the standard is thus dependent only on its length offset *l*. This offset arises since there is, for mechanical reasons, a length *l* between the reference plane and the short. The loss due to this length offset can be neglected in most cases. Accordingly, we can calculate the reflection coefficient at the reference plane as follows:

$$s_{11} = -e^{-j4\pi l/\lambda} \qquad \textit{(3.3-1)}$$

Modeling the short in the network analyzer usually requires entry of only its electrical length *l*. In exceptional cases, the model can be extended using the polynomial coefficients L_0 to L_3 for parasitic inductance. Polynomial coefficients L_0 to L_3 are based on the assumption that at frequencies over a few multiples of 10 GHz, the short is dominated

by its inductive behavior with a frequency-dependent inductance $L_e(f)$. This frequency dependence is approximated using a series expansion up to the third order ($L_3 f^3$).

$$L_e(f) = L_0 + L_1 f + L_2 f^2 + L_3 f^3 \qquad \textit{(3.3-2)}$$

Formula (3.3.1) is then replaced by the following equation:

$$s_{11} = \frac{j2\pi f L_e(f) - Z_0}{j2\pi f L_e(f) + Z_0} e^{-j4\pi l/\lambda} \qquad \textit{(3.3-3)}$$

Typical technical specifications for a short are as follows:

Offset length	Phase uncertainty		Polynomial coefficients for the inductance			
	0 to 8 GHz	8 to 26.5 GHz	L_0	L_1	L_2	L_3
5.00 mm	≤0.5°	≤2.0°	0 pH	0 pH/GHz	0 pH/GHz²	0 pH/GHz³

Table 3.3.1 Typical technical specifications for a short.

Open (O)

When implementing a coaxial open standard, we must use a closed design since radiation effects can occur otherwise. At the open end of the inner conductor, a frequency-dependent **fringing capacitance** is formed. Its frequency dependence can be approximated using a cubic equation as follows:

$$C_e(f) = C_0 + C_1 f + C_2 f^2 + C_3 f^3 \qquad \textit{(3.3-4)}$$

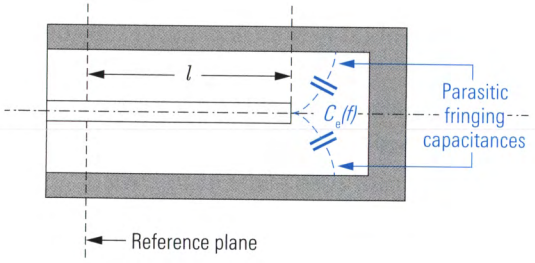


Fig. 3.3.4 Open

The reflection coefficient for the open at the reference plane is then calculated as follows:

$$s_{11} = \frac{1 - j2\pi f Z_0 C_e(f)}{1 + j2\pi f Z_0 C_e(f)} e^{-j4\pi l/\lambda} \qquad \textit{(3.3-5)}$$

Even if we were able to manufacture an open standard with a length $l = 0$, the fringing capacitances would result in a negative imaginary part for s_{11} at higher frequencies. Typical technical specifications are as follows:

Offset length	Phase uncertainty		Polynomial coefficients for the fringing capacitance			
	0 to 8 GHz	8 to 26.5 GHz	C_0	C_1	C_2	C_3
5.00 mm	≤0.5°	≤2.5°	13.6348 fF	−0.2164 fF/GHz	0.0189 fF/GHz^2	−0.00028 fF/GHz^3

Table 3.3.2 Typical specifications for an open.

Match (M) (or load)

A match is a precision broadband impedance with a value that corresponds to the system impedance. Fig. 3.3.5 shows an implementation in which the inner conductor terminates into a resistively coated substrate. Using trimming holes inserted with a laser, for example, it is possible to optimize the impedance value so that return losses of, say, 45 dB are attainable (at frequencies up to 4 GHz).

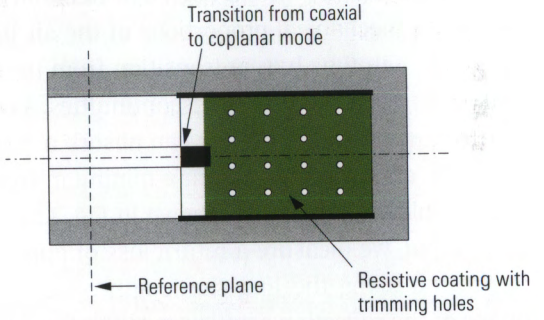


Fig. 3.3.5
Terminating impedance

In the past, it was common to assume an ideal match ($\Gamma = 0$) in the calibration process. Accordingly, it was not modeled in greater detail. More recently, network analyzers also allow to include the non-ideal properties of a match.

DC resistance	Return loss		Max. power
	0 to 4 GHz	4 to 26.5 GHz	
50 Ω ± 0.5 Ω	≥40 dB	≥30 dB	0.5 W (27 dBm)

Table 3.3.3 Typical specifications for a match.

Sliding Match (or sliding load, sliding termination)

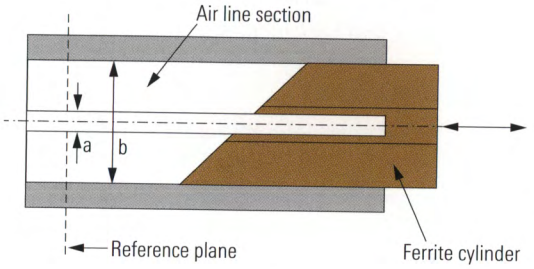


Fig. 3.3.6
Schematic design of a sliding match.

An air line with a specified characteristic impedance can be manufactured significantly more accurately than the match shown in Fig. 3.3.5. The characteristic impedance Z_C of the air line is calculated from the diameter a of the inner conductor and the inner diameter b of the outer conductor as follows:[1)]

$$Z_C = \frac{60\,\Omega}{\sqrt{\varepsilon_r}} \ln\left(\frac{b}{a}\right) \approx 60\,\Omega \cdot \ln\left(\frac{b}{a}\right) \qquad \textit{(3.3-6)}$$

Since formula (3.3-6) contains the ratio of values a and b and not their absolute values, we have a degree of freedom that makes it possible to match the mechanical proportions of the air line to the connector type in use. We can thus have a transition from the connector to the air line that is as free as possible of discontinuities. A cylindrical ferrite rod that is introduced into the air line can absorb a large portion of the electromagnetic energy starting from a minimum frequency of approx. 2 GHz. The complete arrangement shown in Fig. 3.3.6 is called a sliding match. At its input, we measure a return loss of approx. 20 dB. If we move the ferrite rod along the line, the length offset changes and so does the phase of the reflection coefficient at the line input.

At a constant frequency, the reflection coefficient of the sliding match lies at a different point on the circle shown in green in Fig. 3.3.7 depending on the length offset. Three points on the circle are sufficient to construct the center point Γ_C. This point corresponds to the characteristic impedance of the air line section as measured by the network analyzer. This quantity is used instead of the impedance from Fig. 3.3.5 for calibration purpose. The circle radius Γ_r is determined by the return loss of

1) As an approximation, a relative permittivity $\varepsilon_r = 1$ is assumed for air. The exact value at 101.325 kPa pressure and 23°C is $\varepsilon_r = 1.000649$.

the ferrite rod. It is shown in Fig. 3.3.7. The value Γ_r has been choosen larger than normally for illustration purposes.

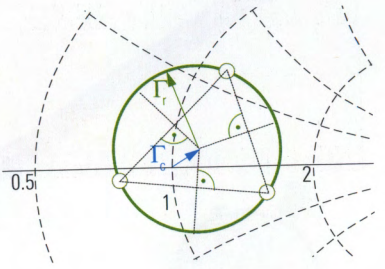


Fig. 3.3.7
Enlarged excerpt from the Smith chart for determining the characteristic impedance.

In practice, a frequency sweep is usually performed instead of a measurement at a fixed frequency. In this case, it no longer suffices to just measure three positions of the ferrite rod. If we increase the frequency in Fig. 3.3.7, the three points move along the circumference of the circle. However, the frequency dependence of the points is not identical, meaning they move at different speeds. At certain frequencies, two or three points will overtake one another and merge into a single point. In constructing the center point, however, at least three separate points that are distributed over the circumference of the circle are needed. To ensure this in the frequency range specified for the sliding match, a total of six positions are selected on the sliding match. It is not necessary to approach a certain position in the Smith chart entirely accurately. Instead, it is sufficient if we orient ourselves during the calibration procedure towards the marks on the sliding match (see Fig. 3.3.8 ❶ to ❻). Using the pin and the fastening nut ❼ and ❽, we can adjust the offset of the inner conductor with respect to the reference plane for the connector ❾. Our adjustment device is a pin depth gage ❿. The zero point on the pin depth gage is generally aligned prior to using the device with a suitable gage master. If the inner conductor of the sliding match protrudes too far the connectors can be damaged, when connecting the sliding match and the test port. On the other hand, if the inner conductor is seated too low, this will produce an unnecessary air gap. The influence of this sort of cylindrical constriction of the inner conductor is described in [MS66], for example. For an N-type connector at 18 GHz and an air gap width of 0.15 mm, [BS98] specifies a parasitic reflection coefficient of 0.023 ($\triangleq$ 33 dB). For an air gap that is reduced to 0.067 mm, we calculate a reflection of 40 dB. The parasitic reflection influences the Γ_C value in Fig. 3.3.7 and thus the accuracy of the calibration. Accordingly, we should seek to minimize the air gap.



Fig. 3.3.8 *Sliding match and pin depth gage.*

We cannot use the sliding match at frequencies below 2 GHz. However, almost all known calibration techniques can be performed using a combination of a sliding match and a fixed match to circumvent a frequency range restriction at low frequencies. Particularly at frequencies over 8 GHz, the sliding match is significantly more accurate than a match. The TOM-X technique described in section 3.4.4 cannot be performed using a sliding match.

Through (T)

A through is a two-port standard that allows direct connection of two test ports with low loss. If we are using connectors of the same type but with a different gender (or sexless connectors of the same type), we can directly interconnect the two test ports to produce a through connection. By definition, this special case of a through has an electrical insertion length of 0 mm. If we are using connectors of the same type and same gender, a through will consist of a small line section making a suitable electrical connection between the two test ports. The ideal case is represented by a small air line without an inner conductor support. Due to its greater ease of use, a line with a fixed inner conductor (inner conductor support in Fig. 3.3.9) is generally preferred.

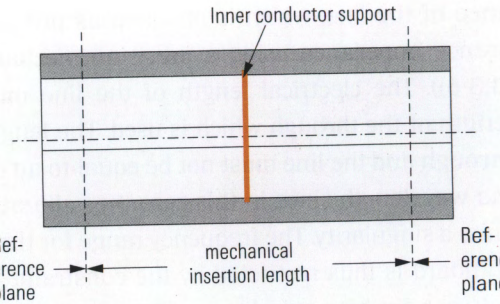


Fig. 3.3.9 Through

The characteristic impedance for this connection should match the reference impedance as closely as possible. The characteristic quantities of a through are the insertion loss and the electrical insertion length. In most common calibration techniques the through is assumed to be ideally matched.

Electrical insertion length	Return loss		Insertion loss	
	0 to 4 GHz	4 to 26.5 GHz	0 to 4 GHz	4 to 26.5 GHz
23.20 mm ± 0.12 mm	≥32 dB	≥26 dB	≤0.08 dB	≤0.18 dB

Table 3.3.4 *Typical specifications for a through.*

The following calibration standards are reserved for the techniques presented in section 3.4.2 as "7-term" techniques. All these standards have only partially known S-parameters.

Reflect standard (R)
This one-port standard exhibits a reflection coefficient $|\Gamma| > 0$. The exact reflection value is not needed for the TRL technique described in section 3.4.2. However, it must be identical at both test ports. The reflect standard is used in the so-called **self-calibration techniques**. In these techniques, the known information for the standards is used to determine the unknown properties of the standards. To pin down the sign, the phase of the reflect standard must be known to $\pm 90°$. To characterize the standard its behavior at low frequencies is said to be either **more capacitive** or **more inductive**. If the length offset of the standard causes the phase to lie outside the interval $0°$ to $-90°$ or $0°$ to $+90°$, the length offset must also be known approximately.

Line standard (L)
The line standard is a two-port standard. In coaxial systems, it is implemented as an **air line**. The critical quantity is the characteristic imped-

ance of the line, which is matched as precisely as possible to the reference impedance using a precision mechanical design (see formula (3.3-6)). The electrical length of the line must be different than the length of the through which is used. The length difference between the through and the line must not be equal to an integral multiple of half of the wavelength since in this case the calibration techniques would produce a singularity. The frequency range for the calibration using the line standard is thus restricted by the constraint that the phase difference between the transmission coefficients of the through and line must be sufficiently different from 0° and 180° (e.g. it must be between 20° and 160° or the ratio of the stop and start frequencies may have a maximum ratio of 8:1). To extend this frequency range, we can measure two lines with different lengths. In addition, the lower frequency limit that is dictated by the longer of the two lines can be extended down to 0 Hz by measuring the fixed match. Since air lines do not have supports, a vertical position is necessary during usage at least for longer air lines. Otherwise, the weight would displace the inner conductor from its central position.

Symmetrical network (N)

The symmetrical network standard is a two-port standard which exhibits symmetrical reflection coefficients $s_{11} = s_{22}$ for its two ports. This reflection coefficients may have any value except for $s_{11} = s_{22} = 0$. Like for the reflection standard, it is necessary to know whether the reflection coefficients are more like an open or a short. The transmission coefficient s_{21}, s_{12} of the standard is irrelevant and can take on any value; even values of 0 and 1 are permitted. To a certain extent, this standard represents a generalization of the reflection standard.

Attenuator standard (A)

The attenuator standard is a two-port standard which must be well matched on both sides. Unlike the through standard, it should have an insertion loss which does not have to be known exactly. The attenuation difference compared to the through should be at least 10 dB. The insertion loss of the attenuator standard should not exceed 55 dB.

Unknown through (U)

Any two-port network whose S-parameters fulfill the reciprocity condition $s_{21} = s_{12}$ can be used as an unknown through for the UOSM technique described in section 3.4.2. To pin down the sign in the calcula-

tion, the phase shift produced by the unknown through must be known approximately. If a two-port DUT fulfills the reciprocity condition, it can also serve as an unknown through in accordance with [FP92].

3.3.2 Waveguide calibration standards

For calibration of waveguide measurements, basically the same type of standards are used than within a coaxial geometry. Various cross-sectional geometries are possible for waveguides. The most common involve rectangular cross-sections with the ratio of the sides equal to 2 : 1. The standardized dimensions can be found in Table 8.5.1.

Waveguides can transport a number of different propagation modes. For a mode to be capable of propagation, the stimulus frequency f must be greater than the critical frequency of the relevant mode. As the stimulus frequency increases, the number of modes capable of propagation increases. Assuming we have a rectangular waveguide with inner dimensions a, b where $a > b$, then the H_{10} mode is the mode with the lowest critical frequency. This frequency is also known as the cutoff frequency f_c of the waveguide. It is associated with the cutoff wavelength $\lambda_c = c_0 / f_c$ where $c_0 = 2.99792458 \cdot 10^8$ m/s (speed of light in vacuum).

$$\lambda_c = 2a\sqrt{\varepsilon_r \mu_r} \qquad \textit{(3.3-7)}$$

The volume (assumed to be homogenous) enclosed by the waveguide is characterized by its relative permittivity ε_r (for air, $\varepsilon_r \approx 1$) and relative permeability μ_r (typically, $\mu_r = 1$).

For a ratio of the sides equal to 2 : 1, further modes (H_{20}, H_{01} mode) can propagate starting at twice the cutoff frequency f_c. To avoid mode interference (i.e. operation with several modes propagating simultaneously), in practice we use exclusively the H_{10} mode in the following frequency range:

$$f = 1.2 \cdot f_c \text{ to } 1.9 \cdot f_c \qquad \textit{(3.3-8)}$$

Waveguides are dispersive by nature, i.e. the waveguide wavelength $\lambda_W(f)$ is frequency-dependent. For the H_{10} mode, the waveguide wavelength is calculated as follows:

$$\lambda_W(f) = \frac{1}{\sqrt{\varepsilon_r \mu_r}} \frac{\lambda_0}{\sqrt{1-\left(\frac{\lambda_0}{\lambda_c}\right)^2}} \qquad \textit{(3.3-9)}$$

where the free-space wavelength is:

$$\lambda_0 = c_0/f \qquad \textit{(3.3-10)}$$

To give an impression of typical waveguide wavelength corresponding to the frequencies given by formula (3.3-8) we can assume $\varepsilon_r = 1$, $\mu_r = 1$ and use equation (3.3-9) and (3.3-7) to calculate λ_W by the inner dimension a of the waveguide as $\lambda_W(1.2 \cdot f_c) = 3.016 \cdot a$ to $\lambda_W(1.9 \cdot f_c) = 1.238 \cdot a$.

Short (S)

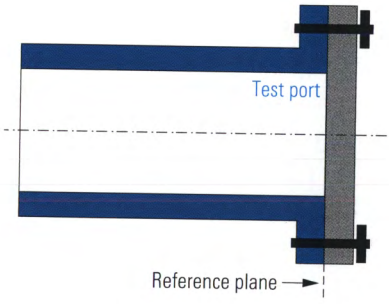


Fig. 3.3.10
Waveguide short

In a waveguide system, a short can be realized using a conductive plate mounted on the waveguide flange of the test port. The short thus has a length offset $l = 0$. The uncertainty in determining the length l disappears.

Offset short

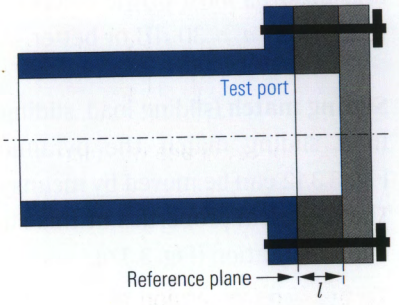


Fig. 3.3.11
Waveguide offset short.

A waveguide that is open on the end is not usable as an open. Due to radiation effects at the open end of a waveguide, only a small amount of energy is reflected. This results in a typical return loss of 6 dB to 25 dB. In other words a waveguide that is open on the end behaves much more like a bad match and cannot be used as an open standard. Another realization for the open standard has to be chosen. Since waveguides are operated exclusively in a narrowband fashion (see formula (3.3-8)), we can use a so-called offset short (see Fig. 3.3.11). The length offset l should be selected so that the short is transformed into an open at the center frequency of the waveguide, i.e. at 1.55 f_c.

$$l = \frac{1}{4}\lambda_W(f) \quad \text{where} \quad f = 1.55 f_c \qquad \textit{(3.3-11)}$$

Using equations (3.3-7), (3.3-9), (3.3-11) and assuming $\varepsilon_r = 1$, $\mu_r = 1$ we calculate the necessary length offset l as a function of the inner waveguide dimension a as $l = 0.422 \cdot a$.

Match (M) (or load)

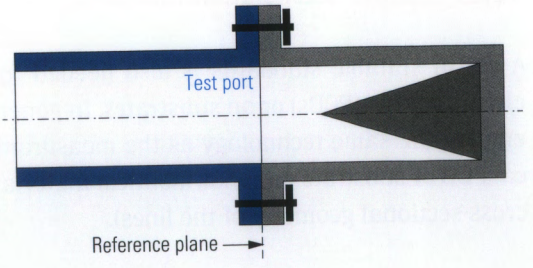


Fig. 3.3.12
Waveguide match

A pyramid- or cone-shaped structure made of ferrite material is fixed in a waveguide piece that is closed at the end. At typical waveguide

frequencies $f > 4$ GHz, the material acts as an attenuating terminator and absorbs most of the electromagnetic energy, helping to achieve a return loss $a_r = 30$ dB or better.

Sliding match (sliding load, sliding termination)
In a sliding match, the pyramid-shaped ferrite material shown in Fig. 3.3.12 can be moved by means of a sliding element in the waveguide. The electrical properties of this standard are comparable to its coaxial implementation (Fig. 3.3.6).

Through (T)
The through consists of either a piece of waveguide or, if the mechanical conditions allow it, a direct connection between the test ports. A waveguide through is characterized in the same manner as a coaxial through.

Adapters between coaxial and waveguide systems
Most network analyzers have coaxial connectors, necessitating an adapter for transmission from coaxial connectors to the waveguide system and vice versa. Capacitive or inductive coupling (or a combination of the two) is used to couple the signal into and out of the waveguide. A variety of those adapters is commercially available. Alternatively an example design of a coaxial to rectangular waveguide adapter that is scaled to meet different waveguide dimensions can be found e.g. in [Ku01]. With waveguides working at much higher frequencies (normally higher than the network analyzer's maximum frequency) another method of interconnection is used. It is described in section 8.5 of this book.

3.3.3 Microstrip calibration standards

A set of suitable standards is also needed for measuring on **printed circuit boards** (PCBs) or on **substrates**. In general, these standards must employ the same technology as the measurement substrate (i.e. identical PCB or substrate material, identical material thickness and the same cross-sectional geometry of the lines).

The following figure shows a typical implementation on a PCB[1]. For all of the standards, a minimum length is required so that the higher modes that arise at the transition from the coaxial system to the PCB can fade away. In the setup used here, a length $l = 20$ mm to 50 mm is sufficient.

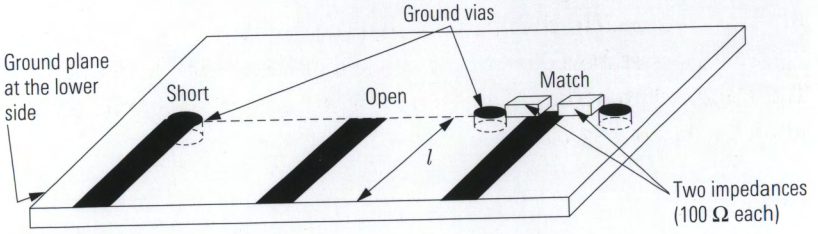


Fig. 3.3.13 One-port standards in microstrip technology.

If all of the standards are implemented with the same length offset l and the through connection has the length $2l$, the length offset can be specified as being equal to 0 in the characteristic data. The line section with their length l are then treated during the calibration process as if they belonged to the test port cable. To locate the reference plane exactly at the component edge, the measurement substrate must have a length l between each contact point and the DUT's pads (see Fig. 3.3.14).

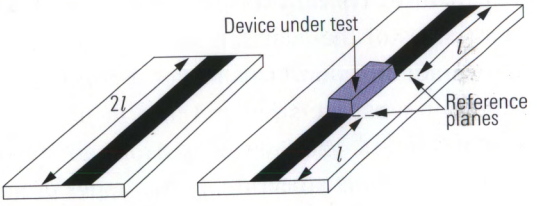


Fig. 3.3.14 Through standard and measurement substrate.

The contact between the calibration standards and the DUTs involves, e.g. fixed coaxial adapters such a SMA solder-in connectors. A **test fixture** is generally preferred for substrates.

1) In this case, DiClad 527 from the company Arlon was used as the raw material for the printed circuit board. It has a thickness $H = 0.38$ mm and a relative permittivity $\varepsilon_r = 2.54$. The copper layer was increased on both sides to 35 µm. For microstrip lines having a characteristic impedance of 50 Ω, we can calculate a width of 1 mm. This line width is suitable particularly for the component size SMD 0603 used in the example. For manufacturing low-inductance ground vias, a hole diameter of 1 mm has proven practical under the specified conditions. The setup was used up to a frequency of 8 GHz.

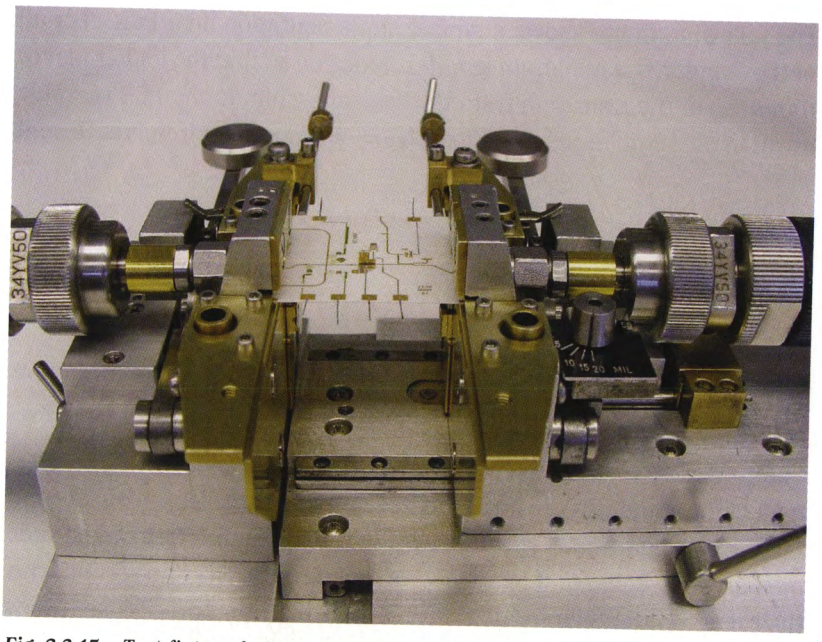


Fig. 3.3.15 *Test fixture for RF substrates (Rohde&Schwarz).*

Handling of a test fixture

Test fixtures typically have two clamping jaws that are placed on the front sides of the substrate. One of the jaws can be adjusted in both horizontal directions. This makes it possible to adapt the test fixture to different substrate sizes and also to handle diagonally arranged contact interfaces. The jaws consist of two strip-shaped ground contacts that squeeze the substrate. In the upper ground contact, an inner conductor is opened up to adapt to a coaxial line system. The ground connection is normally made by the metallized bottom of the substrate, i.e. by the lower strip-shaped ground contact of the clamps. The implementation shown in Fig. 3.3.15 is specially configured for manual measurement of large production lots. Two mechanical stops are provided to simplify inserting the substrate. The DC probes (not shown here) are automatically raised and lowered when unlocking and locking the clamping jaws. Test fixtures must have a very rugged mechanical design. However, their fine inner conductor and the often brittle substrates necessitate careful handling. Never activate the mechanism for locking and unlocking the jaws in a jerky manner. To ensure correct measurement results, make sure that the inner conductor contacts the substrate at the proper position. During storage many test fixtures (but not the one shown here) require a suitable spacer clamped into the jaws, because otherwise the

inner conductors would be damaged. Be sure to follow the manufacturer's instructions for the test fixture that you use.

The following figure shows customer-specified calibration substrates as are needed for three-port calibration in a special test fixture. The match standard (in the top right of the figure) is realized with thin-film resistors. The substrate is based on an aluminum oxide ceramic material (Al_2O_3).

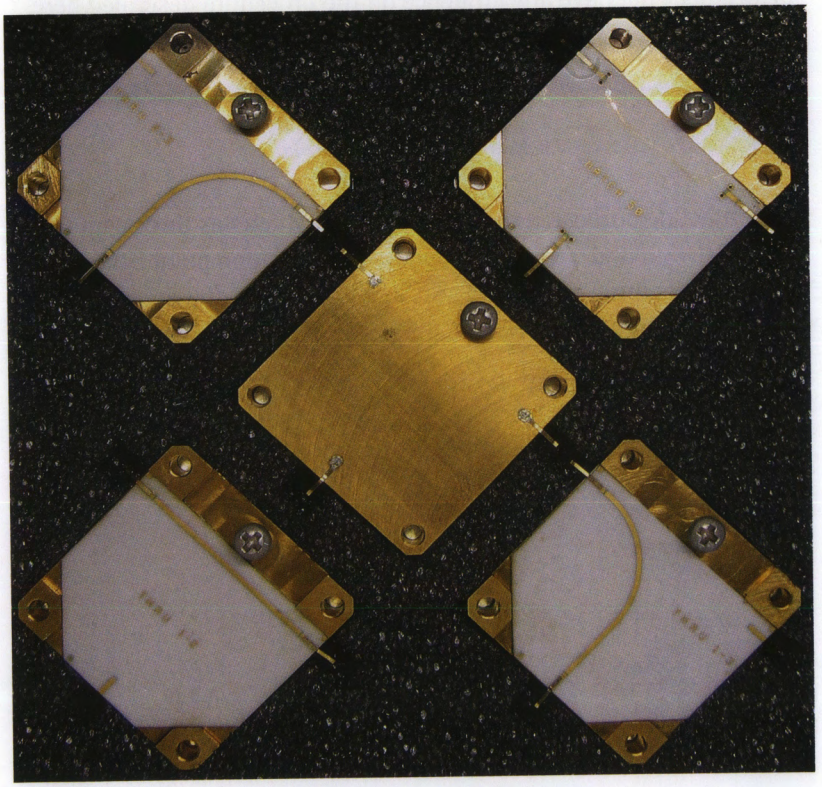


Fig. 3.3.16 Customer-specified calibration substrates.

3.3.4 Coplanar calibration standards

To obtain the best possible adaptation between the coaxial line system of the network analyzer and the substrate, a suitable ground connection must be provided at the transition point. In Figs. 3.3.13 and 3.3.14, the substrates were accessible via the front sides. This enabled a ground connection at the contact point through the clamping jaws. Particularly

in case of **on-wafer measurements** in the semiconductor industry, the circuits are accessible only via the top side of the substrate. To ensure a suitable ground connection at the contact point nonetheless, it is implemented as a **coplanar waveguide** (CPW). In this structure, all of the lines (i.e. signal and ground) are situated on the same side of the substrate. Besides this cross-section that is shown in Fig. 3.3.17, a structure with an additional ground surface on the lower side is also commonly used. This is known as a grounded coplanar waveguide (GCPW). This additional surface does not have any influence on the procedure described here.

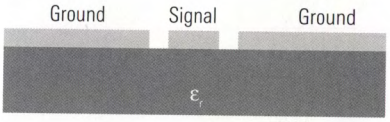


Fig. 3.3.17
Cross-section of a coplanar line.

For contacting purpose, a **GSG probe** is used which has three contact fingers (GSG = ground - signal - ground). The transition point between the coaxial mode (provided by the test port cable) and the coplanar mode (needed for contacting) is inside the **probe**. Fig. 3.3.18 shows the design of several typical calibration standards using coplanar technology.

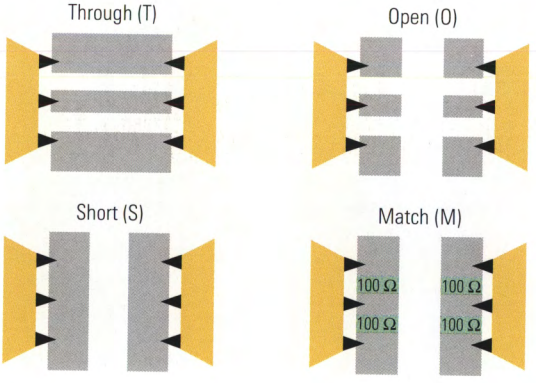


Fig. 3.3.18
Different calibration standards in coplanar technology.

Due to the very small dimensions of the contacts, a highly precise mechanical implementation is required. This device is shown in Fig. 3.3.19 and is known as a **wafer prober**.

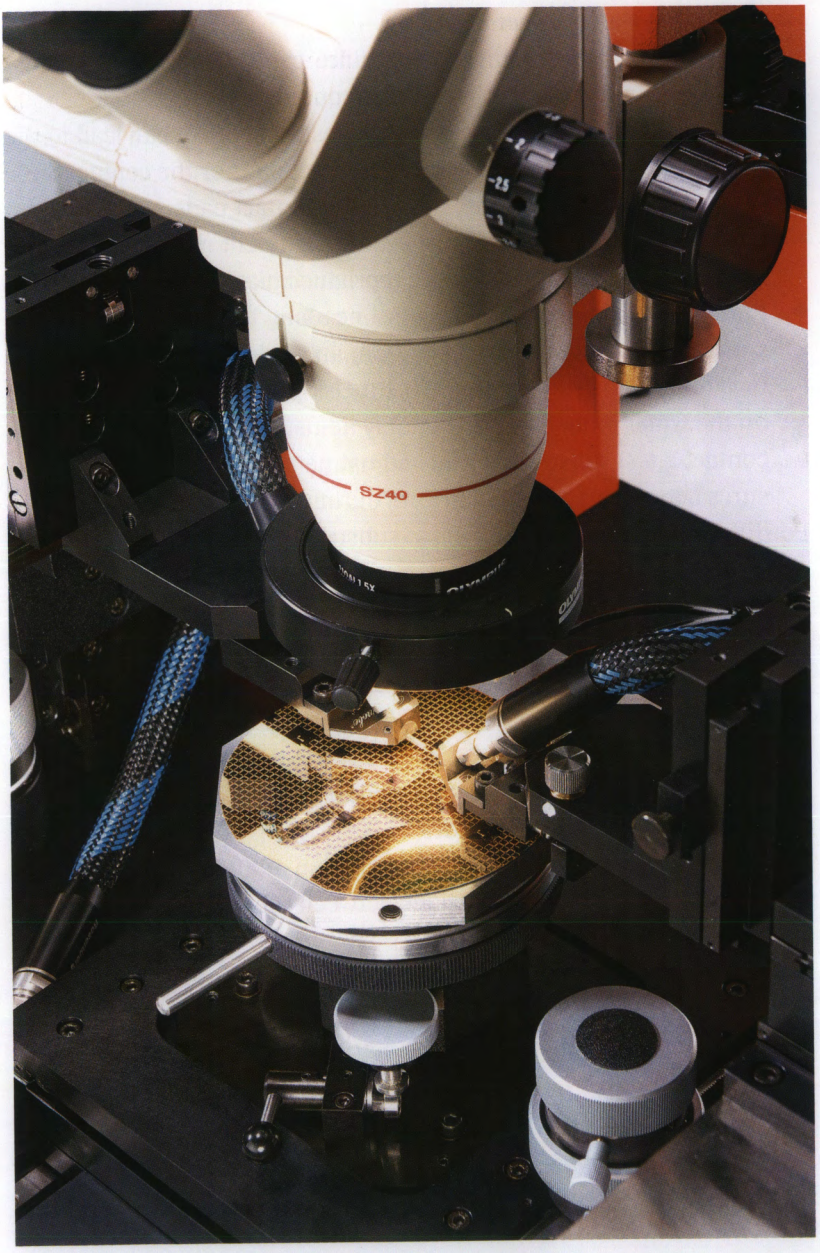


Fig. 3.3.19 *Manual wafer prober: SUSS MicroTec PM8.*

Handling of a wafer prober

The probes are very sensitive and require careful mechanical adjustment. Here, we can distinguish two main quantities:

- **Offset**: Under unfavorable circumstances, a big offset can lead to the contact fingers lying outside of the contact pads or the contact fingers causing a short circuit between the contact pads. However, even a much smaller offset can be relevant. Alignment marks help to minimize the offset and improve the reproducibility of the contact.
- **Planarity**: If we imagine a straight line along the tips of the three contact fingers, it should ideally form an angle $\varphi = 0°$ with the surface of the substrate (see Fig. 3.3.20). Using the binoculars installed on the wafer prober, we can suitably adjust the probe tips. We can contact a test substrate for fine adjustment. When different pressures act on the individual fingers during contact, they will leave different impressions on the test substrate.

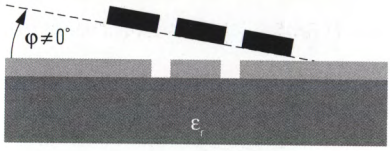


Fig. 3.3.20
Probe tips with insufficient planarity.

Prior to beginning the calibration procedure, it is necessary to set up the wafer prober. This process includes:

- Visual inspection, cleaning and installation of all of the GSG probes
- Creation of electrical connections between the GSG probes and the network analyzer
- Precision arrangement of the relative position of the GSG probes with respect to one another
- Inspection and adjustment of the planarity of all probes tips

The calibration substrate is inserted into a receptacle plate, which is known as device chuck. The first measurement position is visited. During contact (which normally involves raising the device chuck), the substrate must not be moved. To ensure a good electrical contact, the device chuck is raised slightly more than necessary for the very first mechanical contact. This additional (probe-dependent) motion of the device chuck is called overtravel. It causes the probes break through any surface oxidation and assures a good low resistance contact. When using the wafer prober, be sure to follow the manufacturer's instructions for your model as well of the instructions for the probes you are using.

Automatic wafer probers are used for series production. Once the system is configured, the different contact points are automatically visited under program control. The ear-like extensions of the device chuck serve as a support for the calibration substrates (see Fig. 3.3.21). This allows the test station to perform automatic calibration under its own control and repeat this at regular intervals. Frequent calibration can help to prevent problems associated with electrical drift.

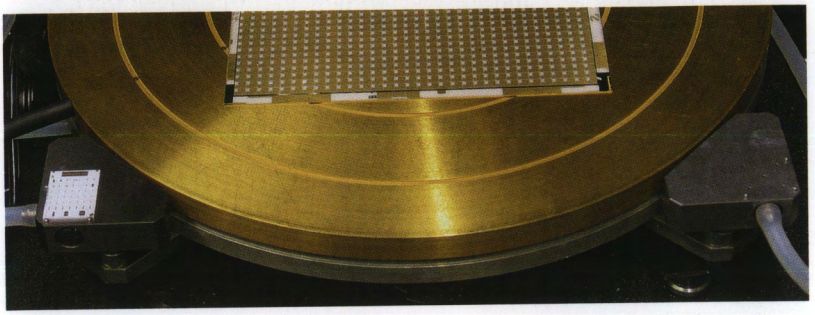


Fig. 3.3.21 Device chuck on an automatic wafer prober (SUSS MicroTec PA200).

3.3.5 The uniform model of the calibration standards

The properties of all calibration standards can be characterized using a uniform model. This allows more or less uniform handling of the coefficients for the standards on the network analyzer. The coefficients that are not relevant for the applicable standard are set to a neutral value (as shown in Table 3.3.5).

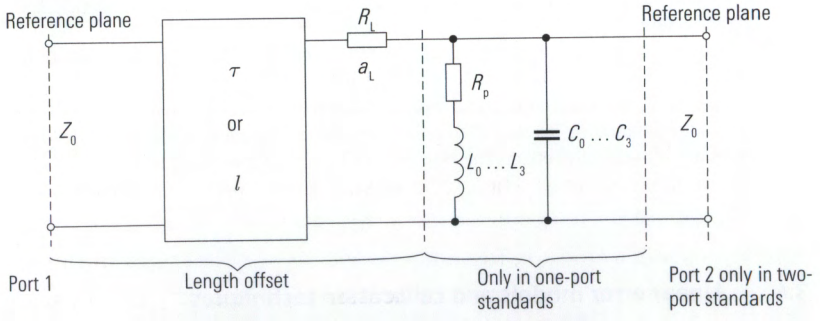


Fig. 3.3.22 The uniform model of the calibration standards.

Standard	Number of ports	Length offset or two-port standard			Single-port standard		
		Delay τ in s or electrical length l in m	Loss resistance R_L in Ω/s or[4] loss attenuation a_L in dB/$\sqrt{\text{Hz}}$	Reference impedance Z_0 in Ω	Inductance coefficients $L_0 \ldots L_3$ in H ... H/Hz^3	Capacitance coefficients $C_0 \ldots C_3$ in F ... F/Hz^3	Parallel resistance R_p in Ω
Short	1	✓	✓	50	✓ [1]	0 ... 0	0
Open	1	✓	✓	50	0 ... 0	✓	∞
Match	1	✓[2]	✓[2]	50	0 ... 0	0 ... 0	✓[2]
Sliding match	1	Not specified	0	50	0 ... 0	0 ... 0	Not specified
Through	2	✓	✓	50	Not specified	Not specified	Not specified
Reflection standard	1	✓	0	50	✓ [3]	✓ [3]	∞
Line standard	2	✓	✓	50	Not specified	Not specified	Not specified
Symmetrical network	2	✓	0	50	✓ [3]	✓ [3]	0
Attenuator standard	2	Not specified	Not specified	50	Not specified	Not specified	Not specified

Table 3.3.5 *Parameters in the uniform model of the calibration standards.*

Notes for the table:

✓ This value must be entered.

1) The inductance coefficients are often neglected (= 0).

2) If an ideal match $R_p = Z_0$ can be assumed, the delay τ and the attenuation a_L can be assumed to be equal to 0.

3) Here, L_0 or C_0 is required (meaning more capacitive or more inductive).

4) The loss resistance R_L is referred to a fixed frequency, typically 1 GHz, while the loss attenuation a_L directly reflects the skin effect which is proportional to $\sqrt{f}$.

3.4 Linear error models and calibration techniques

During the past four decades a total number of more than one hundred reports and articles on the topic of vector network analyzer calibration techniques have been published. Because there is no uniform notation among this material, it is difficult to group and to compare all the pro-

vided possibilities. The main aim of this section is to give a systematic introduction using four different error models to group the most important calibration techniques and to compare these techniques. This section can help readers to find the calibration technique that is the best for their individual requirements.

The error models and calibration techniques will be illustrated using a two-port network analyzer. The more general case with N ports is considered subsequently. Practical examples on the topic can be found in section 4.1 and 4.2.

3.4.1 3-term error model (OSM technique)

Section 2.2.6 presented an error model for a one-port network analyzer. This is shown in Fig. 3.4.1 below. All of the systematic measurement errors are modeled using an error two-port. This makes it possible to combine the generator and reference receiver into an ideal wave source ($b_1 = 0$). Together with the ideal measurement receiver ($a_3 = 0$), we obtain a one-port device that can be termed an ideal network analyzer.

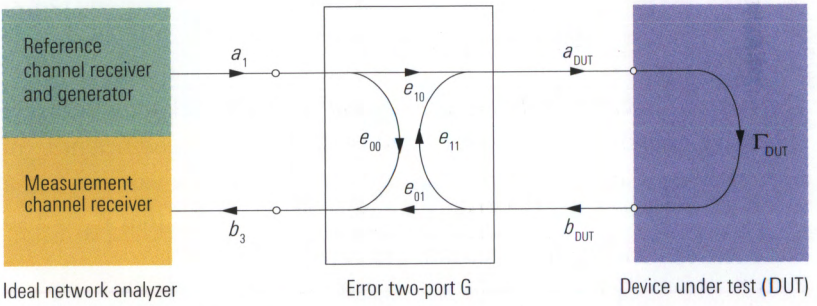


Fig. 3.4.1 Error two-port with e terms.

In the **OSM technique**[1], the following standards are measured to determine the model parameters for the error two-port: Open (O), short (S) and match (M). To simplify the computation, an ideal match ($\Gamma_M = 0$) is assumed.[2] The open and short are characterized by their reflection coef-

1) In the relevant literature, this technique is sometimes also referred to as OSL (open, short, load) or SOL. To avoid confusion with the line standard (L), we suggest using the name given here (OSM).

2) Modern network analyzers are able to model a non-ideal match.

ficients Γ_O and Γ_S. The error two-port G is characterized by the following S-parameter equation.

$$\begin{bmatrix} a_{DUT} \\ b_3 \end{bmatrix} = \mathbf{S}_G \begin{bmatrix} b_{DUT} \\ a_1 \end{bmatrix} \text{ where } \mathbf{S}_G = \begin{bmatrix} e_{11} & e_{10} \\ e_{01} & e_{00} \end{bmatrix} \qquad \textit{(3.4-1)}$$

The configuration consisting of the device under test (DUT) and the error two-port G can be combined to a one-port with the reflection coefficient Γ_E. We refer to this one-port as the equivalent load.

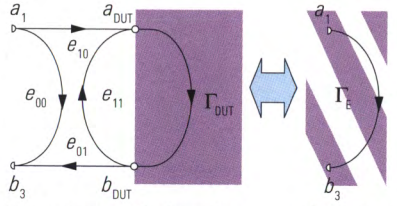


Fig. 3.4.2
Transformation into an equivalent load.

The reflection coefficient Γ_E of the equivalent load can be calculated from the matrix $\mathbf{S}_G$ and the reflection coefficient Γ_{DUT}.

$$\Gamma_E = \frac{b_3}{a_1} = e_{00} + \frac{e_{01}e_{10}\Gamma_{DUT}}{1 - e_{11}\Gamma_{DUT}} \qquad \textit{(3.4-2)}$$

The measured value M determined by the ideal network analyzer from the wave quantities a_1 and b_3 is equal to the reflection coefficient Γ_E of the equivalent load. The parameter e_{01} will never have a value of 0 in a functional network analyzer. Accordingly, we can assume that $e_{01} = 1$ so as to incorporate its influence into the parameter e_{10}.

$$M = e_{00} + \frac{e_{10}\Gamma_{DUT}}{1 - e_{11}\Gamma_{DUT}} \qquad \textit{(3.4-3)}$$

If we carry out the measurement with the standards open ($\Gamma_{DUT} = \Gamma_O$), short ($\Gamma_{DUT} = \Gamma_S$) and match ($\Gamma_{DUT} = \Gamma_M = 0$), we obtain the measured values $M = M_O$, $M = M_S$ and $M = M_M$. By combining them with the characteristic data Γ_O, Γ_S and Γ_M in formula (3.4-3), we obtain the following equations:

$$M_O = e_{00} + \frac{e_{10}\Gamma_O}{1 - e_{11}\Gamma_O} \qquad \textit{(3.4-4)}$$

$$M_S = e_{00} + \frac{e_{10}\Gamma_S}{1 - e_{11}\Gamma_S} \tag{3.4-5}$$

$$M_M = e_{00} \tag{3.4-6}$$

We can now solve these equations for the error terms e_{00}, e_{10} and e_{11}.

$$e_{00} = M_M \tag{3.4-7}$$

$$e_{10} = \frac{(\Gamma_O - \Gamma_S)(M_O - M_M)(M_S - M_M)}{\Gamma_O \Gamma_S (M_O - M_S)} \tag{3.4-8}$$

$$e_{11} = \frac{\Gamma_S (M_O - M_M) - \Gamma_O (M_S - M_M)}{\Gamma_O \Gamma_S (M_O - M_S)} \tag{3.4-9}$$

The error terms e_{00}, e_{10}, e_{11} are also referred to as the raw system data for the one-port network analyzer.

Table 3.4.1 Interpretation of the error terms in the 3-term error model.

Raw system data	Error term
Reflection tracking	e_{10}
Directivity	e_{00}
Test port match	e_{11}

The error terms e_{00}, e_{10}, e_{11} determined in this manner are used to perform system error correction for the one-port measurement. We solve formula (3.4-3) for Γ_{DUT} and insert the complex error terms along with the complex measured value M_{DUT}. We obtain a corrected value Γ_{DUT}.

$$\Gamma_{DUT} = \frac{M_{DUT} - e_{00}}{e_{10} + e_{11}(M_{DUT} - e_{00})} \tag{3.4-10}$$

3.4.2 7-term error model (TOM, TRM, TRL, TNA, UOSM techniques)

The 7-term error model is based on a two-port network analyzer which has four receivers. The 7-term error model includes two error two-ports known as G and H. During the measurement, the two test ports are

operated as active test ports in alternation. The generator in the network analyzer is connected to the currently active test port via a source switch and a state in which $a_{G2} \neq 0$ or $a_{H2} \neq 0$ is produced (Fig. 3.4.3). Since there are four receivers, the wave quantities a_{G2} and a_{H2} are both measured regardless which test port is operated as an active port. In this manner, the non-ideal properties of the source switch (like its crosstalk) can be monitored during all measurements and thus be corrected.

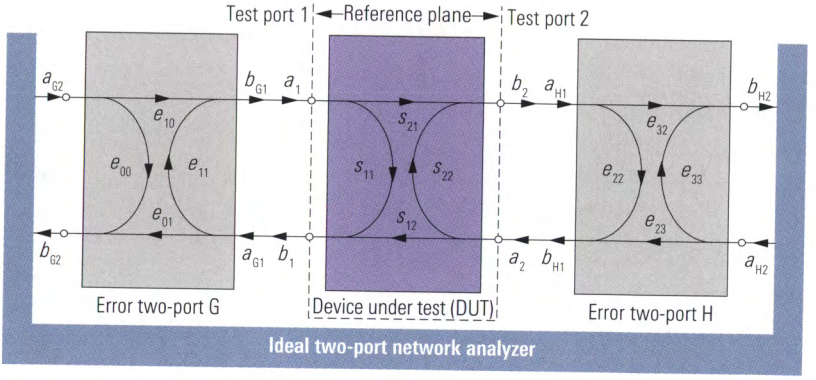


Fig. 3.4.3 7-term error model.

The error two-ports G and H are characterized by the following scattering parameter equations.

$$\begin{bmatrix} b_{G1} \\ b_{G2} \end{bmatrix} = \mathbf{S}_G \begin{bmatrix} a_{G1} \\ a_{G2} \end{bmatrix} \text{ where } \mathbf{S}_G = \begin{bmatrix} e_{11} & e_{10} \\ e_{01} & e_{00} \end{bmatrix} \qquad \textit{(3.4-11)}$$

$$\begin{bmatrix} b_{H1} \\ b_{H2} \end{bmatrix} = \mathbf{S}_H \begin{bmatrix} a_{H1} \\ a_{H2} \end{bmatrix} \text{ where } \mathbf{S}_H = \begin{bmatrix} e_{22} & e_{23} \\ e_{32} & e_{33} \end{bmatrix} \qquad \textit{(3.4-12)}$$

Out of the total of eight e terms, in a functional network analyzer the terms e_{10}, e_{32}, e_{23} and e_{01} will never have a value equal to 0. Accordingly, we can set one of these four terms to any arbitrary value that is not equal to 0. In this manner, the absolute values of the wave quantities change, but the relevant wave quantity ratios are unaffected. In the following, we assume that the parameter $e_{32} = 1$. This means that the error model contains seven independent e terms, hence its name. For the forward and reverse measurements, the same identical error model is used (Fig. 3.4.3). The following table is used to assign the e terms to the raw system data. Certain system data (transmission tracking, load match and isolation) are needed to correct the transmission measurements.

Raw system data	Forward measurement	Reverse measurement
Directivity	e_{00}	e_{33}
Reflection tracking	$e_{10}e_{01}$	$e_{23}e_{32}$
Source match	e_{11}	e_{22}
Transmission tracking	$e_{10}e_{32}$	$e_{01}e_{23}$
Load match	e_{22}	e_{11}

Table 3.4.2 Interpretation of the parameters in the 7-term error model.

The reflection parameter e_{11} or e_{22} which is assigned to each test port is independent of whether the port is active or passive. Accordingly, the load match for the reverse measurement is the same as the **source match** for the forward measurement and vice versa (rows with blue background in Table 3.4.2).

During the calibration procedure, one and two-port standards are measured. If we measure a one-port standard Γ_{Std} at test port 1, the measured value $M = b_{G2}/a_{G2}$ corresponds to the input reflection coefficient of the equivalent load consisting of error two-port G and standard Γ_{Std} (see Fig. 3.4.2 and formula (3.4-2)).

$$\frac{b_{G2}}{a_{G2}} = e_{00} + \frac{e_{01}e_{10}\Gamma_{Std}}{1 - e_{11}\Gamma_{Std}} \qquad \textit{(3.4-13)}$$

The same considerations apply analogously to a measurement of the one-port standard at test port 2.

$$\frac{b_{H2}}{a_{H2}} = e_{33} + \frac{e_{23}e_{32}\Gamma_{Std}}{1 - e_{22}\Gamma_{Std}} \qquad \textit{(3.4-14)}$$

Measurement of a one-port standard at both test ports thus gives two independent equations for the seven unknown e terms.

To characterize the measurement of a two-port standard, we convert the matrices $\mathbf{S}_G$ and $\mathbf{S}_H$ into the corresponding chain transfer matrices $\mathbf{T}_G$ and $\mathbf{T}_H$. Furthermore we assume that the two-port standard is characterized by its chain transfer matrix $\mathbf{T}_{Std}$. For the cascaded network consisting of error two-port G, two-port standard and error two-port H, we obtain the following equation:

$$\begin{bmatrix} b_{G2} \\ a_{G2} \end{bmatrix} = \mathbf{T}_{tot} \begin{bmatrix} a_{H2} \\ b_{H2} \end{bmatrix} \text{ where } \mathbf{T}_{tot} = \mathbf{T}_{G} \cdot \mathbf{T}_{Std} \cdot \mathbf{T}_{H} \qquad \textit{(3.4-15)}$$

In this manner, the reverse measurement of the two-port standard is modelled. To describe its forward operation, we calculate the matrix $\mathbf{T}_{tot,rev}$ from the matrix $\mathbf{T}_{tot}$ using a suitable matrix transformation.

$$\begin{bmatrix} b_{H2} \\ a_{H2} \end{bmatrix} = \mathbf{T}_{tot,rev} \begin{bmatrix} a_{G2} \\ b_{G2} \end{bmatrix} \qquad \textit{(3.4-16)}$$

For a two-port standard, we obtain with formulas (3.4-15) and (3.4-16) exactly four equations for the e terms. Overall, the calibration process should produce at least seven equations so that the seven unknown terms in the error model can be determined. If all of the standards are fully known, calibration will yield eight equations, meaning that the procedure is overdetermined. Accordingly, in the 7-term error model, we can also use standards with characteristics that are partially unknown for calibration purposes. Table 3.4.3 shows how many equations are yielded by the individual standards. This allows us to determine the total number of equations for each of the techniques.

	TOM	TRM	TRL	TNA	UOSM
One-port standards					
Open (O)	2	–	–	–	2
Short (S)	–	–	–	–	2
Match (M)	2	2	–	–	2
Reflect (R)	–	1	1	–	–
Two-port standards					
Through (T)	4	4	4	4	–
Line (L)	–	–	2	–	–
Network (N)	–	–	–	1	–
Attenuator (A)	–	–	–	2	–
Unknown through	–	–	–	–	1
Total number of equations	8	7	7	7	7

Table 3.4.3 Number of equations in the 7-term techniques.

Five calibration techniques that are based on the 7-term error model are presented below.

TOM technique [ES91]

The TOM technique uses fully characterized standards. Together, the two one-port standards open (O) and match (M) produce four equations that

are supplemented by four additional equations for the through standards (T). Since the technique is overdetermined, it enables an implicit plausibility check. Connection of incorrect or defective standards as well as faulty electrical connections can be detected immediately after the calibration is finished.

TRM technique [ES91]
In the TRM technique, a reflect standard is used instead of the open. The same reflect standard must be used for all of the test ports.

TRL technique[1] [EH97]
As its reference impedance, the TRL technique uses a line with a corresponding characteristic impedance. This replaces the match. Ideally, it is an air line. Since it is possible to manufacture the characteristic impedance of the air line to be more precise than the impedance of a fixed match, this technique primarily produces higher effective directivity, which also results in an improvement in the test port match. For more information about the line length and the related boundary conditions for the start and stop frequencies, see the discussion of line standards in section 3.3.1.

TNA technique [ES91]
The reflection standard used in the previous techniques is replaced here by a two-port network known as a symmetrical network. Instead of the line standard, an attenuator standard is used with very good matching (but arbitrary attenuation). The attenuation should fall within the range described in section 3.3.1 in the discussion about attenuator standards.

UOSM technique [FP92]
In the UOSM technique, a complete one-port calibration is performed using the standards open, short and match on each of the test ports that are involved in the process. Individual one-port standards can be used for each test port depending on the connector type and gender. The relevant characteristic data must be taken into account. An unknown through replaces the through used in the TOM technique. Since, for example, a conventional adapter meets the requirements of the unknown through,

1) In the relevant literature, the LRL technique is also described which uses another line (L) in place of the through (T). In the **LRL technique**, the boundary condition for the start and stop frequencies is calculated from the length difference between the two lines instead of from the length difference between the line and through.

the UOSM technique is particularly well suited for calibration involving different connector types at the test ports.

Calibration of an N-port network analyzer

To measure the wave quantities a_1, a_2, a_3, ... on all the test ports including the passive ports, a total number of $2N$ measurement receivers are provided in a full N-port network analyzer (cf. Fig. 2.1.1). For each test port, the error model has four additional parameters. Taking into account the normalization $e_{32} = 1$, the resulting system of equations has the following number of unknown terms:

$$4N - 1 \qquad \textit{(3.4-17)}$$

Based on a complete two-port calibration using one of the techniques described above, we can define seven parameters in the two-port model. It is useful to split the calibration of an N-port network analyzer into a series of two-port calibrations. The required through standards are generally available. In an N-port network analyzer, $\frac{1}{2}N(N-1)$ different two-port through connections are possible. One different approach is to take advantage of the UOSM technique. A power divider[1] that connects all the test ports can serve as an unknown (multi port) through. During all the through measurements involved in the calibration the standard remains attached unchanged to all the test ports, leading to an ease of work. With each partial measurement $(N-1)$ through measurements are done. With a total number of 3 and more test ports this will be faster than a pair-wise approach and reduce the calibration time. Of course this technique has its limits in the total number of test ports.

Calibration of a network analyzer using an N-port switching matrix

In the configuration shown in Fig. 2.10.2 only two ports of the DUT are routed to the network analyzer at a time. As a consequence all transmission measurements needed for calibration have to be done in a pair-wise manner. Please consider that the load match of the ports not routed to the analyzer is determined by the properties of the switching matrix. With DUTs having more than two ports and insufficient load match multiple path propagation will occur. In Fig. 3.4.4, two signal paths leading from port 1 to port 2 are shown in red as an example of this phenom-

1) E.g. a 4-port power divider made of four resistors each $Z_0/2$ or a 3-port power divider made of tree resistors each $Z_0/3$.

enon. Suitable calibration techniques must be used to take this into account (e.g. described in [RS05]).

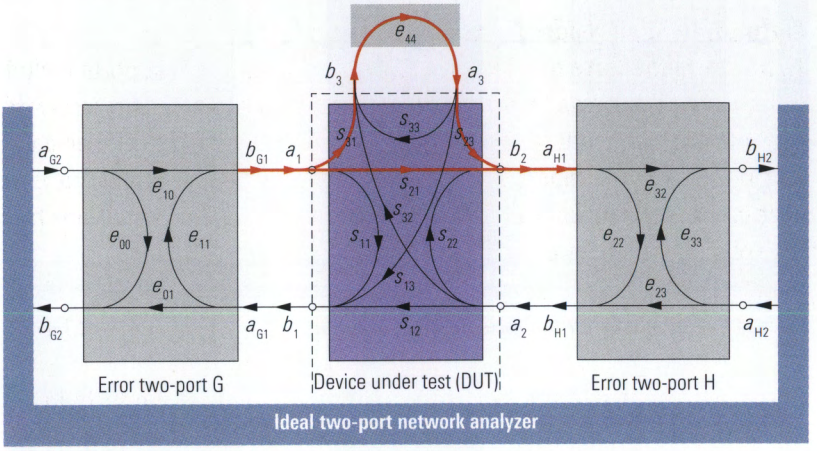


Fig. 3.4.4 Multipath propagation of the transmission.

3.4.3 10-term and 12-term error models (TOSM technique)

Not every two-port network analyzer has the four receivers needed for the 7-term technique. In terms of its design, a network analyzer with three receivers corresponds to the concept illustrated in Fig. 2.10.1 where $N = 2$. The stimulus wave is measured prior to the source switch regardless of the measurement direction. This means that the switch is part of the test set and accordingly, the error model must take the switch into account. Two separate submodels are needed for the forward and reverse directions. We have to assume that the reproducibility of the source switch characteristics is ideal in this error model. Each submodel has seven parameters. Like in the 7-term error model, in each submodel we can assume that one of the quantities e_{10}, e_{01}, e_{23}, e_{32} or e'_{01}, e'_{23}, e'_{32}, e'_{10} is equal to 1. In the following, this will be the quantities e_{10} and e'_{23}, which leads to $e_{10} = 1$ or $e'_{23} = 1$. Accordingly, both submodels together have 12 independent parameters. By omitting the crosstalk terms e_X and e'_X, we obtain the 10-term error model that is examined below. In practice the complete 12-term error model is used only in exceptional cases.

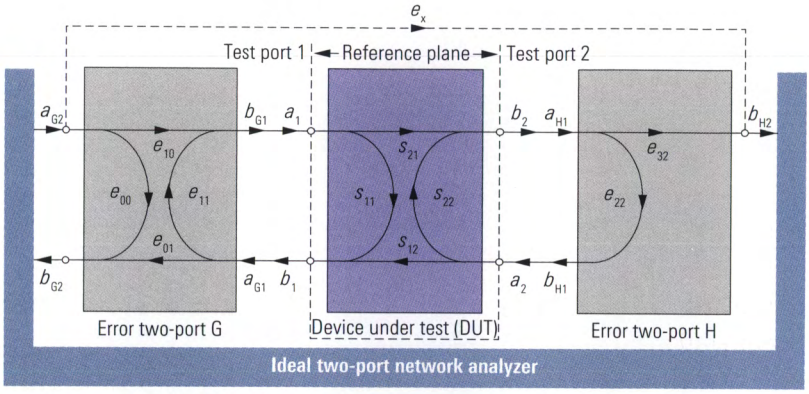


Fig. 3.4.5 Submodel for the forward measurement.

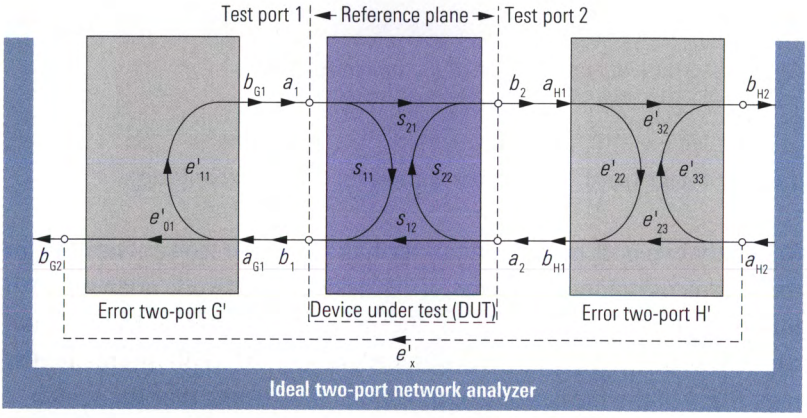


Fig. 3.4.6 Submodel for the reverse measurement.

The relevant relationships in the forward submodel can be described using the matrix $\mathbf{S}_G$ and two additional equations.

$$\begin{bmatrix} b_{G1} \\ b_{G2} \end{bmatrix} = \mathbf{S}_G \begin{bmatrix} a_{G1} \\ a_{G2} \end{bmatrix} \quad \text{where} \quad \mathbf{S}_G = \begin{bmatrix} e_{11} & e_{10} \\ e_{01} & e_{00} \end{bmatrix} \tag{3.4-18}$$

$$b_{H2} = e_{32} a_{H1} \tag{3.4-19}$$

$$b_{H1} = e_{22} a_{H1} \tag{3.4-20}$$

The reverse submodel is described analogously.

$$b_{G2} = e'_{01} a_{G1} \tag{3.4-21}$$

$$b_{G1} = e'_{11} a_{G1} \tag{3.4-22}$$

$$\begin{bmatrix} b_{H1} \\ b_{H2} \end{bmatrix} = \mathbf{S}_{H'} \begin{bmatrix} a_{H1} \\ a_{H2} \end{bmatrix} \text{ where } \mathbf{S}_{H'} = \begin{bmatrix} e'_{22} & e'_{23} \\ e'_{32} & e'_{33} \end{bmatrix} \tag{3.4-23}$$

As was the case with the 7-term error model, we can provide a general interpretation of the model quantities.

Raw system data	Forward measurement	Reverse measurement
Directivity	e_{00}	e'_{33}
Reflection tracking	$e_{01}\, e_{10}$	$e'_{32}\, e'_{23}$
Source match	e_{11}	e'_{22}
Transmission tracking	$e_{32}\, e_{10}$	$e'_{01}\, e'_{23}$
Load match	e_{22}	e'_{11}

Table 3.4.4 *Interpretation of the parameters in the 12-term error model.*

Unlike the 7-term error model (Table 3.4.2), for each test port we distinguish separate reflection parameters e_{11}, e'_{11} or e_{22}, e'_{22} for the forward and reverse direction as well as the active and passive state. The relationship between the load and source match on a test port that is known from the 7-term error model is no longer applicable here.

The calibration process using one-port standards at test port 1 or test port 2 can be described analogously to the 7-term error model as an equivalent load.

$$\frac{b_{G2}}{a_{G2}} = e_{00} + \frac{e_{01} e_{10} \Gamma_{Std}}{1 - e_{11} \Gamma_{Std}} \tag{3.4-24}$$

$$\frac{b_{H2}}{a_{H2}} = e'_{33} + \frac{e'_{23} e'_{32} \Gamma_{Std}}{1 - e'_{22} \Gamma_{Std}} \tag{3.4-25}$$

Measurement of a one-port standard at the test ports thus leads to precisely 2 equations that have the e or e′ terms as their unknowns.

The only way to obtain the transmission tracking ($e_{32}e_{10}$, $e'_{01}e'_{23}$) and the load match (e_{22}, e'_{11}) is via a transmission measurement. Measurement of a through produces, for the forward direction, the measured quantities b_{H2}/a_{G2} and b_{G2}/a_{G2}, and for the reverse direction, the measured quantities b_{G2}/a_{H2} and b_{H2}/a_{H2}. The associated equations can be specified as functions of the e and e' terms and the S-parameters for the through. The S-parameters for the through are known from their characterization. Accordingly, for the measurement of the through, we obtain four equations that have the e and e' terms as their unknowns.

TOSM technique[1] [Re74], [KS71]

To determine the ten unknown error terms (e_{00}, $e_{01...}$ e'_{33}), we use the **TOSM technique**. This involves measuring the one-port standards open (O), short (S) and match (M) on both ports of the analyzer. It is also necessary to measure the through (T).

	TOSM
One-port standards	
Open (O)	2
Short (S)	2
Match (M)	2
Two-port standards	
Through (T)	4
Total number of equations	10

Table 3.4.5
Standards in the TOSM technique.

If we would like additionally to determine the crosstalk parameters e_X and e'_X, a measurement is necessary in the forward and reverse directions in which both test ports have a match. This is a way of forcing $s_{21} = s_{12} = 0$. Measurement of the quantity b_{H2}/a_{G2} or b_{G2}/a_{H2} thus leads directly to the crosstalk parameter e_X or e'_X.

Calibration of an N-port network analyzer

In an N-port network analyzer, each one-port standard from Table 3.4.5 is measured on each test port. This produces $3N$ one-port measurements. In addition, for each two-port combination of the N ports, a measurement with the through standard is required. This produces $\frac{1}{2}N(N-1)$

1) In the relevant literature, this technique is also known as SOLT (short, open, load, through) or OSLT. To avoid confusion with the line standard (L), we suggest using the name TOSM.

two-port measurements. The total number of measurements is as follows:

$$\tfrac{1}{2}N(N+5) \qquad \textit{(3.4-26)}$$

In an 8-port network analyzer, this means 52 measurements, for example. By combining certain ports into port groups and modifying the procedure for measuring the transmission, [HS04] proposes a reduction to

$$4N \qquad \textit{(3.4-27)}$$

For example, in an 8-port network analyzer, this involves 32 measurements.

Calibration of a network analyzer using a switching matrix

Based on Fig. 3.4.5 and Fig. 3.4.6, what we need is a error model that is dependent on the measurement direction and thus has a separate submodel for all of the $(N-1)N$ two-port measurement directions. The $N-2$ test ports of a submodel for which the match is implemented in the switching matrix can be described using a reduced load port equivalent circuit with load impedances $Z_{\text{Matrix, k}}$ where $k = 1, 2, \ldots, N$.

3.4.4 15-term error model (TOM-X technique)

The design of the previous error models always incorporated a certain amount of information about the configuration of the instruments. It is also possible, however, to design the error network as a blackbox without any prior information.

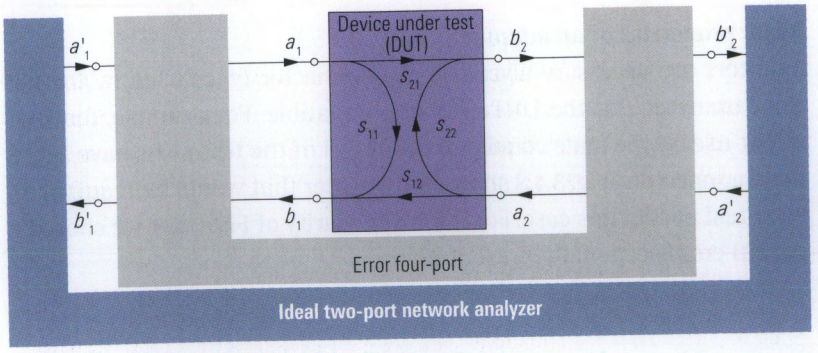


Fig. 3.4.7 15-term error model.

This forces us to associate all of the measurement and wave quantities. In the case of a two-port network analyzer, we would assume an error four-port. Out of the 16 S-parameters in this network, one can be defined to be equal to 1 like we previously saw in section 3.4.2. Accordingly, for a two-port instrument, we need 15 error terms. The advantage of this error model is that it is complete. It is the only error model, which is capable of modeling crosstalk that is dependent on the DUT.

TOM-X technique [SG03]
Based on the 15-term error model, we can derive the TOM-X calibration technique. The name reflects the standards used in the calibration process. The letter X stands for the ability to eliminate crosstalk. Determination of the 15 parameters requires a significantly higher number of calibration measurements than were required, for example, using a 7-term technique. In the case of a two-port network analyzer, we need a through, two opens and two matches. Contact with the one-port standards is always made in pairs as open-open, open-match, match-open and match-match. All of the standards must be fully known. So far, TOM-X calibration technique has been used only in exceptional cases.

Calibration of an N-port network analyzer
In an N-port network analyzer, we replace the error four-port with an error 2N port network with $(2N)^2$ scattering parameters.

3.4.5 Adapters and noninsertable DUTs

In this section, we will see how adapters interact with the calibration process.

Basic properties of an adapter
Adapters are necessary whenever the connector types used by the network analyzer and the DUT are not compatible. For example, the DUT might use N-type male connectors while all of the test ports have PC3.5 male connectors. Fig. 3.4.8 shows the adapter that would be required on each test port in this case. For improved clarity of Fig. 3.4.8, we chosed a partial cross-sectional representation.

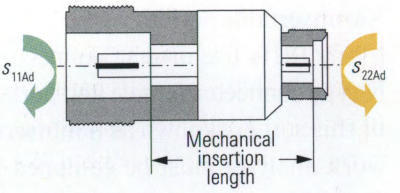


Fig. 3.4.8
Adapters N female to PC3.5 female with mechanical insertion length and the reflection coefficients s_{11Ad} and s_{22Ad}.

The maximum available frequency for the test ports can be restricted through the use of a different connector types. In this example, we experience a reduction to 18 GHz due to the usage of the N connector type. At certain frequencies above 18 GHz, resonance effects occur inside the N connectors which impair the measurement results. Some connector types are compatible with each other, so that no adapter is needed. For additional information, see Table 3.4.6. To ensure that the measurement accuracy is affected as little as possible, calibration should be performed after the adapters and test instrument have been connected using an calibration kit (e.g. N-type calibration kit in our example).

Connector type	Frequency range	Mechanically compatible with	Note
N	DC to 18 GHz	–/–	1)
14 mm	DC to 8.5 GHz	–/–	For high power levels
7-16	DC to 8.5 GHz	–/–	For high power levels
PC7	DC to 18 GHz	–/–	Sexless
PC3.5	DC to 34 GHz	SMA; K	–/–
SMA	DC to 18 GHz	3.5 mm; K	2)
2.92 mm (K)	DC to 40 GHz	SMA; 3.5 mm	–/–
2.4 mm	DC to 50 GHz	1.85 mm	–/–
1.85 mm (V)	DC to 67 GHz	2.4 mm	–/–
1.00 mm	DC to 110 GHz	–/–	–/–

Table 3.4.6 *Typical coaxial RF connector types with a 50 Ω characteristic impedance.*

Notes for the table:

1) These connectors are available alternatively in a 75 Ω version. Direct connection of the 50 Ω and 75 Ω versions is not possible and can even damage the connectors!
2) This family of connectors represents an economical alternative to the PC3.5-type connector. It uses a PTFE dielectric in the connector. Therefore it is more rugged in mechanical terms. However, it has worse RF data than the PC3.5-type connector.

Noninsertable devices

Many DUTs use diverse connector types on their RF ports (e.g. port 1: N-type connector, female and port 2: PC3.5-type connector, female). DUTs of this sort are known as **noninsertable devices.** The test ports of the network analyzer must be equipped with suitable adapters (see Fig. 3.4.9). To avoid unnecessarily impairing the measurement accuracy, the adapters should be introduced into the test setup prior to the calibration process. A suitable calibration kit is then required for each of the different types of connectors. The calibration kit must include, as a minimum, the standards that are needed for one-port calibration. The through connection between the test ports that use different connector types must be made using adapters. One problem here is that unlike through standards, the adapters are usually not characterized. However, conventional adapters do fulfill the requirements for an unknown through (section 3.3.1). The precondition is that the network analyzer must allow the use of unknown throughs as calibration standards.

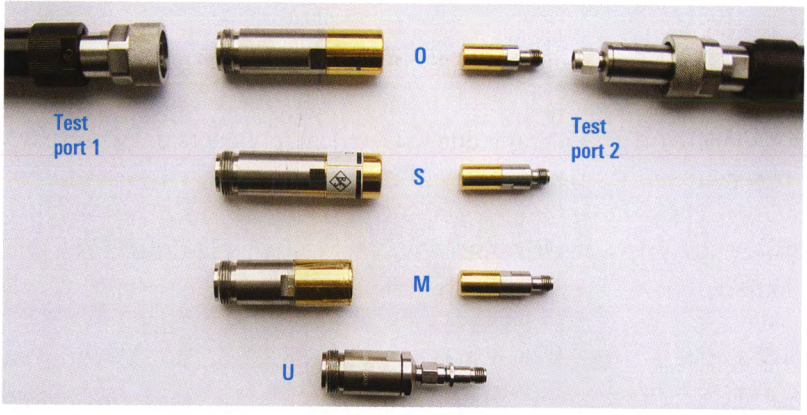


Fig. 3.4.9 Calibration standards and adapters for a noninsertable DUT.

In some network analyzers, a multi-step procedure known as "adapter removal" is performed instead of the procedure described above. It is based on the 12-term error model so that it can also be implemented with an N+1 receiver analyzer (Fig. 2.10.1).

Subsequent introduction of adapters

If none of the procedures described above can be performed (e.g. due to the lack of a suitable calibration kit), the adapter can be introduced into the test setup only after the calibration process. However, this results in degradation of the effective system data. The residual effective direc-

tivity can be calculated using formula (2.4-3). Here, $s_{\pi 11}$ corresponds to the quantities $s_{11\mathrm{Ad}}$ or $s_{22\mathrm{Ad}}$ (see Fig. 3.4.8). The insertion loss of an adapter can usually be neglected in computations using formula (2.4-3) since $s_{\pi 21} = 1$. The rule of thumb is as follows: The effective directivity is degraded by the adapter to a value of $s_{11\mathrm{Ad}}$ or $s_{22\mathrm{Ad}}$. Accordingly, the adapter should exhibit the best possible match at its ports. Here, there are considerable quality differences between more economical models and high-performance adapters with, say, 32 dB return loss at 4 GHz. To be able to measure with the proper phase, the network analyzer must take into account a suitable length offset. The mechanical insertion length of the adapter (see Fig. 3.4.8) is a function of the distance between the reference planes from the adapter input to the adapter output and not the outer (mechanical) dimensions of the adapter. The electrical insertion length that should be set on the analyzer is usually somewhat greater than the mechanical insertion length. The reasons for this are as follows:

- A lengthening by a factor of $\sqrt{\varepsilon_r}$ due to a dielectric with $\varepsilon_r > 1$ arranged in the adapter.
- An increase in the length due to discontinuities of the inner and outer conductor.
- An increase in the length due to electrical influences of the inner conductor support.

3.4.6 Incomplete calibration techniques

The calibration techniques discussed in the previous sections were based on complete error models. All relevant linear systematic measurement errors have been taken into account. In order to use these correction techniques with DUTs having more than one port, the analyzer must measure all of the S-parameters for the DUT, regardless of which of them are displayed or not. Therefore, it is not possible to reduce the measurement time by choosing only a set of wisely selected S-parameters. Furthermore technical constraints can make it impossible to measure all of the S-parameters. We will see some examples where this is true later in this section. As a matter of fact a network analyzer provides some incomplete calibration techniques for these cases.

Normalization

If the accuracy requirements are low, we can use normalization instead of the techniques described in sections 3.4.1 to 3.4.4. We can distinguish the following cases:

- Reflection normalization which normalizes one of the reflection parameters s_{11}, s_{22}, etc. Here, the reference is a calibration measurement with a short or an open.
- Transmission normalization which normalizes one of the transmission parameters s_{21}, s_{12}, etc. Here, the reference is a calibration measurement with a through.

We can see that for each normalized S-parameter, exactly one calibration measurement with one standard is required. The characteristic data of the standard are provided to the network analyzer for comparison. The analyzer only measures the S-parameters that it displays. Additional S-parameters are not required when applying the normalization. Of course, normalization covers only a portion of the linear systematic errors, meaning it cannot be used for full error correction. The result is a larger residual measurement uncertainty. Transmission normalization is suitable, for example, for the test setup shown in Fig. 7.1.5. Due to the source booster present at test port 1, it can only be operated unidirectionally, i.e. only the measurement direction from port 1 to port 2 makes sense. Furthermore, measurement of the S-parameter s_{11} at the DUT is also not possible. This means that full system error correction, which requires measured quantities including s_{12} and s_{11}, is not practicable for a s_{21} measurement. But a transmission normalization is possible.

One-path two-port calibration technique

Like the calibration technique discussed above, the measurement direction is preserved during a one-path two-port calibration (unidirectional operation). For all of the measurements needed in this technique, the same test port is always operated as the active port. A complete one-port calibration is performed only on this active test port. Transmission calibration from the active test port to the other (passive) test port is performed using a through connection. In this technique, a directional element is not necessary at the passive test port. We have a similar situation in Fig. 2.4.1 at test port 2. Due to the 13 dB attenuator pad provided there, there is a poor value for the directivity (see Table 2.4.1). Using one-path two-port calibration, measurement errors that arise as a result of multiple reflections between the passive test port and the DUT output

are not corrected. As a result, in case of DUTs that are mismatched on their output side, the passive test port must be well matched (e.g. an attenuator pad with good return loss in Fig. 2.4.1). When it comes to an s_{11} measurement with the active test port, the influence of the passive test port is dependent on the transmission characteristics of the DUT. With unidirectional DUTs (e.g. ideal microwave isolators), the matching conditions at the passive test port are not relevant. However, this is not the case for a bidirectional DUT, particularly if it has low insertion loss (e.g. a low-loss transmission line). The calibration technique described here can be extended for use with N test ports. Accordingly, transmission calibration is required from the active test port to each of the $(N-1)$ passive test ports. Since only one test port is operated as the active test port, all of the necessary measurements can be performed within one partial measurement (see section 2.7.1). For the network analyzer schemes shown in Fig. 2.1.1 and Fig. 2.10.1, the pure measurement time required for the DUT is thus identical for normalization and one-path two-port calibration.

3.4.7 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 like it is discussed in section 3.5. 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 should help you to properly assess the relevant influence factors.

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 3.4.7 *Influence of various actions on the calibration.*

Notes for the table:
1) Assuming that the network analyzer and DUT 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 shown in Fig. 2.4.5 and Fig. 2.4.6.

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.

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 3.4.8 *Summary of properties for different error models and calibration techniques.*

Notes for the table:

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.

3.5 Verification

As was already discussed at the start of chapter 3, system error correction is only capable of compensating systematic measurement errors. The correction process cannot do anything about random measurement errors. Furthermore the accuracy of system error correction is limited by the uncertainty with which we know the characteristic data of the calibration standards. In modern network analyzers, this is usually the limiting factor when making extremely precise measurements.

Of course, we must distinguish this from faulty calibrations that arise as a result of operator errors, defective network analyzers and damaged standards. The implicit plausibility check provided in the TOM calibration technique helps to detect a faulty calibration early on in the process. The T-check and the Beatty standard presented in section 3.5.1 serve the same purpose.

For verification purposes, we cannot simply reuse the standards (open, short, match) that were used for calibration. As shown in the following computational example, this sort of **recognition** of the standards is also not suited to detecting faulty calibrations. Let's assume that the match provided in a calibration kit is defective and has, instead of 50 Ω, an impedance of 25 Ω. We will consider the OSM calibration technique for this example. During the calibration procedure, the defective match is measured by the network analyzer, which measures the reflection coefficient belonging to 25 Ω as the raw measured value $M_M = -1/3$. Using formula (3.4-7), the analyzer computes the error term $e_{00} = M_M = -1/3$. After finishing the calibration process the defective match is measured instead of a DUT. The analyzer measures the raw value $M_{DUT} = M_M = -1/3$. During the system error correction using formula (3.4-10) the "corrected" reflection coefficient $\Gamma_{DUT} = (M_{DUT} - e_{00})/(\ldots) = (-1/3 + 1/3)/(\ldots) = 0$ is calculated. This is precisely the value that the user expects for a match that is not defective. But if we use a faultless match from another calibration kit, as a verification DUT the analyzer would measure a raw value $M_{VER} \approx 0$ and calculate a "corrected" result $\Gamma_{VER} >> 0$ due to the wrong e_{00} value. This would be a clear indication for a faulty calibration, a faulty standard or a faulty analyzer because we would expect a $\Gamma_{VER} \approx 0$ with a faultless match. In practice we have to assume misalignments that are much smaller than in this numerical example. Typically a dB scale is used to display the reflection a_{VER} instead of the coefficient Γ_{VER}. Assum-

ing a common measurement setup a reflection $a_{VER} > -30$ dB would be a first indication for a possible fault. However we simply do not know in advance whether one or more standards are defective and if so, which standards are defective. If we double-check the one port calibration by measuring the three standards open, short and match of a second calibration kit, we are able to calculate the effective system data. But it is necessary to take all the characteristic data of the standards used for double-checking and their uncertainties in to account. Since this uncertainties are normally in the same order of magnitude as the expected effective system data this method is only an academically example. In practice the second calibration kit has to be replaced by verification standards like described in section 3.5.2. These standards are especially designed to meet the high accuracy demands.

3.5.1 T-check and Beatty standard

T-check

The application note [1EZ43] presents the so-called T-check, which serves as a **measurement accuracy test** for network analyzers. It is based on a T-junction that has an impedance Z at its port 3. We refer to this network as a **T-checker** below.

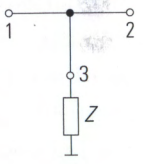


Fig. 3.5.1
Basic circuit diagram for a T-checker.

The network analyzer is connected to the ports 1 and 2 of the T-checker. As long as the length dimensions of the T-junction are small enough, we can consider it to be lossless on its own. This assumption holds for most commercially available T-junctions. The S-matrix for a lossless three-port network fulfills what is known as the unitary condition $\mathbf{S}(\mathbf{S}^*)^{\mathrm{T}} = 1$. Since this condition is fulfilled, we can derive the following equations by taking the product of the matrices $\mathbf{S}(\mathbf{S}^*)^{\mathrm{T}}$ and performing a few transformations.

$$|s_{13}| \cdot |s_{23}| = \left| s_{11} s_{21}^* + s_{12} s_{22}^* \right| \qquad (3.5\text{-}1)$$

$$|s_{13}| \cdot |s_{23}| = \sqrt{\left(1 - |s_{11}|^2 - |s_{12}|^2\right)\left(1 - |s_{21}|^2 - |s_{22}|^2\right)} \qquad (3.5\text{-}2)$$

From the right side of formulas (3.5-1) and (3.5-2), we define the following ratio:

$$c_T = \frac{\left| s_{11}s_{21}^* + s_{12}s_{22}^* \right|}{\sqrt{\left(1 - |s_{11}|^2 - |s_{12}|^2\right)\left(1 - |s_{21}|^2 - |s_{22}|^2\right)}} \qquad \textit{(3.5-3)}$$

This scalar quantity is known as the **T-check parameter**. As we can see directly from the left side of formulas (3.5-1) and (3.5-2), this quantity has a theoretical value of one although the numerator and denominator terms in formula (3.5-3) cannot be directly transformed into one another. Formula (3.5-3) does not contain the reflection coefficient for the impedance connected at port 3. In other words, the impedance at port 3 does not have any influence on the T-check parameter c_T. Accordingly, it can assume any arbitrary value $0 << Z << \infty$, e.g. 50 Ω. The S-parameters measured at ports 1 and 2 are sufficient for the T-check. They are measured by the network analyzer. However, they are subject to measurement errors and as a result, formula (3.5-3) produces a value that is different than $c_T = 1$ for them. The deviation from the ideal value indicates the presence of measurement uncertainties in the network analyzer under test.

It is possible to build your own T-checker using fairly simple means. Another advantage of the T-check is that it does not require any characteristic data. To perform a T-check, first measure the four S-parameters for the T-checker with the network analyzer under test. They are saved in a Touchstone®-compatible data format. Based on these data, the T-check program available with the application note [1EZ43] computes the T-check parameter c_T and displays it vs. the frequency axis as shown in Fig. 3.5.2. Deviations of up to ±10 % are considered small. In an analyzer that has been carefully calibrated, deviations of more than 15 % should not occur.

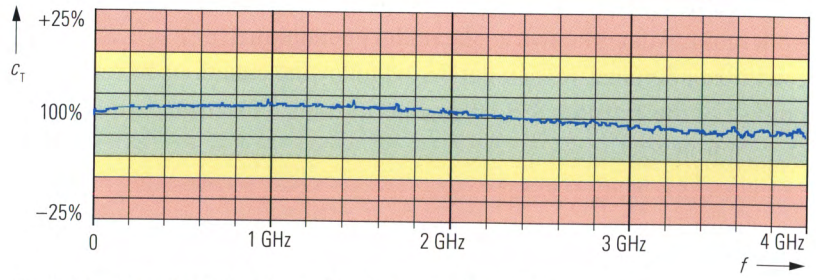


Fig. 3.5.2 Sample evaluation using the T-checker program.

Beatty standard

Named after Robert W. Beatty (NIST), this verification standard consists of a 50 Ω air line that contains a 25 Ω line section.

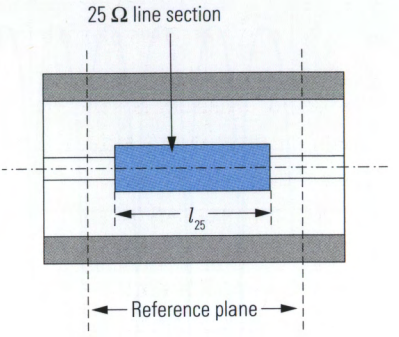


Fig. 3.5.3
Beatty standard

The Beatty standard is connected as a two-port DUT to the network analyzer. During the reflection measurement on ports 1 and 2, the 25 Ω line section causes a frequency-dependent transformation. The theoretical expected reflection values can be determined using the Smith chart as shown below.

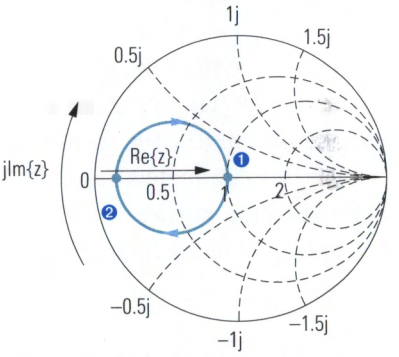


Fig. 3.5.4
Smith chart for the Beatty standard.

For $f = 0$ Hz, no transformation is to be expected; this corresponds to point ❶ in Fig. 3.5.4 where $\Gamma = 0$. As the frequency increases, the reflection coefficient moves in the clockwise direction to point ❷, which it reaches for $f_0 = c/(4l_{25})$. The 25 Ω section then produces exactly a $\lambda/4$ transformation and we read off $Z/Z_0 = 0.25$ which corresponds to $\Gamma = -0.6$. As the frequency increases further, the wavelength λ continues to drop so that finally between the length of the 25 Ω section and the wavelength, the relationship $l_{25} = \lambda/2$ holds so that the point ❶ is reached again. Further rotation about the circle causes a repetition of the same process described here. Fig. 3.5.5 shows a simulation result for the magnitudes of the reflection (blue) and the transmission (black). As

the frequency increases, the influence of the loss along the line increases. Typically, the overall loss of the line is equal to several times 0.1 dB.

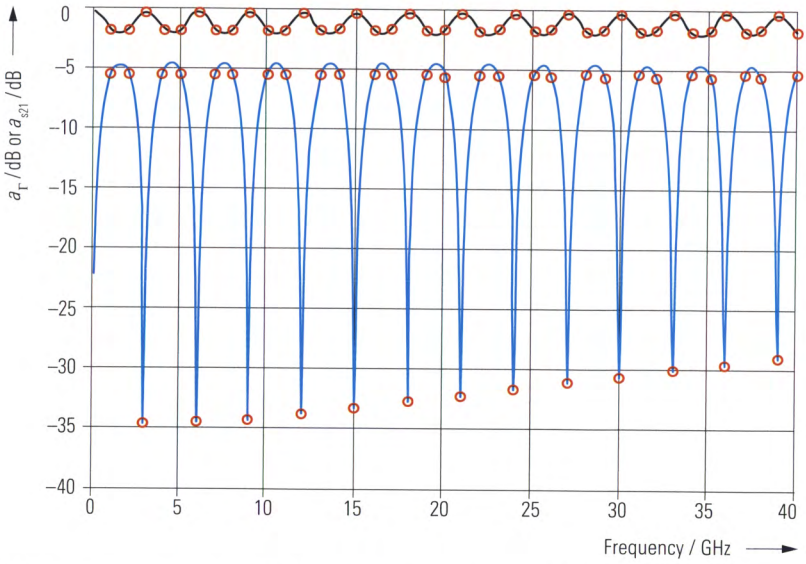


Fig. 3.5.5 Transmission (black) and reflection (blue) with losses neglected.

As we can see from the blue curve in Fig. 3.5.5, the Beatty standard covers a relatively large range for the reflection ranging from $a_r = -4.44$ dB to approx. -30 dB. The transmission a_{s21} fluctuates between 0 dB and -1.94 dB. To evaluate a result for the network analyzer under test, characteristic data must be provided with the Beatty standard. These data are indicated in the figure above using red circles.

3.5.2 Measurement of the effective system data

The accuracy information derived by the accuracy tests of the previous section are only valid for a certain DUT (the T-check or the Beatty standard). They cannot be assumed for an arbitrary DUT. The objective of a **full verification** is to assess the residual measurement uncertainties that arise after calibration for any arbitrary linear DUT. The main bases for this error calculation are the effective system data. By nature, the effective system data $e_{00\text{eff}}$, $e_{01\text{eff}}$, $e_{10\text{eff}}$, $e_{11\text{eff}}$, $e_{22\text{eff}}$, $e_{32\text{eff}}$, $e'_{01\text{eff}}$, $e'_{11\text{eff}}$, $e'_{22\text{eff}}$, $e'_{23\text{eff}}$, $e'_{32\text{eff}}$, $e'_{33\text{eff}}$ correspond to the 12 error terms in the 12-term error model described in section 3.4.3. Of course, they relate to the residual

errors on an instrument that has already been calibrated, while the error terms relate to the uncalibrated network analyzer.

Effective directivity

The test setup shown below is connected to the test port of interest. The **mismatch** that is used should have a return loss of about 20 dB. A mismatch with significantly better or worse match is not suitable for this measurement!

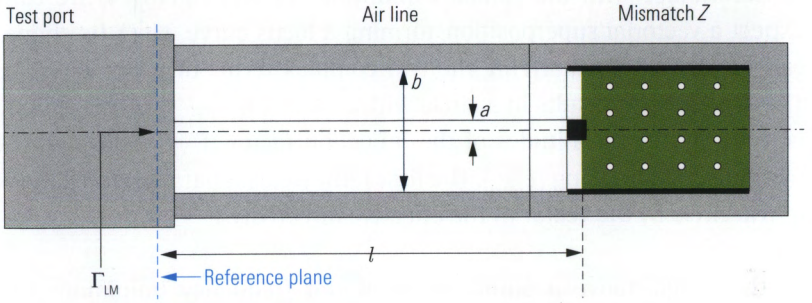


Fig. 3.5.6 Configuration for determining the effective directivity.

As was mentioned in section 3.3.1, it is possible to manufacture an air line with a specified characteristic impedance Z_C with great precision. Based on the ratio of the diameters a and b, we can use formula (3.3-6) to check this impedance and trace it back to a national length standard. In the current test setup, the air line has dimensions such that its characteristic impedance Z_C corresponds as exactly as possible to the reference impedance Z_0. Accordingly, we will assume from here on that $Z_C = Z_0$.

Before doing any further considerations we have to detail the reflection coefficient Γ_{LM} that is provided by our verification standard: The mismatch Z is transformed by the air line corresponding to its length l and the frequency f. The related complex reflection coefficients Γ_{LM} occurring at different frequencies lie on a circle. Since $Z_C = Z_0$, the center of the circle is located at the origin $\Gamma = 0$ of the complex reflection coefficient plane.

When measuring the reflection coefficient Γ_{LM} with a calibrated vector network analyzer the effective system data will influence the measurement result M. The relationship between the reflection coefficient of a DUT Γ_{DUT}, the error terms e_{00}, e_{10}, e_{11} and the measured value M

was described in section 3.4 using formula (3.4-3). This formula can also be used for the effective system data $e_{00\,\mathrm{eff}}$, $e_{10\,\mathrm{eff}}$, $e_{11\,\mathrm{eff}}$ and the current reflection coefficient Γ_{LM}. Since $|\Gamma_{\mathrm{LM}}| \approx 0.1$ (20 dB) and $|e_{11\,\mathrm{eff}}| < 1$ the product of Γ_{LM} and $e_{11\,\mathrm{eff}}$ fulfills the condition $|\Gamma_{\mathrm{LM}} \cdot e_{11\,\mathrm{eff}}| << 1$ and we can make the following approximation.

$$M = e_{00\,\mathrm{eff}} + e_{10\,\mathrm{eff}} \cdot \Gamma_{\mathrm{LM}} \qquad \textit{(3.5-4)}$$

In accordance with the complex quantities of formula (3.5-4) we can expect a vectorial superposition, forming a locus curve $M(f)$. Its shape can be derived by analyzing the two complex terms of formula (3.5-4). The second term results in a circle with a radius $|\Gamma_{\mathrm{LM}} \cdot e_{10\,\mathrm{eff}}|$. This radius is given by the magnitudes of the reflection factor $|\Gamma_{\mathrm{LM}}|$ and the effective reflection tracking $|e_{10\,\mathrm{eff}}|$. The first term yields a parallel translation of the circle by the vector of the effective directivity $e_{00\,\mathrm{eff}}$.

In the range above a sufficient minimum frequency (minimum frequency is defined by the length *l* of the air line), we can assume that the phase of the reflection coefficient Γ_{LM} changes significantly faster versus frequency than any other frequency dependent behavior that might occur in the quantities $e_{00\mathrm{eff}}$ and $e_{10\mathrm{eff}}$. Thus we are able to assume the translation given by the vector $e_{00\mathrm{eff}}$ and the radius given by $|e_{10\mathrm{eff}}\,\Gamma_{\mathrm{LM}}|$ are constant values within the vicinity of one rotation around the circle. The following figure illustrates the situation assuming the condition $|e_{00\mathrm{eff}}| < |e_{10\mathrm{eff}}\,\Gamma_{\mathrm{LM}}|$. This condition is fulfilled because the effective directivity is normally much smaller than our mismatch with its 20 dB return loss.

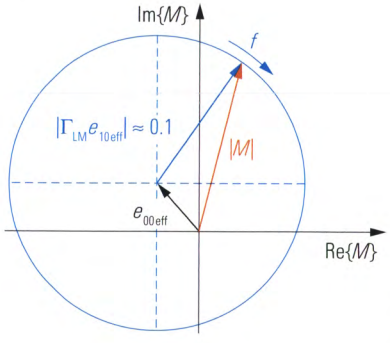


Fig. 3.5.7
Complex reflection coefficient plane with locus curve M(f).

The calibrated network analyzer performs a frequency sweep that covers the range to be verified. The magnitude of the reflection coefficient is displayed as the measured quantity using linear scaling. When looking at Fig. 3.5.7 we can imagine that the length of the vector M will change during the frequency sweep. Meaning the magnitude $|M|$ will fluctuate between $|e_{10\text{eff}}\,\Gamma_{\text{LM}}| - |e_{00\text{eff}}|$ and $|e_{10\text{eff}}\,\Gamma_{\text{LM}}| + |e_{00\text{eff}}|$. To demonstrate this on a real-word system, the described measurement was carried out yielding Fig. 3.5.8.

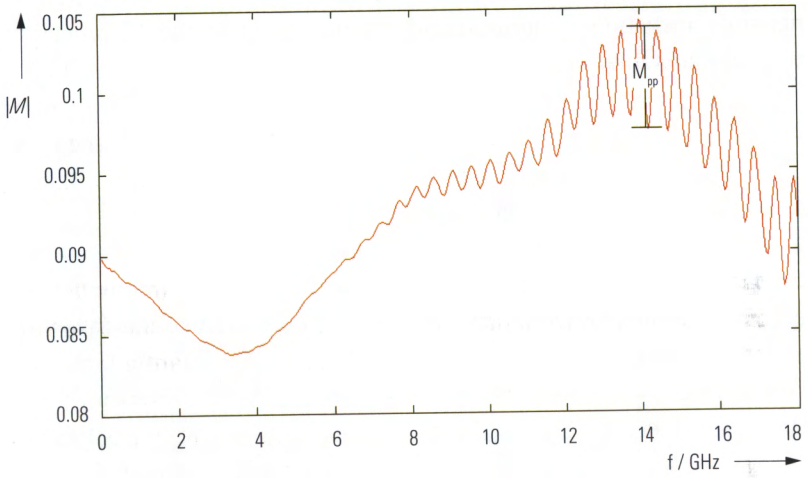


Fig. 3.5.8 Measured quantity $|M|$ vs. frequency f.

Because the magnitude $|M|$ varies by $\pm e_{00\text{eff}}$ the effective directivity $e_{00\text{eff}}$ can be calculated from the peak-to-peak ripple M_{pp} of the trace in Fig. 3.5.8.

$$|e_{00\text{eff}}| = M_{\text{pp}}/2 \qquad \textit{(3.5-5)}$$

A totally different situation happens if we do the measurement with an uncalibrated network analyzer in order to determine its raw directivity e_{00}. Under unfavorable conditions the assumption $|e_{00\text{eff}}| < |e_{10\text{eff}}\,\Gamma_{\text{LM}}|$ cannot be kept with the corresponding error terms e_{00} and e_{10}. As a consequence a complex formatted measurement result M is essential to avoid any ambiguity. In addition, raw system data such as e_{00} and e_{10} exhibit greater frequency dependence than the effective system data used in the paragraphs above. Accordingly, measuring the raw directivity e_{00} must be done at a fixed frequency. Instead of sweeping the frequency the length l has to be varied to yield a circle. For this purpose

we recommend using a sliding match (see section 3.3.1). Using its different adjustment positions at a fixed frequency, it is possible to determine individual points on the circle in the complex reflection coefficient plane (see Smith chart in Fig. 3.3.7). The center of the circle yields the vector e_{00}.

Effective source match

For this verification measurement, we use a setup like the one shown in Fig. 3.5.6. In place of the mismatch, however, we use a short. For the reflection coefficient of the short, we have $\Gamma_S \approx -1$. For the cascaded network consisting of the short and air line, we determine the reflection coefficient as follows:

$$\Gamma_{LS} = \Gamma_S \cdot e^{-j4\pi l/\lambda} \qquad \textit{(3.5-6)}$$

Assuming that the line loss is low, the reflection coefficient Γ_{LS} fulfills the condition $|\Gamma_{LS}| \approx 1$. Formula (3.4-3) from section 3.4 can be used in the same manner it was used to determine the effective directivity. However, due to the change in the proportions, we must use the following approximation:

$$M = e_{00\,\mathrm{eff}} + e_{10\,\mathrm{eff}}(\Gamma_{LS} + e_{11\,\mathrm{eff}}\,\Gamma_{LS}^2) \qquad \textit{(3.5-7)}$$

This involves an expansion of the quotient $1/(1 - e_{11\,\mathrm{eff}} \cdot \Gamma_{LS})$ from formula (3.4-3) into a power series. The series expansion is truncated after the quadratic term. This means that only the reflections shown in Fig. 3.5.9 are considered between the test port and the verification standard.

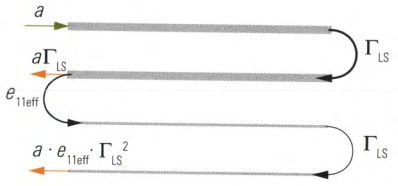


Fig. 3.5.9
Illustration of formula (3.5-7).

In our further analysis, we will assume that the contribution of the effective directivity $e_{00\,\mathrm{eff}}$ in formula (3.5-7) can be neglected. In addition, the effective reflection tracking is set as $e_{10\,\mathrm{eff}} = 1$. Applying formula (3.5-6), we obtain the following equation:

$$M = \Gamma_S\, e^{-j4\pi l/\lambda} + e_{11\,\mathrm{eff}}\,\Gamma_S^2\, e^{-j8\pi l/\lambda} \qquad \textit{(3.5-8)}$$

We can illustrate the vector superposition of the two terms from formula (3.5-8) in the complex reflection coefficient plane (see Fig. 3.5.10). The first term characterizes a vector of length $|\Gamma_S|$ that is rotating in a clockwise direction at $4\pi l/\lambda$. The second vector (added vectorially) has a length $|e_{11\text{eff}}\Gamma_S^2|$ and is rotating at twice the angular velocity $8\pi l/\lambda$.

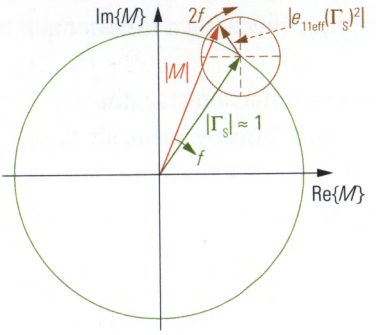


Fig. 3.5.10
Vectorial superposition for formula (3.5-8).

The magnitude of the vector M, i.e. $|M|$, is displayed by the calibrated network analyzer using a linear scaling.

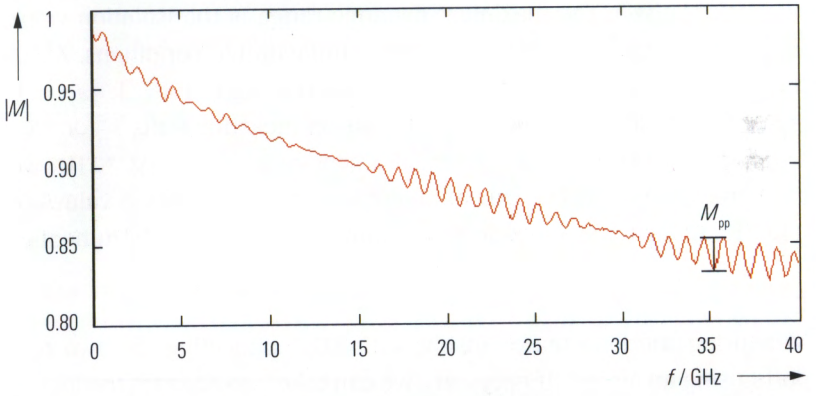


Fig. 3.5.11 Measured quantity |M| vs. frequency f.

The rotation of vector Γ_S does not affect the magnitude $|M|$ since Γ_S rotates about the center point $M = 0$. However, the rotation of vector $e_{11\text{eff}}(\Gamma_S)^2$ does affect the magnitude $|M|$. It causes the ripple $M_{pp} = 2|e_{11\text{eff}} \cdot \Gamma_S^2|$. Accordingly, we calculate the effective source match as follows:

$$|e_{11\,\text{eff}}| = \frac{M_{pp}}{2} \cdot \frac{1}{|\Gamma_S|^2} \qquad \textbf{(3.5-9)}$$

Since the value $|\Gamma_S|$ is usually unknown, we use the average of the measured value $|M|$ in the vicinity of M_{pp}. The disadvantage of this approach is that the quantities e_{00eff} and e_{10eff} are neglected. In practice they cause the periodic fluctuation effects also observed in Fig. 3.5.11. The quantities e_{00eff} and e_{10eff} cause considerable measurement uncertainty when determining the effective source match e_{11eff}. We will see a more precise technique involving a time-domain transformation in section 6.4.

Effective reflection tracking

We cannot directly measure the reflection tracking. If we use the test setup described above (air line with short), the only way to distinguish the components e_{01} and e_{11} is by using the time-domain transformation described in chapter 5. For further details, see the measurement example in section 6.4. Typical values for the effective reflection tracking are in the range ± 0.01 dB, making them difficult to verify.

Isolation

The isolation is measured with the maximum possible test port power. It is dependent on the reflection coefficients s_{11}, s_{22}, s_{33}... attached to the network analyzer. The maximum dynamic range is the isolation value exhibited by the test instrument under unfavorable conditions. All of the test ports are terminated with a short standard $|s_{11}| = 1$, $|s_{22}| = 1$, $|s_{33}| = 1$ and so on. By making transmission measurements s_{21}, s_{12}, s_{31}, s_{13}, s_{23}, s_{32} with a very small measurement bandwidth, e.g. 10 Hz, we can directly read off the maximum dynamic range. Isolation values of -140 dB and better are commonly encountered in modern instruments.

Effective load match

Measuring the load match involves directly connecting the two test ports using an air line. If necessary, we can take into account the loss of the air line as a measurement uncertainty. The S-parameter s_{11} is measured as the load match for the forward measurement, i.e. e_{22}. Likewise, the parameter s_{22} is measured as the load match for the reverse measurement e'_{11}.

3.5.3 A primer to statistics

Normal distribution (Gaussian distribution)
We can record the random errors of a measured quantity X by making repeated measurements. This involves N statistically independent observations. In general, a measured quantity X is subject to many statistically independent influence factors and is considered to have a normal distribution.[1] In the following discussion, we will first assume that $N \rightarrow \infty$. If we plot the relative frequency with which a certain measured value x_n occurs on a Cartesian coordinate system, we will obtain the bell-shaped normal distribution shown in Fig. 3.5.12. Based on $N \rightarrow \infty$ individual measurements x_n, we can calculate the **true value** x_{True} of the measured quantity X as follows:

$$x_{True} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_n \qquad \textit{(3.5-10)}$$

The **standard deviation** σ serves as a measure of the width of the bell-shaped curve. The integral over the range $x_{True} \pm \sigma$ of the bell-shaped curve corresponds to the probability that a single measurement result x_n will lie within the range $x_{True} \pm \sigma$. For a normal distribution, this probability is equal to approx. 68.3%.

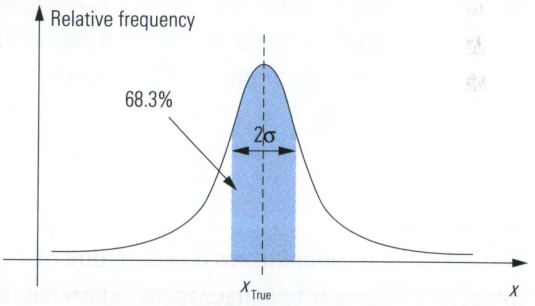


Fig. 3.5.12
Normal distribution

In practice, the number of measurements is finite, i.e. $N < \infty$. Instead of the true value x_{True}, we obtain the **estimated value** $\bar{x}$ of the measured quantity X. It is defined as the **linear average** of the individual measurement results x_n.

$$\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n \qquad \textit{(3.5-11)}$$

1) This conclusion is known as the central limit theorem in statistics e.g. [BS91] (see p. 677).

Since $N < \infty$, we obtain the **experimental standard deviation** $s(x)$ instead of the standard deviation σ.

$$s(x) = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(x_n - \bar{x})^2} \qquad \textit{(3.5-12)}$$

The usefulness of the quantities $\bar{x}$ and $s(x)$ for approximating x_{True} and σ depends on the number of measured values N. According to the theory of observational errors developed by C. F. Gauss, the quantities $\bar{x}$ and $s(x)$ represent the best possible estimation of the parameters x_{True} and σ relating to the bell-shaped curve. We want to know how closely the estimated value $\bar{x}$ comes to the value x_{True}. This is given by the experimental standard deviation of the mean. It is defined as:

$$u(x) = \frac{1}{\sqrt{N}}\, s(x) \qquad \textit{(3.5-13)}$$

The more individual measurements we make, the more accurately the estimated value $\bar{x}$ will approximate the true value x_{True}. If $N \to \infty$, we have $\bar{x} = x_{True}$ and $u(x) = 0$. The experimental standard deviation of the mean $u(x)$ provides the statistical measurement uncertainty corresponding to the normal distribution. An individual measurement x lies within the interval $x = \bar{x} \pm u(x)$ with a probability of 68%. To determine an expanded measurement uncertainty, we multiply the interval by the coverage factor k:

$$x = \bar{x} \pm k \cdot u(x) \qquad \textit{(3.5-14)}$$

The associated probabilities as a function of k are given in the following table. It is common to calculate an expanded uncertainty where $k = 2$.

k	Probability
0.675	50.00%
1.000	68.27%
1.500	86.64%
2.000	95.45%
2.500	98.76%
3.000	99.73%

Table 3.5.1
Expanded measurement uncertainties.

Rectangular distribution

Certain errors such as rounding errors have a **uniform distribution (rectangular distribution)**. In such cases, the limits a, b are known (e.g. for the rounding error resulting from truncation of all of the digits to the right of the decimal point, we have $a = x_w - 0.5$; $b = x_w + 0.5$).

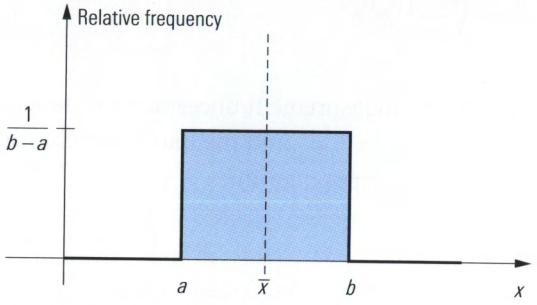


Fig. 3.5.13
Rectangular distribution

We can determine the estimated value as follows:

$$\bar{x} = \frac{b + a}{2} \qquad \textit{(3.5-15)}$$

The standard deviation is calculated by:

$$s(x) = \frac{b - a}{\sqrt{12}} \qquad \textit{(3.5-16)}$$

Combined standard measurement uncertainty

In many cases, a measured quantity Y is composed of different input quantities X_1, X_2, ... X_M, where the related uncertainties u_1, u_2, ... u_M are known. We would like to determine a overall uncertainty u_c of the measured quantity Y. Assume that the measured quantity Y is calculated using the following model equation:

$$Y = f(X_1, X_2, \ldots, X_M) \qquad \textit{(3.5-17)}$$

From this, the **partial derivatives** with respect to the input quantities X_1, X_2, ... X_M can be calculated as follows:

$$c_m = \frac{\partial f(X_1, X_2, \ldots, X_M)}{\partial X_m} \text{ where } m = 1, 2, \ldots M \qquad \textit{(3.5-18)}$$

After calculating all partial derivatives the estimated values $\bar{x}_1$, $\bar{x}_2$, ... $\bar{x}_M$ of the input quantities have to be inserted as X_1, X_2 ... X_M in formula (3.5-18). The quantities c_m where $m = 1, 2, \ldots M$ are then known

as the **sensitivity coefficients**. The combined standard measurement uncertainty u_c which is associated with the quantity Y is calculated as follows:

$$u_c = \sqrt{\sum_{m=1}^{M} (u_m c_m)^2} \qquad \textit{(3.5-19)}$$

The overall measurement uncertainty u_c depends on the sensitivity coefficients $c_1, c_2, \ldots c_m$ and the measurement uncertainties $u_1, u_2, \ldots u_m$ for all of the input quantities $X_1, X_2, \ldots X_M$.

3.5.4 Analysis of the measurement uncertainty

If we would like to specify the measurement uncertainty for an arbitrary linear DUT using the effective system data we must carry out a detailed analysis of all relevant influence quantities. The main input quantities are as follows:

- The effective system data obtained during the verification process ($e_{00eff}, e_{01eff}, e_{10eff}, e_{11eff}, e_{22eff}, e_{32eff}, e'_{01eff}, e'_{11eff}, e'_{22eff}, e'_{23eff}, e'_{32eff}, e'_{33eff}$)
- The uncertainties of the characteristic data for the verification standards that are used:
  - The mechanical tolerances of the air line
    → Uncertainty of the characteristic impedance
  - The air gap in the inner conductor contact of the connectors
    → Uncertainty of the reflection coefficient
- The random measurement uncertainties of the network analyzer under test (since they determine to what extend the result of the verification measurements can be applied to other measurements):
  - The temperature drift of the network analyzer in terms of the quantities $e_{00eff}, e_{01eff}, e_{10eff}, e_{11eff}, e_{22eff}, e_{32eff}, e'_{01eff}, e'_{11eff}, e'_{22eff}, e'_{23eff}, e'_{32eff}, e'_{33eff}$
  - The effective low-level noise (system noise floor)
  - The effective high-level noise (phase noise of the source)
  - The change in the reflection and transmission of the test port cables in case of bending
  - The repeatability of the connectors (reflection coefficient, transmission coefficient)
  - The nonlinearity of the network analyzer

The measured quantities are the magnitude or phase of the S-parameters. Depending on the particular measured quantity, the input quantities can be assembled from the quantities listed above. A complete treatment of measurement uncertainty analysis would exceed the scope of this book. As an example, however, we will consider the measurement uncertainty of the magnitude of the reflection coefficient.

If we assume we have a one-port DUT for the sake of simplicity, the remaining passive test ports of the network analyzer do not have any influence on the measurement result. We will not use any test port cables. The effective system data form an error two-port that is located between the calibrated network analyzer and the DUT. In section 3.4.1, we chose $e_{01} = 1$; this is also possible for the effective system data as $e_{01\text{eff}} = 1$.

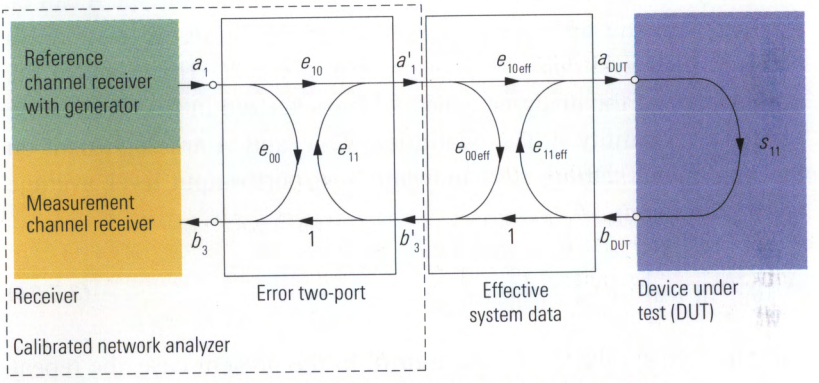


Fig. 3.5.14 Calibrated network analyzer with effective system data.

The model of the effective system data shown in Fig. 3.5.14 is characterized by formula (3.4-10) from section 3.4.1. In this formula, all we need to do is replace the error terms e_{00}, e_{10} and e_{11} with the relevant effective system data $e_{00\text{eff}}$, $e_{10\text{eff}}$ and $e_{11\text{eff}}$. Taking the place of the measured value M_{DUT} is the reflection coefficient $s_{11\text{M}}$ displayed on the calibrated network analyzer.

$$s_{11} = \frac{s_{11\text{M}} - e_{00\,\text{eff}}}{e_{10\,\text{eff}} + e_{11\,\text{eff}}(s_{11\text{M}} - e_{00\,\text{eff}})} \qquad \textit{(3.5-20)}$$

Computation of the sensitivity coefficients requires calculating the partial derivatives from formula (3.4-20) for $e_{00\text{eff}}$, $e_{10\text{eff}}$ and $e_{11\text{eff}}$ at the point where $e_{00\text{eff}} = 0$; $e_{10\text{eff}} = 1$; $e_{11\text{eff}} = 0$.

$$\frac{\partial s_{11}}{\partial e_{00\,\mathrm{eff}}} = \frac{-e_{10\,\mathrm{eff}}}{\left[e_{10\,\mathrm{eff}} + e_{11\,\mathrm{eff}}(s_{11\,\mathrm{M}} - e_{00\,\mathrm{eff}})\right]^2} \Rightarrow_{e_{00\,\mathrm{eff}}=0;e_{10\,\mathrm{eff}}=1;e_{11\,\mathrm{eff}}=0} -1 \qquad \textit{(3.5-21)}$$

$$\frac{\partial s_{11}}{\partial e_{10\,\mathrm{eff}}} = \frac{-(s_{11\,\mathrm{M}} - e_{00\,\mathrm{eff}})}{\left[e_{10\,\mathrm{eff}} + e_{11\,\mathrm{eff}}(s_{11\,\mathrm{M}} - e_{00\,\mathrm{eff}})\right]^2} \Rightarrow_{e_{00\,\mathrm{eff}}=0;e_{10\,\mathrm{eff}}=1;e_{11\,\mathrm{eff}}=0} -s_{11\,\mathrm{M}} \qquad \textit{(3.5-22)}$$

$$\frac{\partial s_{11}}{\partial e_{11\,\mathrm{eff}}} = \frac{-(s_{11\,\mathrm{M}} - e_{00\,\mathrm{eff}})^2}{\left[e_{10\,\mathrm{eff}} + e_{11\,\mathrm{eff}}(s_{11\,\mathrm{M}} - e_{00\,\mathrm{eff}})\right]^2} \Rightarrow_{e_{00\,\mathrm{eff}}=0;e_{10\,\mathrm{eff}}=1;e_{11\,\mathrm{eff}}=0} -s_{11\,\mathrm{M}}^2 \qquad \textit{(3.5-23)}$$

In the remainder of the calculation, only the magnitudes of the partial derivatives are relevant. Measurement uncertainties of type B[1] are covered fully in this manner. Measurement uncertainties of type A can be grouped into the following influences: For the noise, we distinguish between high-level noise N_1 and low-level noise N_2. There is also the nonlinearity of the analyzer, which is characterized in first approximation by the quantity A. The quantities N_1, N_2 and A are dependent on the instrument settings (IF bandwidth, test port output level, average factor and so on).

$$N = |N_1||s_{11\,\mathrm{M}}| + |N_2| + |A||s_{11\,\mathrm{M}}| \qquad \textit{(3.5-24)}$$

The effective system data show a drift. In the present case, the repeatability of the connection can also be included in the computation of this drift. The relevant sensitivity coefficients are obtained from formulas (3.5-21) to (3.5-23).

$$\Delta = |\Delta_{\mathrm{e00\,eff}}| + |\Delta_{\mathrm{e10\,eff}}||s_{11\,\mathrm{M}}| + |\Delta_{\mathrm{e11\,eff}}||s_{11\,\mathrm{M}}|^2 \qquad \textit{(3.5-25)}$$

Formulas (3.5-21) to (3.5-25) can be combined to build a model equation:

$$|s_{11}| = |s_{11\,\mathrm{M}}| + |e_{00\,\mathrm{eff}}| + |e_{10\,\mathrm{eff}}||s_{11\,\mathrm{M}}| + |e_{11\,\mathrm{eff}}||s_{11\,\mathrm{M}}|^2 \qquad \text{(systematic)}$$

$$+ |N_1||s_{11\,\mathrm{M}}| + |N_2| \qquad \text{(noise)}$$

1) For help in distinguishing types A and B, see chapter 3, first paragraph.

$$+|A||s_{11\mathrm{M}}| \qquad \text{(linearity)}$$

$$+|\Delta_{\mathrm{e}00\,\mathrm{eff}}|+|\Delta_{\mathrm{e}10\,\mathrm{eff}}||s_{11\mathrm{M}}|+|\Delta_{\mathrm{e}11\,\mathrm{eff}}||s_{11\mathrm{M}}|^2 \qquad \text{(drift)} \qquad \textbf{(3.5-26)}$$

Symbol	Quantity
$\lvert s_{11}\rvert$	Magnitude of the reflection coefficient (measured quantity)
$\lvert s_{11\mathrm{M}}\rvert$	Magnitude of the reflection coefficient displayed on the test instrument
$\lvert e_{00\mathrm{eff}}\rvert$	Magnitude of the effective directivity
$\lvert e_{10\mathrm{eff}}\rvert$	Magnitude of the effective reflection tracking
$\lvert e_{11\mathrm{eff}}\rvert$	Magnitude of the effective source match
$\lvert A\rvert$	Magnitude of the display linearity for the reflection
$\lvert N_1\rvert$	Magnitude of the effective high-power noise
$\lvert N_2\rvert$	Magnitude of the effective low-power noise
$\lvert \Delta_{\mathrm{e}00\mathrm{eff}}\rvert$	Magnitude of the drift of the effective directivity
$\lvert \Delta_{\mathrm{e}10\mathrm{eff}}\rvert$	Magnitude of the drift of the effective reflection tracking
$\lvert \Delta_{\mathrm{e}11\mathrm{eff}}\rvert$	Magnitude of the drift of the effective source match

Table 3.5.2 Symbol table

In a sample computation, we assume a measured value of -30.46 dB. We first need to convert this value to its linear format $|s_{11\mathrm{M}}| = 0.03$. To evaluate formula (3.5-19), it is helpful to use Table 3.5.3 below. All of the quantities contained in the table are linear (i.e. not in dB). For all of the input quantities, we need to determine the magnitudes shown under standard measurement uncertainty using suitable verification measurements. We can read off the sensitivity coefficients from formula (3.5-26). The uncertainty contribution $u_{\mathrm{m}}c_{\mathrm{m}}$ for each of the input quantities are calculated from the standard measurement uncertainty u_{m} and the sensitivity quotient c_{m}. As described by formula (3.5-19), the overall uncertainty can be determined in the last column of the table. The final result of the computation is the expanded measurement uncertainty for $k = 2$ according to formula (3.5-14). In this example, the true value of the measured quantity lies within the interval 0.03 ± 0.0034 with a probability of about 95%. We can also convert these linear values to their logarithmic equivalents -30.46 dB $+ 0.48$ dB to -30.46 dB $- 0.5$ dB. The standard measurement uncertainties given in column 4 of Table 3.5.3 are typical values and do not refer to any particular network analyzer.

Quantity	Estimated value	Distribution	Standard measurement uncertainty	Sensitivity coefficient c_i	Uncertainty contribution
$\lvert s_{11M} \rvert$	0.03				
$\lvert e_{00eff} \rvert$	0	Normal	0.00123	1	0.0012
$\lvert e_{10eff} \rvert$	0	Rectangular	0.00365	$\lvert s_{11M} \rvert$	0.0001
$\lvert e_{11eff} \rvert$	0	Normal	0.00306	$\lvert s_{11M} \rvert^2$	0.0000
$\lvert A \rvert$	0	Rectangular	0.00033	$\lvert s_{11M} \rvert$	0.0000
$\lvert N_1 \rvert$	0	Normal	0.00025	$\lvert s_{11M} \rvert$	0.0000
$\lvert N_2 \rvert$	0	Normal	0.00002	1	0.0000
$\lvert \Delta_{e00eff} \rvert$	0	Rectangular	0.00121	1	0.0012
$\lvert \Delta_{e10eff} \rvert$	0	Rectangular	0.00121	$\lvert s_{11M} \rvert$	0.0000
$\lvert \Delta_{e11eff} \rvert$	0	Rectangular	0.00144	$\lvert s_{11M} \rvert^2$	0.0000
$\lvert s_{11} \rvert$	0.03				0.0017
		Expanded measurement uncertainty where $k = 2$:			0.0034

Table 3.5.3 *Summary of measurement uncertainty components.*

We can determine the measurement uncertainties for various measured values $\lvert s_{11M} \rvert$ accordingly. The result is shown in the following diagram.

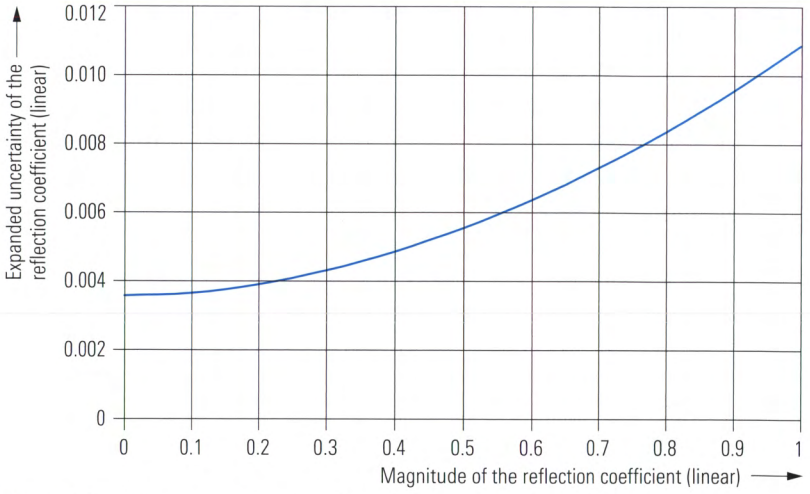


Fig. 3.5.15 *Measurement uncertainty of the reflection coefficient vs. its magnitude.*

We must distinguish the measurement uncertainties described here from the manufacturer's guaranteed values. The guaranteed values are

tested on each newly manufactured instrument. To keep the cost of such testing within bounds, the process is limited to a few selected measured values (see Fig. 3.5.16). Since the tolerance scheme must be fully covered, suitable margin must be available compared to the curve shown in Fig. 3.5.15.

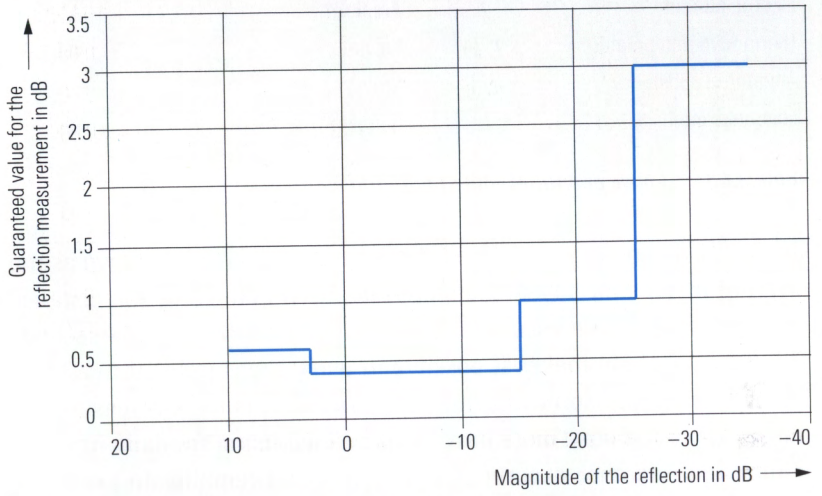


Fig. 3.5.16 Guaranteed values for the reflection coefficient.

3.6 Traceability

Measurements form an essential basis for ensuring the quality of products and services. The accuracy that is specified for measurement results must be verifiable by all of the involved parties. This is based on traceability along with a mandatory system of units.

3.6.1 The International System of Units

Many physical quantities are interrelated and their units can be converted among one another. A **system of units** highlights a set of units as base units. All other units are derived from these base units. The **International System of Units** (French: Système International d'Unités) or SI for short was agreed upon in 1960 and has since been adopted by all of the major industrial nations. It is based on the seven **SI base units** listed in the following table.

SI base quantity		SI base unit	
Name	Variable	Name	Symbol
Length	l	Meter	m
Mass	m	Kilogram	kg
Time	t	Second	s
Electric current	I	Ampère	A
Thermodyn. temperature	T, Θ	Kelvin	K
Amount of substance	n, v	Mole	mol
Luminous intensity	I_v	Candela	cd

Table 3.6.1 *SI base quantities and SI base units.*

Originally, the meter was defined using a material artifact known as the **original meter**. Since 1983, the meter has been defined as the distance that light travels in a vacuum over an interval of 1/299792458 seconds. The new definition makes it possible to avoid measurement uncertainties due to comparison with a material artifact. It is for this reason that almost all of the base units have been defined using **fundamental constants** and reference experiments. The kilogram remains an exception to this day. The following **derived SI units** have been specified for use in the field of electrical engineering.

Quantity		Derived SI unit		Alternative SI representations
Name	Variable	Name	Symbol	
Energy	W	Joule	J	$\mathrm{m^2\,kg\,s^{-2} = Ws}$
Power	P	Watt	W	$\mathrm{m^2\,kg\,s^{-3} = VA}$
Electric potential	U	Volt	V	$\mathrm{m^2\,kg\,s^{-3}\,A^{-1}}$
Electric charge	Q	Coulomb	C	As
Electric capacitance	C	Farad	F	$\mathrm{m^{-2}\,kg^{-1}\,s^4\,A^2 = A\,s\,V^{-1}}$
Electric resistance	R	Ohm	Ω	$\mathrm{m^2\,kg\,s^{-3}\,A^{-2} = V\,A^{-1}}$
Electric conductance	G	Siemens	S	$\mathrm{m^{-2}\,kg^{-1}\,s^3\,A^2 = \Omega^{-1}}$
Magnetic flux	Φ	Weber	Wb	$\mathrm{m^2\,kg\,s^{-2}\,A^{-1} = Vs}$
Magnetic flux density	B	Tesla	T	$\mathrm{kg\,s^{-2}\,A^{-1} = Wb\,m^{-2}}$
Inductance	l	Henry	H	$\mathrm{m^2\,kg\,s^{-2}\,A^{-2} = Vs\,A^{-1}}$
Frequency	f	Hertz	Hz	$\mathrm{s^{-1}}$

Table 3.6.2 *Derived electrical and magnetic quantities and units.*

In the future, **quantum effects** will play an important role in the transfer of the electrical units V, A, Ω due to the higher accuracy they enable. In

this connection, the **Josephson constant** and the **Von Klitzing constant** were elevated to the rank of natural constants in 1990.

The International System of Units is designed so that all of its units are built from exponential **products** of the base units and no additional numerical factors are required. For example, consider the unit Volt $V = m^2 \cdot kg \cdot s^{-3} \cdot A^{-1}$. To avoid unwieldy **numbers**, a set of **SI prefixes** has been defined. These prefixes indicate **decimal fractions** or **decimal multiples** of a given unit (e.g. $0.001\ V = 1\ mV$, $1\ kV = 1000\ V$). The prefixes centi, deci, hecto and deca are to be used only with units with which they are conventionally associated (e.g. cm, hl).

Decimal fractions			Decimal multiples		
Prefix	Symbol	Factor	Prefix	Symbol	Factor
Yocto	y	10^{-24}	Yotta	Y	10^{24}
Zepto	z	10^{-21}	Zetta	Z	10^{21}
Atto	a	10^{-18}	Exa	E	10^{18}
Femto	f	10^{-15}	Peta	P	10^{15}
Pico	p	10^{-12}	Tera	T	10^{12}
Nano	n	10^{-9}	Giga	G	10^{9}
Micro	µ	10^{-6}	Mega	M	10^{6}
Milli	m	10^{-3}	Kilo	k	10^{3}
Centi	c	10^{-2}	Hecto	h	10^{2}
Deci	d	10^{-1}	Deca	da	10^{1}

Table 3.6.3 SI prefixes

Every physical quantity is represented using the following product:

Numerical value · optional SI prefix · SI unit *(3.6-1)*

3.6.2 The pseudo units dB and dBm

The applicable standards specify the usage of a logarithmic form for certain ratios of quantities. For example, this book uses the **power gain figure** a_p and the **power attenuation figure** d_p. Both of these quantities take on the dB (decibel) pseudo unit.

$$a_p = 10 \lg \frac{P_2}{P_1} \text{dB}; \quad d_p = -10 \lg \frac{P_2}{P_1} \text{dB} \qquad \textit{(3.6-2)}$$

In order to display absolute power values logarithmically (like ratio quantities), we normalize the power P to a reference power of $1\ \mathrm{mW}$. What we obtain is the power level L_p in dBm as follows:

$$L_p = 10\lg\frac{P}{1\,\mathrm{mW}}\,\mathrm{dBm} \qquad \textit{(3.6-3)}$$

Besides the **power level** L_p there exist other level definitions in the standards, such as the voltage level with the unit $\mathrm{dB\mu V}$. However, these definitions are fairly irrelevant for network analysis. In this book, instead of power level we generally use the term **level.** A power gain figure a_p can be determined alternatively using the difference between the power levels L_{p2} and L_{p1} associated with P_2 and P_1.

$$\frac{a_p}{\mathrm{dB}} = \frac{L_{p2}}{\mathrm{dBm}} - \frac{L_{p1}}{\mathrm{dBm}} \qquad \textit{(3.6-4)}$$

The reflection coefficient Γ was introduced in chapter 1, formula (1.2-1) as the ratio of the incident and reflected waves. The relevant wave quantities are associated with the power values $|a|^2$ and $|b|^2$. The quantity a_Γ which is known as the **reflection**, is defined like any power gain figure.

$$a_\Gamma = 10\lg\frac{|b|^2}{|a|^2}\,\mathrm{dB} = 20\lg\frac{|b|}{|a|}\,\mathrm{dB} = 20\lg|\Gamma|\,\mathrm{dB} \qquad \textit{(3.6-5)}$$

The power attenuation figure $d_\Gamma = -a_\Gamma$ is known as the **return loss.** In comparison, we can calculate, for example, the **transmission** a_{s21} and the **transmission loss** d_{s21} from the S-parameter s_{21}. For the parameter of the S-matrix, we can generalize this as follows:

$$a_{sik} = -d_{sik} = 20\lg|s_{ik}|\,\mathrm{dB} \text{ where } i = 1, \ldots N \text{ and } k = 1, \ldots N \qquad \textit{(3.6-7)}$$

3.6.3 Some important non-SI units

The mechanical dimensions of rectangular waveguides or substrate raw materials are commonly specified using the "inch", which is an Anglo-American unit. The inch (abbreviation: in) is not an SI unit. To convert it to the SI system, we have the following relationship:

$$1 \text{ in} = 25.4 \text{ mm} \qquad \textit{(3.6-8)}$$

The inch cannot be combined with SI prefixes, e.g. m (milli). However, there does exist the unit mil, which is derived from the inch.

$$1 \text{ mil} = 1/1000 \text{ inch} = 25.4 \text{ µm} \qquad \textit{(3.6-9)}$$

A comprehensive listing of other non-SI units along with conversion formulas can be found, for example, in the relevant literature [Ta95].

3.6.4 An organization for traceability

The **national and international standards** are held on trust by the respective 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. These institutes maintain the standards needed to determine the SI base units in the form of material artifacts (e.g. original kilogram) or defined test setups (e.g. atomic clock). These are commonly referred to as **primary standards**. A measurement result is said to be **traceable** if there exists an uninterrupted chain of comparisons that ends with a primary standard (**calibration chain**). The measurement uncertainty must be calculated for each step in the calibration chain based on generally accepted principles. It must be documented so as to enable computation of an overall uncertainty for the entire chain. The relevant information is included in the **calibration certificate**, for example. The test equipment used must be checked at regular calibration intervals in a traceable manner. The offices and laboratories that are involved in this process are organized hierarchically. To perform a DKD calibration for example (DKD: Deutscher Kalibrierdienst, i.e. German Calibration Service), the DKD laboratory must be **accredited**. For each DKD laboratory, there is an exact specification of which measured quantities, measurement ranges, frequency ranges and so on the laboratory is accredited for.

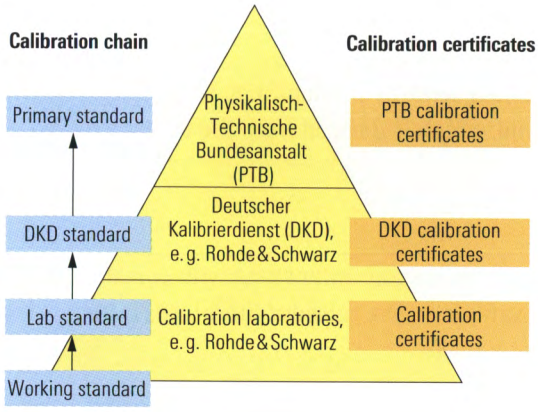


Fig. 3.6.1
Hierarchy of organizations involved in traceability (example: Germany).

3.6.5 Traceability of a network analyzer

Tracing the measurement uncertainty of a network analyzer involves appropriate verification measurements on the calibrated instrument (DUT). The verification process must be supplemented with a measurement uncertainty analysis (see section 3.5.4).

For the reflection measurement, air lines manufactured with great precision are used as the characteristic impedance standards. Section 3.5.2 makes use of such standards for determining the effective system data. The characteristic impedance of an air line can be calculated from the diameters of the inner and outer conductors, as was described in section 3.3.1 with formula (3.3-6). Both of these quantities are measured at the PTB with a resolution of 0.1 µm and traced back to the national length standard [PT04].

For the transmission measurement, the phase is traced back to the length of the air line and the national time standard. The length of the inner and outer conductors of the air line is traced back to the national length standard. Determining the transmission attenuation involves the use of extremely precise attenuator pads. At the PTB, they are characterized, for example, using a coaxial attenuation test station that is based on the **DC power substitution method** [PT04].

4 Linear measurements

This chapter 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.

4.1 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"). The exact location of the reference plane is shown in the cross-sectional drawing of the connectors found in Figs. 3.2.3 and 3.2.4. It can be moved using the corresponding operating functions provided by the analyzer (c.f. section 4.6, 4.9 and 5.4 for example). 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 of the 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, which can be done without reading section 3.

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.



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.

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.
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)).



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. If 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). For more information on coaxial connector types, see Table 3.4.6.

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. Section 3.4.7 describes circumstances under which calibration can lose its validity.



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, O, 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 $z = 0 + 0\mathrm{j}$ 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.



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.

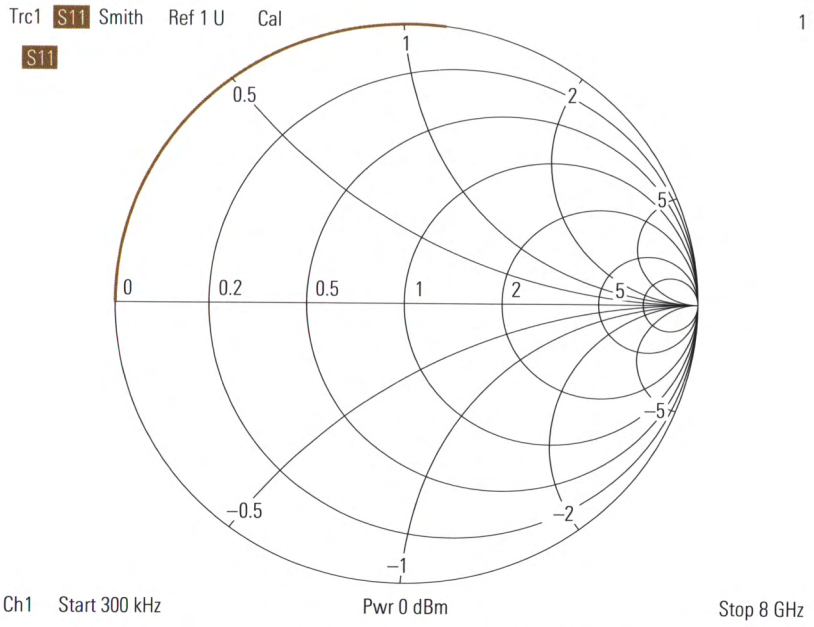


Fig. 4.1.1 Measurement of a short using the calibrated network analyzer.

4.2 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.



Fig. 4.2.1 Microstrip line as a through standard (left) and substrate without a transmission line as a symmetrical network standard (right).

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)]



Test fixtures generally have a mechanically rugged design, but their inner conductors and the often brittle substrates do necessitate careful handling. To avoid damage, pay careful attention to the instructions found in section 3.3.3.

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.
   - Define the through (T) standard by entering its exact electrical length and its losses.
   - The symmetrical network (N) shown in Fig. 4.2.1 is specified as an approximate open.
   - 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.

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.

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. However, this verification has its limitations (see ⓘ on page 159).

4.3 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, two-port 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.

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 detailed description about this is not provided in here. It is part of almost any RF introduction book e.g. see [MG92] pages F15 - F18. 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 $Z_0 = 50\ \Omega$. 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 e.g. used in [MG92] to determine a matching network.

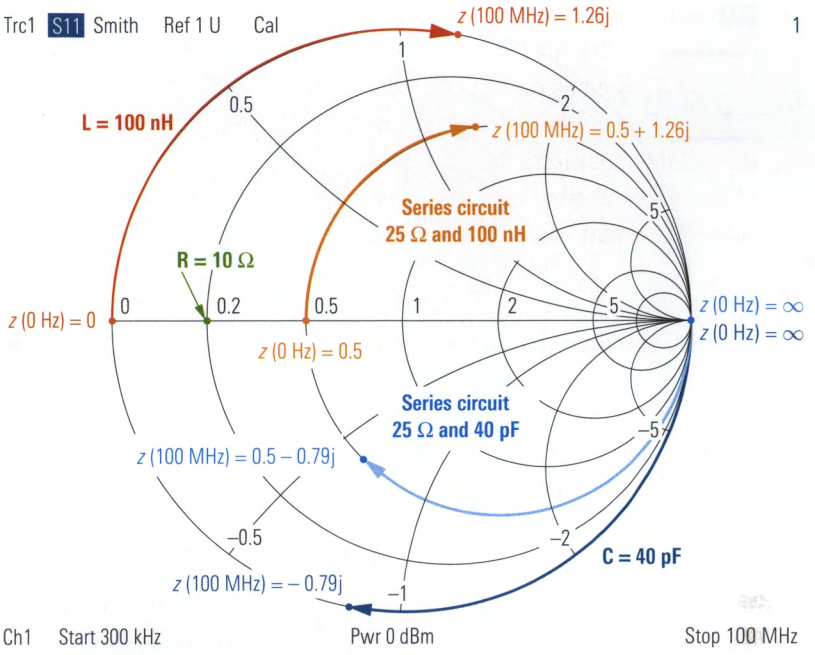


Fig. 4.3.1 Measurement of typical one-port components (reference impedance 50 Ω).

A resistor with a value of 10 Ω 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 $z = 1.26\text{j}$. 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 $z = 0.5 + 1.26\text{j}$. 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.

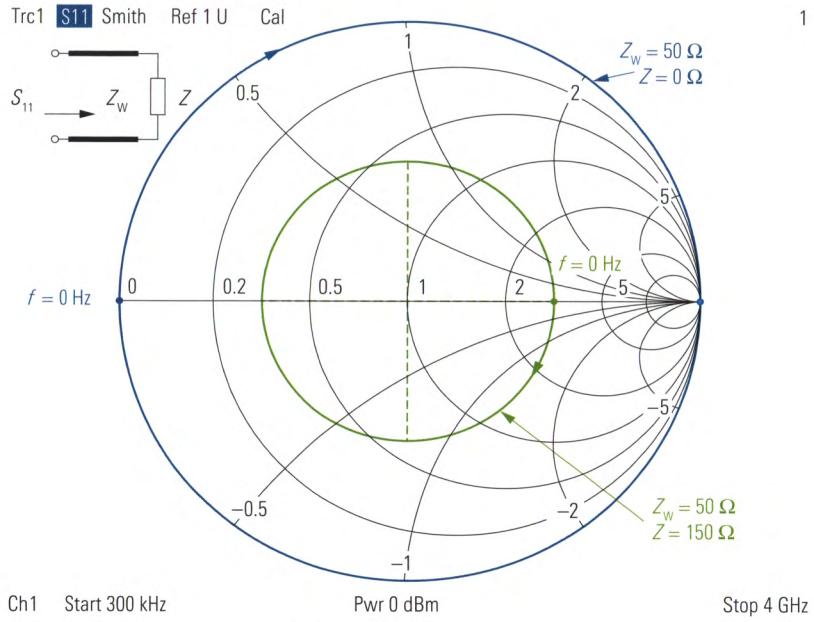


Fig. 4.3.2 Measurement on different transmission line sections (reference impedance 50 Ω).

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 $Z_C = 50\ \Omega$ is terminated with a short $Z = 0\ \Omega$. 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 $z = Z_c/50\ \Omega = 1$. If we terminate the same transmission line with $Z = 150\ \Omega$, 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 towards the point $s_{11} = 0$ (corresponding to $z = 1 + 0\text{j}$).

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 con-

verted reflection coefficient is termed reflection. A typical example is the reflection a_{s11} calculated from the reflection factor s_{11}:

$$a_{s11} = 20\lg|s_{11}|\,\mathrm{dB} \qquad \textbf{(4.3-1)}$$

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.

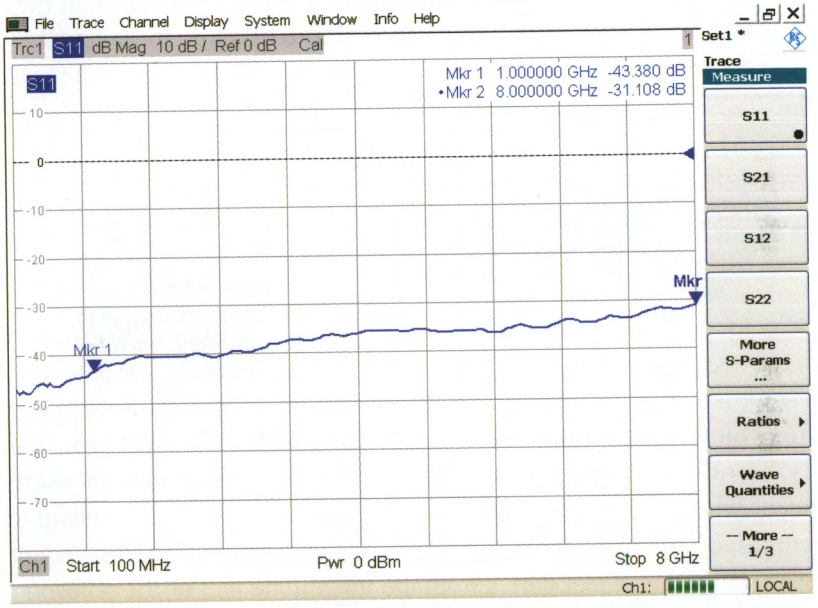


Fig. 4.3.3 Check of the match of a high-performance attenuator pad.

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.

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 $|\Gamma| = 1$, 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 $b = 0$. 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.

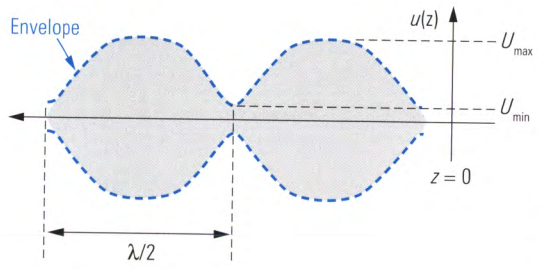


Fig. 4.3.4
Superpostion of the incident and reflected waves.

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

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

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.

Resistance Z in case Z is real and a reference impedance $Z_0 = 50\ \Omega$ 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 2.1.1 *Conversion of characteristic one-port quantities.*

4.4 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.

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

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 highpass 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 $a_{s21} = -\infty$. 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.

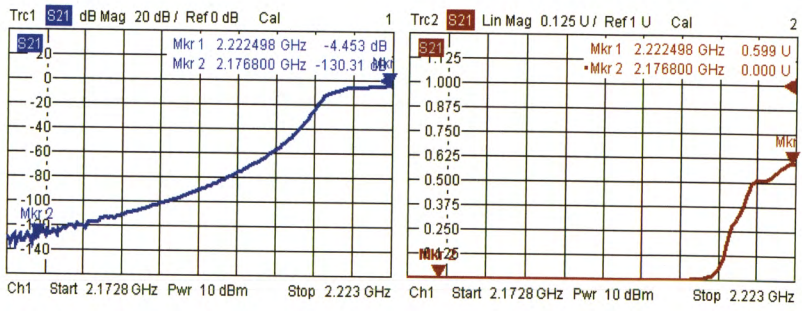


Fig. 4.4.1 Measurement of the transmission of a highpass filter.

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 degrees 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. Fig. 4.6.2 shows a comparison between normal phase display and unwrapped phase.

4.5 Measurement of the group delay

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

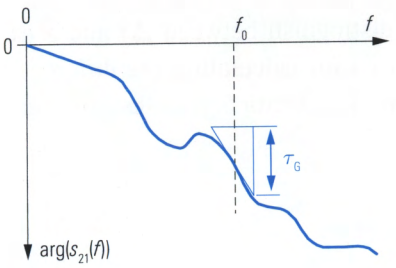


Fig. 4.5.1a
Definition of the group delay.

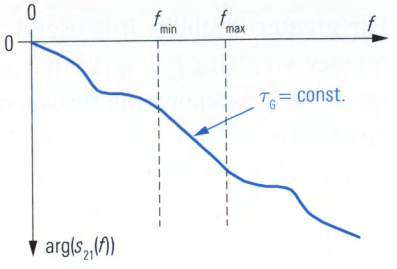


Fig. 4.5.1b
Interval f_{min}, f_{max} where τ_G = const.

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

$$\tau_G(f_0) = -\frac{1}{360°}\frac{d}{df}\arg(s_{21}(f_0)) \qquad \textbf{(4.5-1)}$$

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 $\tau_G(f)$ (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 $f_{min} < f < f_{max}$ to 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 $s_{21}(f) = 0.5 \cdot {f_{CF}}^2/({f_{CF}}^2 - f^2 + jBf)$ which can be directly inserted in to formula (4.5.1) and the derivative (differential quotient $\mathrm{d}/\mathrm{d}f$) can be done by means of 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 $\mathrm{d}/\mathrm{d}f$ must be approximated numerically by a difference quotient.

$$\tau_G(f_0) \approx -\frac{1}{360°} \frac{\arg(s_{21}(f_0 + f_d/2)) - \arg(s_{21}(f_0 - f_d/2))}{f_d} \qquad \textit{(4.5-2)}$$

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

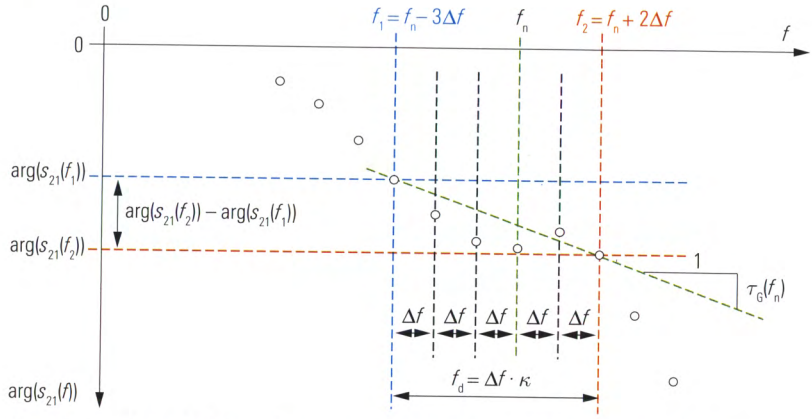


Fig. 4.5.2 Group delay on the network analyzer for aperture stepwidth $\kappa = 5$.

The selected aperture stepwidth $\kappa = f_d / \Delta f$ has an influence on the calculated group delay curve $\tau_G(f)$. 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.

In the following example, a bandpass filter was measured. Fig. 4.5.3 shows the related phase and magnitude curve for the transmission coefficient s_{21}. In Figs. 4.5.4 to 4.5.6, the aperture stepwidth κ was varied. In the present example, an aperture stepwidth of $\kappa = 10$ 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 superim-

posed on the S-parameter and thus reduce the noise on the group delay trace.

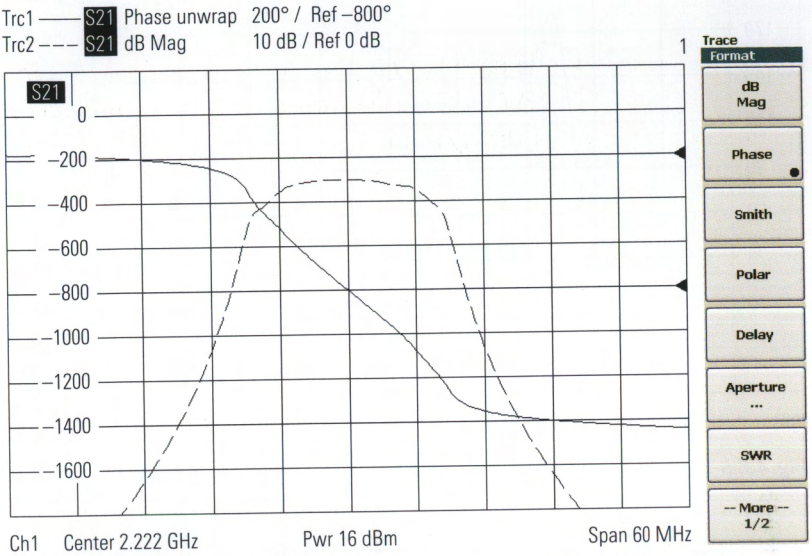


Fig. 4.5.3 Phase and magnitude curve for a bandpass filter.

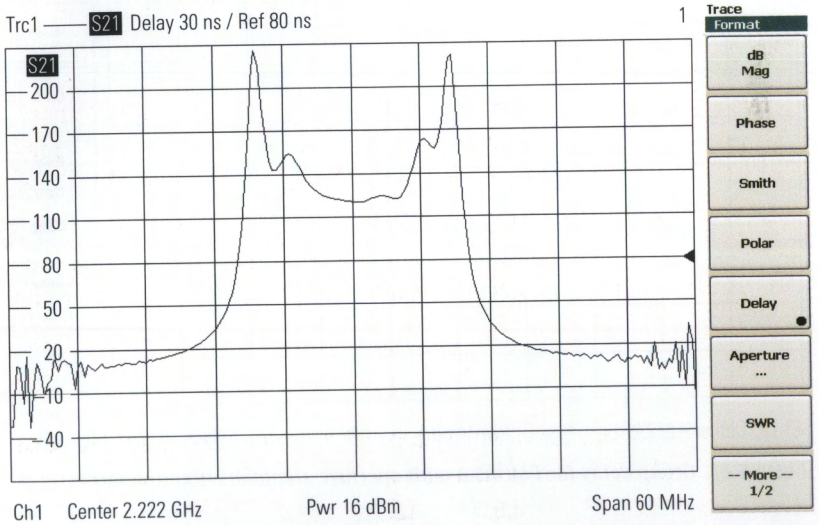


Fig. 4.5.4 Group delay for Fig. 4.5.3 with aperture stepwidth 1 (too small).

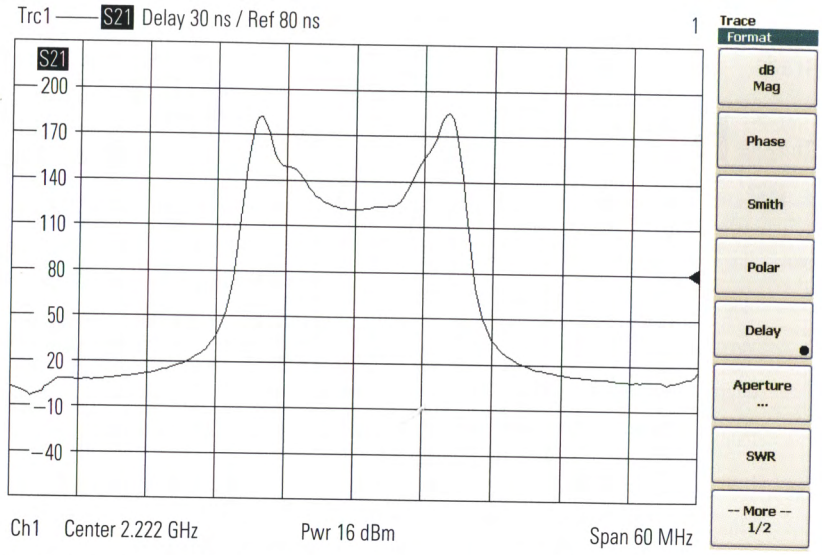


Fig. 4.5.5 Group delay for Fig. 4.5.3 with aperture stepwidth 10 (ideal).

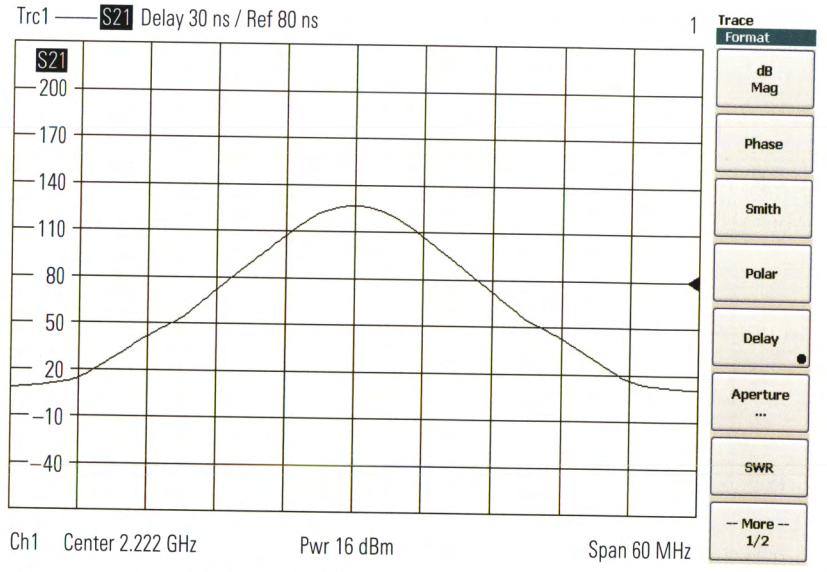


Fig. 4.5.6 Group delay for Fig. 4.5.3 with aperture stepwidth 100 (too large).

4.6 Measurement of the phase delay, auto length

The **phase delay** is a function of the absolute phase at frequency f. It is normally calculated from a transmission coefficient e.g. s_{21}.

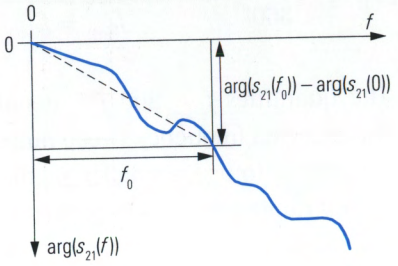


Fig. 4.6.1
Definition of the phase delay.

The phase delay $\tau_P(f)$ is proportional to the phase range covered between 0 Hz and the frequency of interest f divided by the frequency interest. In this computation, the phase at 0 Hz is assumed to be[1] $\arg(s_{21}(0\ \text{Hz})) = 0$.

$$\tau_P(f) = -\frac{\arg(s_{21}(f)) - \arg(s_{21}(0\ \text{Hz}))}{360° f} \qquad \textit{(4.6-1)}$$

In a distortion-free or dispersion-free network, the phase delay remains at a constant value, specific for this network. This value gives the time delay in seconds experienced by any sinusoidal component when passing through the network. Thus the phase delay is strongly related to the length of a transmission lines. Typical transmission lines exhibit a phase delay of some ps up to a few ns for longer lines. The **electrical length** l_{el} of a transmission line can be calculated, for example, from the speed of light in vacuum $c_0 = 2.99792458 \cdot 10^8$ m/s and the phase delay τ_p.

$$l_{el} = c_0 \tau_P \qquad \textit{(4.6-2)}$$

In case of a homogenous, approximately lossless transmission line, we are able to calculate the phase delay from the mechanical length l_{mech} and the relative permittivity ε_r of the dielectric as follows:

$$\tau_P = \frac{l_{mech}\sqrt{\varepsilon_r}}{c_0} \qquad \textit{(4.6-3)}$$

1) This also applies in case a network (e.g. a two-port device) has a transmission coefficient $s_{21} = 0$ for $f = 0$ so that the phase value $\arg(s_{21}(0))$ at $f = 0$ would be, strictly speaking, undefined

Network analyzers typically use the following formula to determine the phase delay. This differs from the definition in formula (4.6-1) in using the start frequency f_{Start} instead of the frequency 0 Hz.

$$\tau_{\mathrm{P}} \approx -\frac{1}{360^\circ} \frac{\arg(s_{21}(f_{\mathrm{Stop}})) - \arg(s_{21}(f_{\mathrm{Start}}))}{f_{\mathrm{Stop}} - f_{\mathrm{Start}}} \qquad \textit{(4.6-4)}$$

The quantities f_{Start} and f_{Stop} denote the start and stop frequencies of the selected frequency sweep range. Some network analyzers also make it possible to use a smaller segment of the trace instead of the entire sweep range (see the lower part of Fig. 4.6.2 with a trace segment from 1 GHz to 2 GHz).

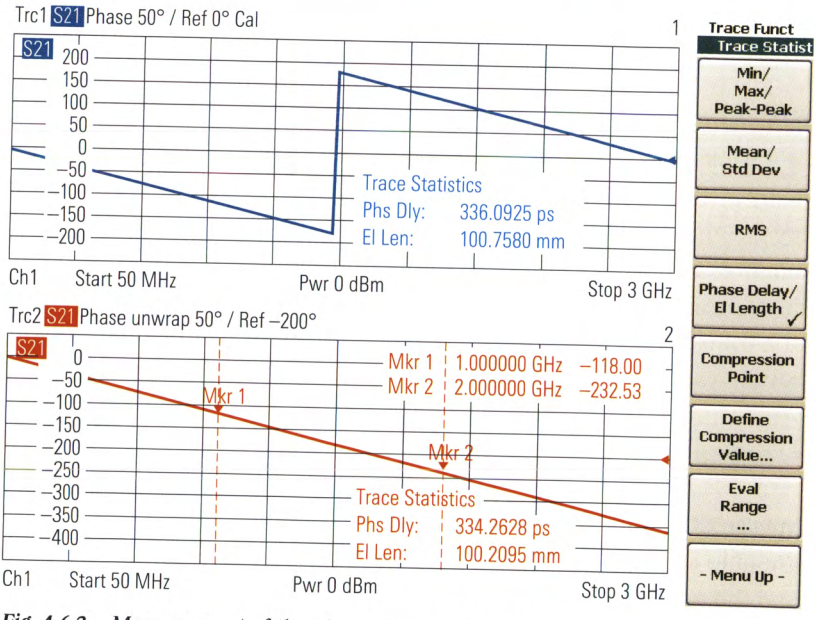


Fig. 4.6.2 Measurement of the phase delay and display with unwrapped phase.

The **auto length** function computes the average electrical length l_{offs} over the entire sweep range. The network analyzer uses this length offset value to correct subsequent measurements. Auto length was activated for the upper trace in Fig. 4.6.3. As a result, the trace shows the deviation from the linear phase. If the length offset is not subtracted off, these deviations cannot be recognized in most cases (see the lower trace in Fig. 4.6.3).

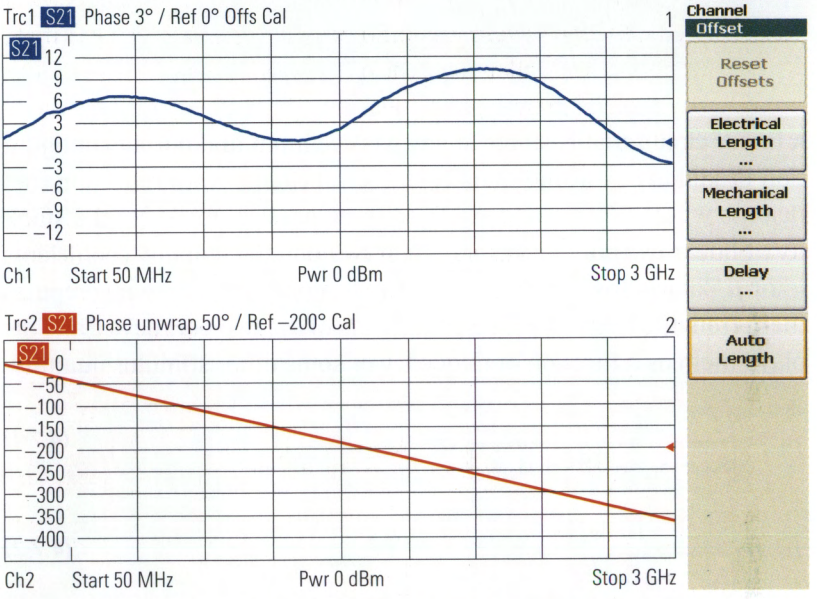


Fig. 4.6.3 Comparison of displays with and without auto length.

4.7 Measurement of the stability

Instability is a basic risk with any amplifier. In an unstable state, the amplifier acts as an oscillator and will produce an (undesired) output signal even if there is no input signal present. In the rest of the frequency range, the oscillation will reduce the gain in most cases.



For amplifiers with large power reserves it is important to ensure prior to the measurement that even the maximum output power cannot do damage to the network analyzer. When the amplifier input power is reduced, its output level will normally decrease, but in the case of instability this drop might be ineffective. The amplifier will act as an oscillator in this case and can produce its maximum output power.

A linear circuit that is stable in any passive input and output configuration is said to be **unconditionally stable.** All circuits with reflection coefficients where $|\Gamma| < 1$ are considered to be passive. To assess the stability of linear two-port device, we can draw upon the stability coefficients μ_1 or μ_2 known from the literature [ES92]. A network analyzer computes them from the four complex S-parameters of the two-port device and displays them as a function of frequency or some other stimulus quantity.

$$\mu_1 = \frac{1 - |s_{11}|^2}{|s_{22} - s_{11}^* \det(\mathbf{S})| + |s_{21} s_{12}|} \qquad \textit{(4.7-1)}$$

$$\mu_2 = \frac{1 - |s_{22}|^2}{|s_{11} - s_{22}^* \det(\mathbf{S})| + |s_{21} s_{12}|} \qquad \textit{(4.7-2)}$$

A necessary and sufficient criterion for unconditional stability of a linear two-port device is the condition $\mu_1 > 1$. As an alternative, it is also possible to check that $\mu_2 > 1$. Many network analyzers only provide the stability coefficient k described by Rollet [Ro62] for use in investigating stability.

$$k = \frac{1 + |\det(\mathbf{S})|^2 - |s_{11}|^2 - |s_{22}|^2}{2|s_{12} s_{21}|} \qquad \textit{(4.7-3)}$$

In the technical literature, this coefficient is much better known than the μ coefficients. However, the condition $k > 1$ is necessary but not sufficient for unconditional stability. We must add one of the following additional conditions to obtain a necessary and sufficient criterion.

$$\left|s_{12}s_{21}\right| < 1 - \left|s_{11}\right|^2 \quad \text{or} \quad \left|s_{12}s_{21}\right| < 1 - \left|s_{22}\right|^2 \qquad \textbf{(4.7-4)}$$

The frequency range for the stability analysis should be selected to cover the entire frequency range of interest. This range extends up to the frequency at which the gain of the amplifier is equal to <1 or <0 dB (transit frequency).

All of the stability coefficients described here are real quantities. They are typically displayed on a linear scale (format: Real). Fig. 4.7.1 below represents a stability analysis using the stability coefficient μ_1. In the frequency range between markers 1 and 2, the amplifier is not unconditionally stable. In most cases it is possible to find some specific input and output configurations for which the amplifier exhibits stable behavior. For these specific Γ values attached to the input and output, the amplifier is said to be **conditionally stable**. But in practice, we typically prefer an unconditionally stable amplifier.

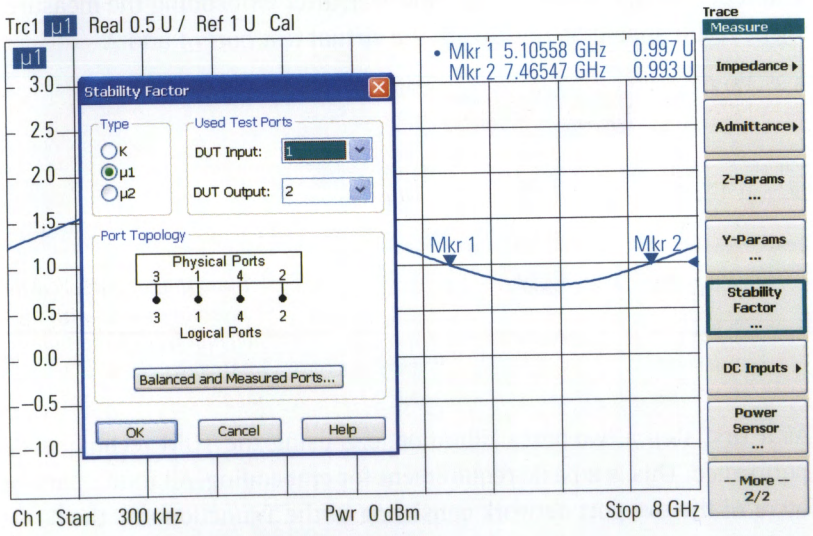


Fig. 4.7.1 Example of a stability analysis.

4.8 Measurement with embedding

With **embedding**, a component is measured as if it were embedded in a defined (but not physically present) circuit environment. A circuit environment of this type is known as a **virtual transformation network**. Typical examples are a matching network for an amplifier or the design of an equalizer for a filter (both determined e.g. using a CAD program). All that is physically present in these examples is the amplifier or the filter.

The figure below shows the circuit for a T-attenuator pad. It has an attenuation of 7.7 dB and all ports are matched $s_{11} = s_{22} = 0$.

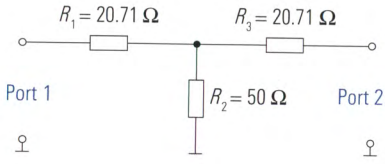


Fig. 4.8.1
T-attenuator pad

For illustration purposes, the resistors R_1 and R_3 are realized through embedding. The DUT which is physically present consists only of a T-junction and a 50 Ω resistor (see Fig. 4.8.2). The connectors of the T-junction form the inner ports. However, after embedding the measurement result should be referred to the virtual test ports 1 and 2.

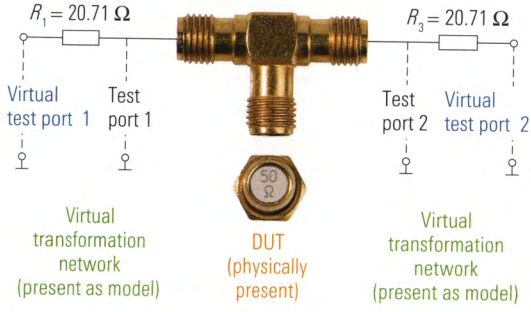


Fig. 4.8.2
T-junction and R_2 prior to assembly, along with the virtual transformation networks.

As a first step a two-port calibration, e.g. using the TOM technique, is performed. This is a basic requirement for embedding. All four S-parameters of the two-port network consisting of the T-junction and the 50 Ω resistor are measured (see Fig. 4.8.3). An attenuation of 3.5 dB is measured. Since the resistors R_1 and R_3 are still missing, the incomplete attenuator pad exhibits a relatively worse return loss of 9.5 dB.

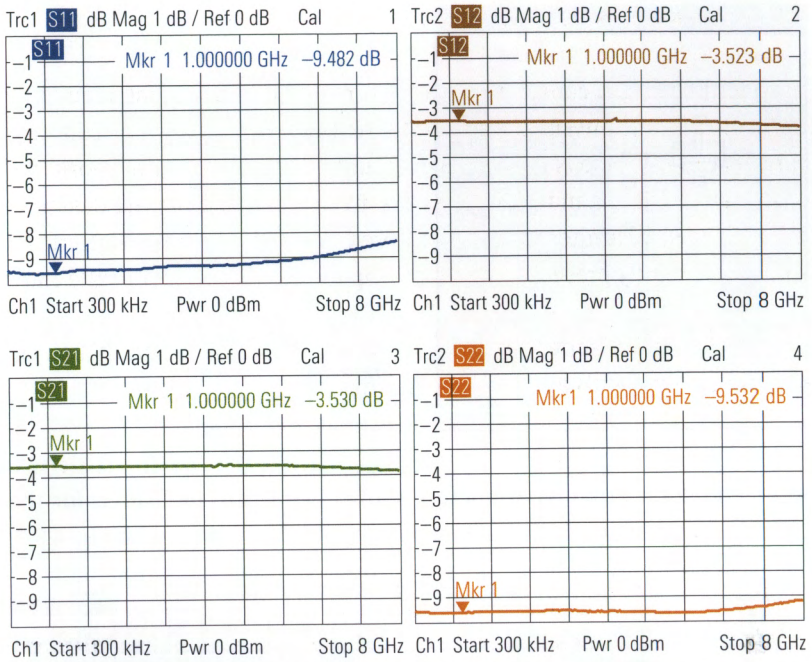


Fig. 4.8.3 Physically present DUT.

Embedding requires complete characterization of the virtual transformation networks that are involved. The network analyzer offers the following options:

- S-parameters for a measured network, e.g. in Touchstone® format as an *.s2p or *.s4p file. The characterization must include all of the S-parameters of the virtual transformation network (two-port characterization, or four-port characterization in case of balanced networks). The necessary frequency range must be covered by the file.
- Selection of a simple circuit model and parameterization (assignment of values to the individual components).

Our example is particularly suitable for use of a parameterized circuit model (see Fig. 4.8.4). The embedding series resistor is set to 20.71 Ω. All the other embedding components are set to neutral values ($C1 = 0$ F, $L1 = 0$ H, $R2 = 10$ MΩ). Embedding can be switched on and off using the "Embed DUT" checkbox. The same input is made for the second test port.

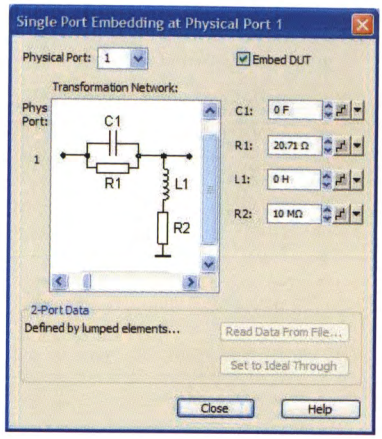


Fig. 4.8.4
Virtual transformation network at test port 1.

After embedding, the S-parameters shown below are measured for the overall circuit between virtual test ports 1 and 2 (compare with Fig. 4.8.2).

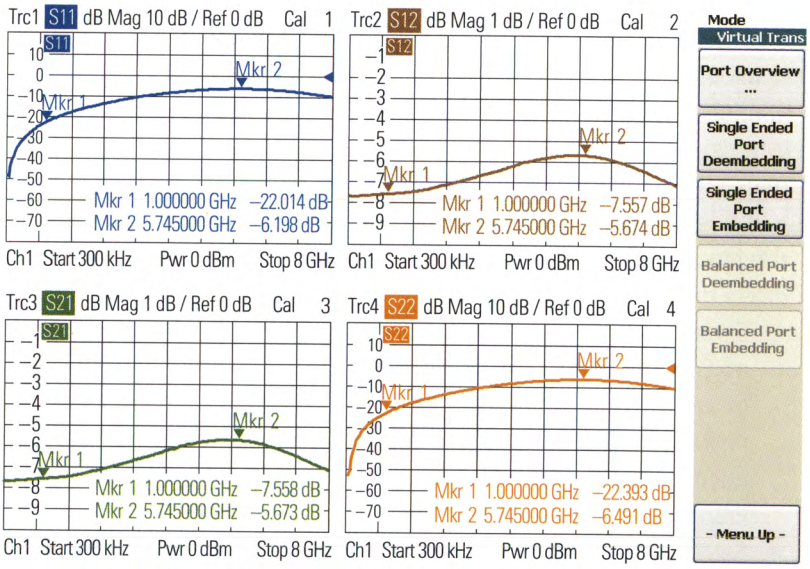


Fig. 4.8.5 *Measurement result with embedding at test port 1 and 2.*

For frequencies up to about 1 GHz, the measurement result meets the expectations. By default the series resistors R_1 and R_3 are assumed to be located in the reference plane of the calibrated network analyzer. As can be concluded from Fig. 4.8.6, this plane lies within the connector. Between the resistors R_1 and R_3 and the node point of the T-junction, there is a transmission line section l_1 or l_2 respectively. Both lengths are shown in Fig. 4.8.6 below. The line section with length l_2 is terminated

with the series circuit of R_3 and the virtual test port 2 (see Fig. 4.8.2). The impedance of this series circuit is equal to about 70 Ω. In other words, this line section is not terminated with its characteristic impedance $Z_c = 50\ \Omega$. This means it causes a frequency-dependent transformation that has an effect on all four S-parameters in Fig. 4.8.5. The same applies to the other line section that is fed with source impedance $R_1 + 50\ \Omega \approx 70\ \Omega$.

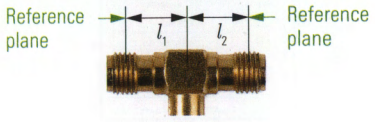


Fig. 4.8.6
Lengths l_1 and l_2 on the T-junction.

By selecting a suitable length offset at test ports 1 and 2, it is possible to shift the reference planes into the T-junction. In the present case, we can determine this length offset by the following conclusions: At the frequency 5.745 GHz (corresponding to a wavelength of $\lambda = 52.22$ mm) the parameter s_{11} has its maximum. At this wavelength, the previously calculated 70 Ω mismatch is transformed by the line section l_2 into a minimum. The actual transformation is based on the assumption of an electrical length $l_2 = \lambda/4$. The electrical length required for compensation is thus equal to $l_2 = 52.22\ \text{mm}/4 = 13.05$ mm. Due to the symmetry, it is suitable for both test ports ($l_1 = l_2$). The measurement result from Fig. 4.8.7 meets our expectations.

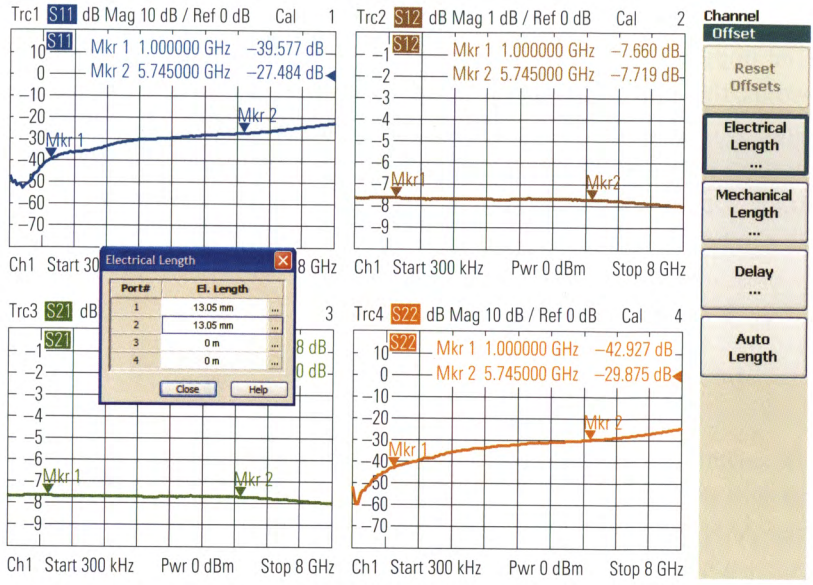


Fig. 4.8.7 Measurement result with embedding and length offset.

Embedding with simple circuit models or with the S-parameters (*.s2p-, *.s4p files) can be combined with additional mechanisms. Possibilities include:

- Taking into account **virtual impedance transformers** or changing the reference impedance to a value that diverges from the instruments physical value which is typically $Z_0 = 50\ \Omega$ or $Z_0 = 75\ \Omega$. This procedure is known as renormalization of the S-parameters.
- Adding homogenous transmission line sections by a **length offset**. A negative length offset $l_{\text{offs}} < 0$ adds a virtual line section to the circuit, while a positive length offset compensates a line section that is physically present. The latter will be covered in the next section (deembedding).

When using these additional mechanisms, it is important to take into account the order of the individual steps in the calculation chain of the network analyzer. This is shown in Fig. 4.8.8 for the instrument family that is employed in this book. The techniques of deembedding and balance-unbalance conversion are included for the sake of completeness and are covered in sections 4.9 and 4.10, respectively.

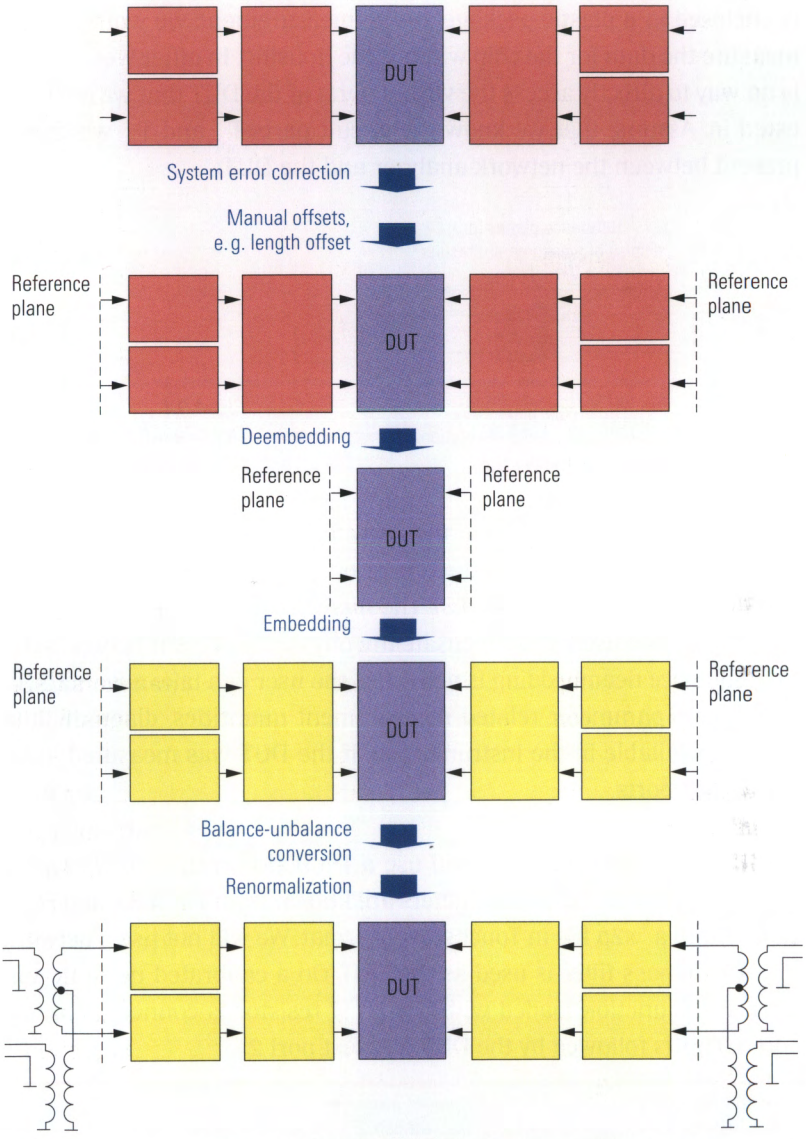


Fig. 4.8.8 Order and effect of the different steps in the calculation chain.

4.9 Measurement with deembedding

The intension of **deembedding** is, to measure a DUT that is embedded in to a physical circuit environment whereas this environment has to be removed mathematically from the measurement result. For example, the DUT might be a **microwave monolithic integrated circuit** which

is enclosed in a plastic package. In the measurement, we would like to measure the data for the chip without the housing. In other words, there is no way to directly access the virtual ports of the DUT that we are interested in. Assume that we know the circuit parts T_{p1} and T_{p2} which are present between the network analyzer and the DUT.

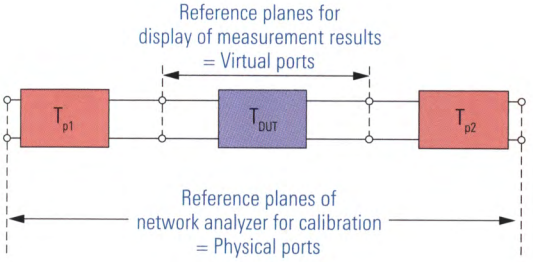


Fig. 4.9.1
Deembedding of a two-port network.

A complete characterization of the networks T_{p1} and T_{p2} is provided to the network analyzer. The characterization methods used here are familiar from embedding (section 4.8). The analyzer computes a transformation that can be used to compensate the physically present networks T_{p1} and T_{p2}. Once deembedding is activated, the user can take advantage of all of the S-parameter related measurement quantities, diagrams and formats available in the instrument as if the DUT was measured at its virtual test ports.

For illustration purposes, we will use a T-junction with a 50 Ω resistor as our network T_{p1}. Its S-parameters are known from Fig. 4.8.3 and were exported to a *.s2p file in Touchstone® format. We will not use a network T_{p2}. A bandpass filter is used as the DUT. On a calibrated network analyzer, the T-junction loaded with the 50 Ω resistor is connected to test port 1. This is followed by the DUT and test port 2.

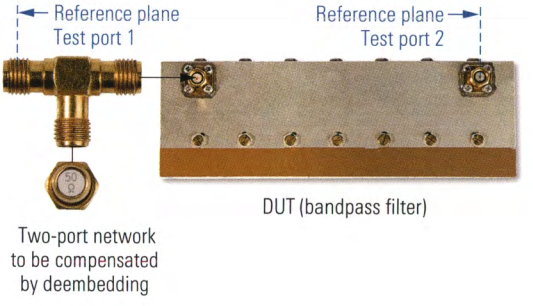


Fig. 4.9.2
Test setup prior to assembly.

Here, deembedding is necessary at test port 1 only. The characterization of the network T_{p1} is assigned as shown in Fig. 4.9.3. For this purpose, the corresponding *.s2p file is loaded. In addition, deembedding at test port 1 can be switched on and off using the "Deembed DUT" checkbox.

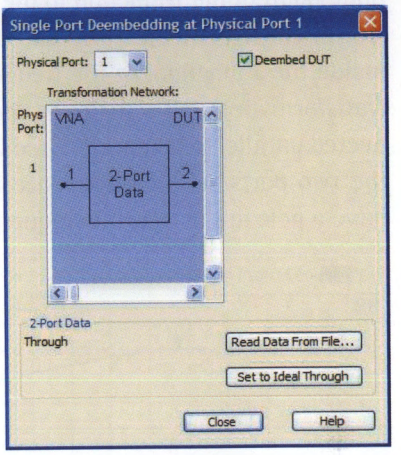


Fig. 4.9.3
Virtual transformation network at test port 1.

As a result of deembedding, the S-parameters of the bandpass filter are displayed as if the network T_{p1} was not present.

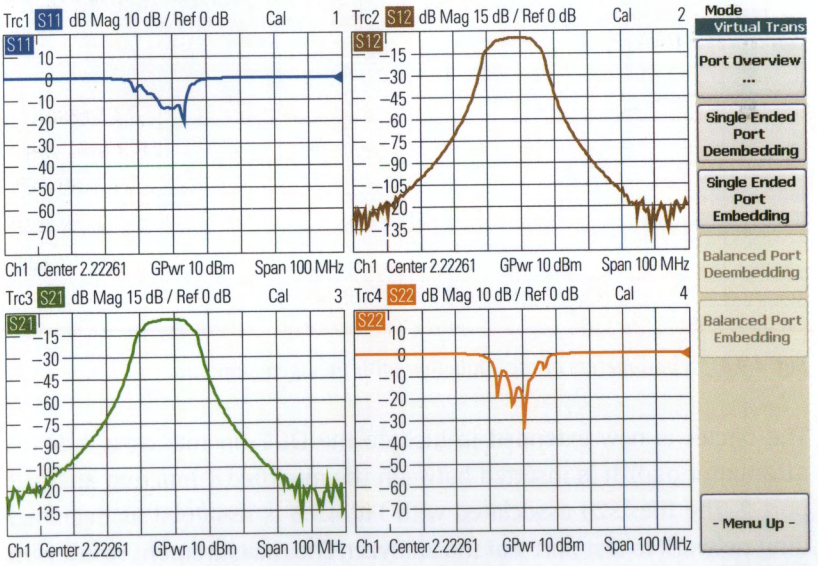


Fig. 4.9.4 Measurement result with deembedding.

We will now modify the T-junction in order to illustrate the limits of deembedding. Instead of the 50 Ω resistor, we connect an open stan-

dard (e.g. from a calibration kit). The T-junction which is modified in this manner is first measured alone as a two-port device. The associated S-parameters are saved to a *.s2p file. They can be seen in Fig. 4.9.5. We can see a pole at 4.22 GHz. At this frequency, the line leading from the node point of the T-junction to the open acts as a $\lambda/4$ stub. The relevant length consists of the electrical length of the open and a line section inside the T-junction. At the frequency 4.22 GHz, the line acts as a $\lambda/4$-transformation and transforms the open into a short. This short is connected parallel to test ports 1 and 2 and prevents transmission between the two ports. As a consequence forward and backward transmission have a pole at this specific frequency (see also s_{12} and s_{21} in Fig. 4.9.5).

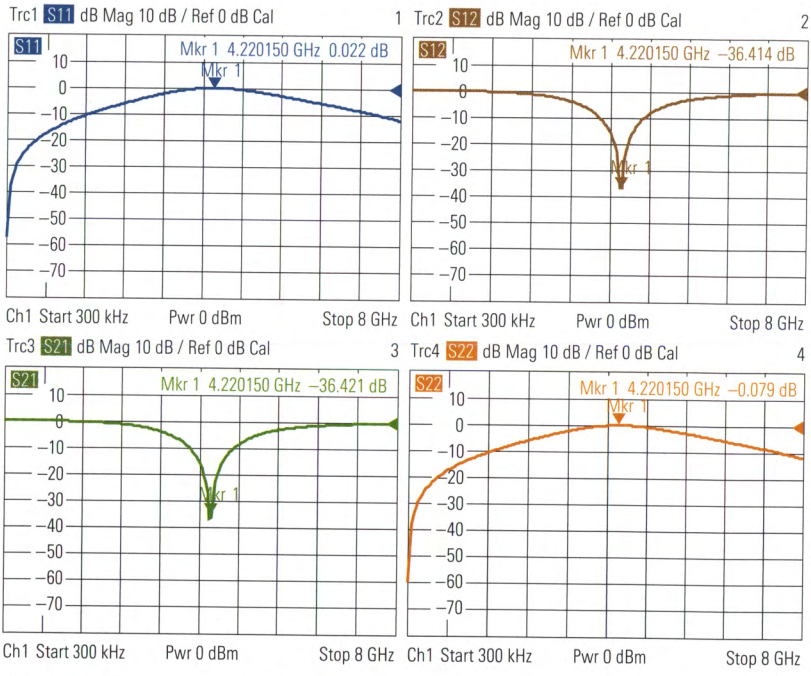


Fig. 4.9.5 S-parameters of the T-junction with the open standard.

The circuit is now extended to include the DUT (in this case, a 10 dB attenuator pad). It is inserted between the modified T-junction and test port 2. The file *.s2p associated with Fig. 4.9.5 is assigned as deembedding network to test port 1 of the analyzer. The measurement result with deembedding exhibits a significant deviation at the frequency 4.22 GHz (see Fig. 4.9.6). As described previously, at this frequency transmission through the network T_{p1} is not possible. The measured values a_1 and b_1 at test port 1 therefore do not contain any signal components that

would allow us to draw a conclusion about the reflection of the DUT. We can perform a similar analysis regarding the transmission at 4.22 GHz. In the vicinity of this frequency, the network analyzer is not capable of computing the S-parameters for the DUT 10 dB attenuator pad from the measured values.

Deembedding is a fascinating tool, but it has physical limits. For example, if the network to be compensated has poles, compensation will not be possible there. If the network to be compensated has high attenuation, this will degrade the effective directivity on the physical port (see section 2.4.1). If the transformation network is known precisely, theoretically no degradation of the effective system data on the virtual test port will occur, but the influence of drift effects will increase. Added to this will be a degradation of the signal-to-noise ratio due to the low levels at the physical port. The noise in the trace will then inevitably increase. Clearly, deembedding cannot solve all your measurement problems.

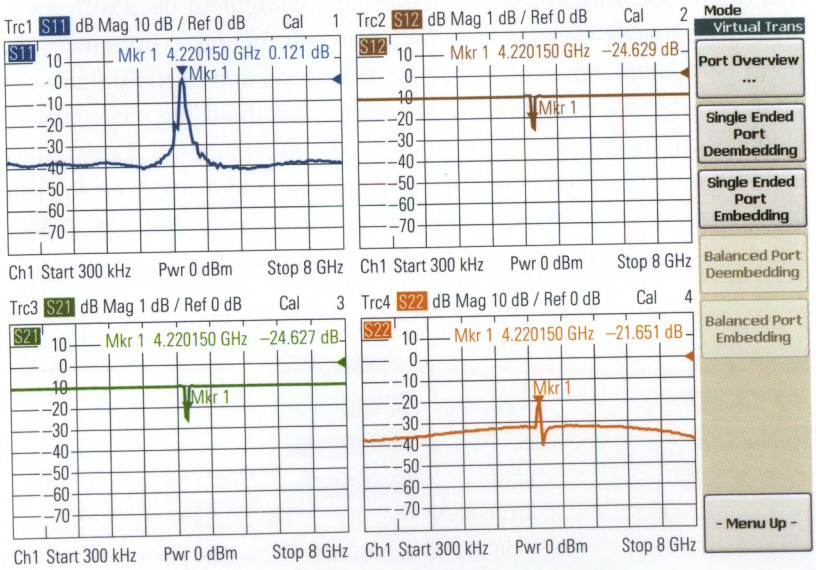


Fig. 4.9.6 Trace for the 10 dB attenuator pad with deembedding.

4.10 Measurement of balanced lines

When transmitting signals over transmission lines, we would like to reduce external interference. Nevertheless, external unwanted signals can be coupled in even when using closed shielding, e.g. due to the finite conductivity of the ground conductor. If we transmit a signal over a pair of **symmetrical coupled conductors**, we can expect the interference to be coupled into both lines equally. On the other hand, if we feed the wanted signal into one conductor with a 0° phase shift and into the other conductor with a 180° phase shift, it is very easy to distinguish the arising differential signal from the common-mode interferences. This configuration is known as a **balanced**[1] **transmission line**. The ideal configuration involves a **twisted pair transmission line** with no ground reference. In the real world, however, a ground is usually present in the vicinity of the line so that besides the desired **differential-mode** signal also, a **common-mode** signal can also propagate along the transmission line. What is undesired is a mode conversion between common and differential mode since clear distinction between the wanted signal and the interference present in the form of a common-mode signal is no longer possible. When assessing transmission lines and equipment used with them such as connector systems, transmitters, receivers and so on, a measurement is required which allows separate observation of the differential and common modes and provides information about the mode conversion. For characterization purposes, the common-mode wave quantities a_c and b_c and the differential wave quantities a_d and b_d are used. In each case, two physical ports are combined to form a balanced logical port. Fig. 4.10.1 shows an example of a balanced two-port device.

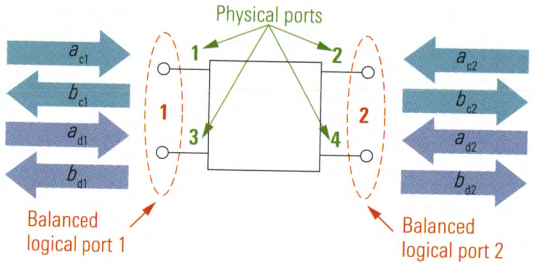


Fig. 4.10.1
Comparison between physical and balanced logical ports.

1) This is also referred to as "differential" in technical literature.

Normally a network analyzer only has **single-ended**[1] test ports. This means that it measures what is known as the **single-ended S-parameters**[2] which were defined in section 1.2. Using suitably configured transformers, these test ports can be adapted to a balanced DUT. Since physical transformers have undesired properties (e.g. a very limited bandwidth) the analyzer simulates ideal transformers, using appropriate circuit models. An arithmetic process similar to embedding (see section 4.8) integrates these ideal transformers into the measurement data processing chain. This process is known as **modal decomposition**. A big advantage of this process is that no special balanced hardware is required for the measurement. Furthermore it is possible to use conventional calibration techniques and calibration standards. A modern network analyzer can automatically perform the modal decomposition and the user only needs to specify the desired test ports as balanced ports. The measurement result is a so-called **mixed-mode S-matrix**. It characterizes the relationships between the balanced wave quantities. For a balanced two-port network, it has a total of 16 entries. Their meaning is clarified using the following measurement examples.

Example 1 The transmission line configuration for the first example consists of two conductors with exactly the same length (electrical length 18.17 mm). The network analyzer used provides a measurement wizard. In its first step, the test port configuration for the balanced measurement is selected (see following Fig.).

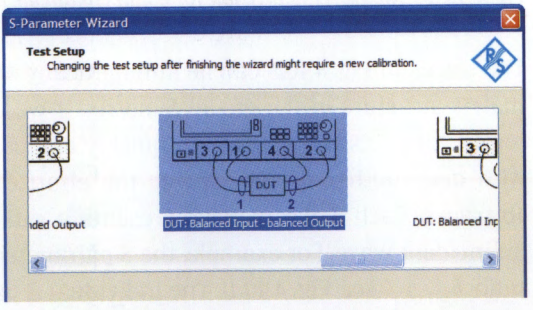


Fig. 4.10.2
Test port configuration for the balanced measurement.

The remaining dialogs in the measurement wizard are used to configure the frequency range, the IF bandwidth and the test port output power. Here, it is okay to simply accept the default settings. Once the measure-

1) In some cases also referred to as unbalanced in technical literature.

2) In technical literature also referred to as "nodal S-parameters".

ment wizard has been configured, the calibration is performed and the DUT is connected. The network analyzer displays the following measurement result:

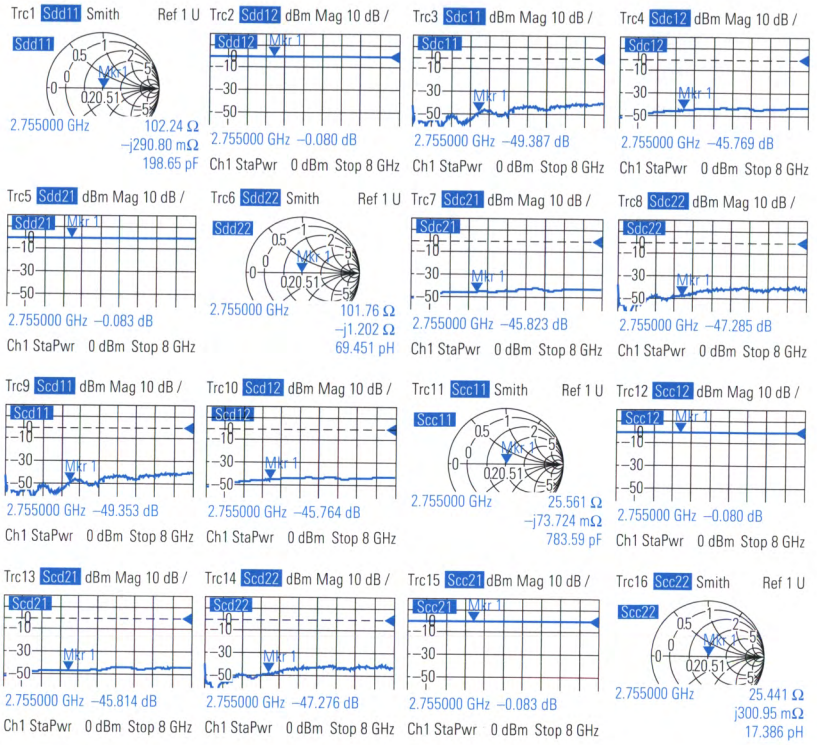


Fig. 4.10.3 Measurement result for equal conductor length (example 1).

The traces in Fig. 4.10.3 can be automatically arranged according to the matrix structure of the modal S-matrix. If we use this structure consistently, it makes it very easy to quickly compare measurement results with one another, thereby saving the effort of searching through the 16 traces. Each S-parameter represents a ratio of a reflected wave to an incident wave. For example, the S-parameter s_{dc21} corresponds to the ratio $b_{\mathrm{d2}}/a_{\mathrm{c1}}$ (see Fig. 4.10.1). The individual S-parameters from the measurement can be interpreted as follows:

- $s_{\mathrm{dd11}} \approx s_{\mathrm{dd22}} \approx s_{\mathrm{dc11}} \approx s_{\mathrm{dc22}} \approx s_{\mathrm{cd11}} \approx s_{\mathrm{cd22}} \approx s_{\mathrm{cc11}} \approx s_{\mathrm{cc22}} \approx 0$
  The transmission line is terminated almost ideally. No power is reflected within a mode and also not between the modes.
- $s_{\mathrm{dd12}} \approx s_{\mathrm{dd21}} \approx s_{\mathrm{cc12}} \approx s_{\mathrm{cc21}} \approx 1$
  The line transmits the common mode and differential mode signals approximately lossless between the two balanced ports.

- $s_{dc12} \approx s_{dc21} \approx s_{cd12} \approx s_{cd21} \approx 0$
  No mode conversion occurs during transmission.

Example 2 In the second example of a transmission line configuration, one of the conductors has been extended. This means that the conductors have electrical lengths of 18.17 mm and 72.80 mm, respectively. As a result of this asymmetry, the mixed-mode S-parameters are changed.

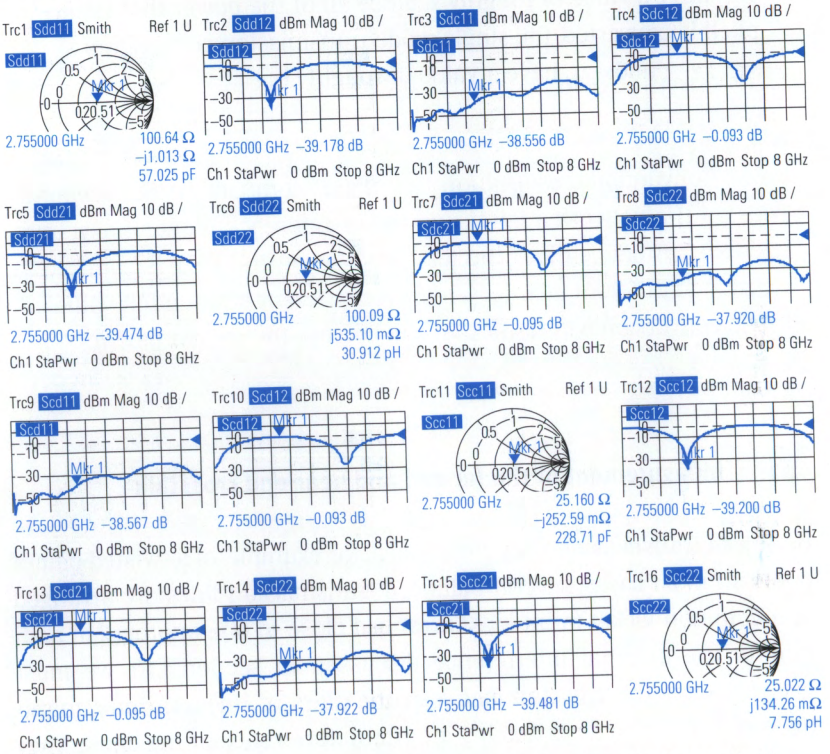


Fig. 4.10.4 Measurement results for different conductors lengths (example 2).

Some of the traces in Fig. 4.10.4 have a maximum or minimum at the frequency $f_0 = 2.755$ GHz. The behavior of the DUT can be read off from the S-parameters for the measurement:

- $s_{dd11} \approx s_{dd22} \approx s_{dc11} \approx s_{dc22} \approx s_{cd11} \approx s_{cd22} \approx s_{cc11} \approx s_{cc22} \approx 0$
  As result of the change in length, the characteristic impedances of the transmission line configuration were not modified in comparison to example 1. Accordingly approximate matching is still obtained.
- s_{dd12}, s_{dd21}, s_{cc12} and s_{cc21} have a value of roughly 0 at f_0, and

- s_{dc12}, s_{dc21}, s_{cd12} and s_{cd21} have a value of roughly 1 at f_0.
  At the frequency f_0, the transmission line extension results in a phase shift of 180° between the two conductor signals. The supplied differential-mode signals are fully converted during transmission into common-mode signals. In comparison, supplied common-mode signals are fully converted into differential mode signals. Accordingly, the mixed-mode transmission coefficients s_{dc12}, s_{dc21}, s_{cd12} and s_{cd21} have a value of roughly 1. Since all of the power that is supplied is converted, the single-mode transmission coefficients s_{dd12}, s_{dd21}, s_{cc12} and s_{cc21} have a value of roughly 0.



Conventional calibration techniques compensate the asymmetries on the network analyzer up to the reference plane. If asymmetries are introduced into the test setup when the DUT is connected, e.g. due to different adapter lengths or different attenuation values, a mode conversion will result that can corrupt the measurement.

4.11 Measurement of the far-end and near-end crosstalk

Balanced transmission lines can consist, for example, of a twisted-pair of wires. In most cases, several balanced transmission lines are arranged in a common jacket (or at least arranged close to one another). In addition to the external interference discussed in section 4.10, we must also consider the crosstalk between the pairs of lines. Various quantities are defined to describe this internal interference of the cable. They are based on true power relationships and refer exclusively to differential mode. When analyzing the crosstalk, we must select two conductor pairs from the cable and terminate all their passive ends in a reflection free manner. For this purpose we must consider both propagation modes present on the line. The following figure shows three implementations for this reflection free termination. Z_{Cd} refers to the differential-mode characteristic impedance and Z_{Cc} to the common-mode characteristic impedance of the transmission line. Typical values are, for example, $Z_{Cd} = 100\ \Omega$ and $Z_{Cc} = 25\ \Omega$.

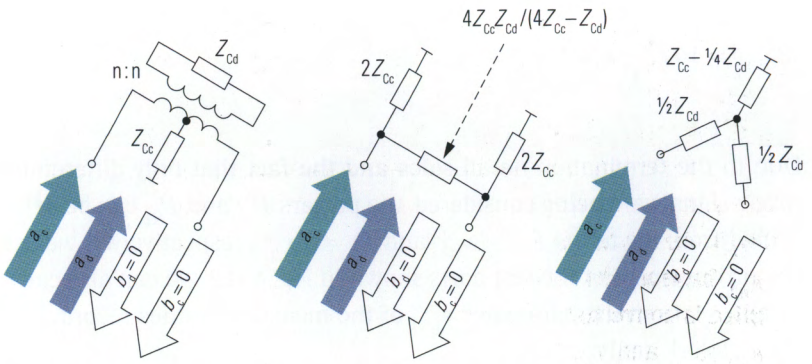


Fig. 4.11.1 Design of the reflection free termination, to be used for figures 4.11.2 to 4.11.4.

The source that is used is also matched to both modes of the transmission line and excites only the differential mode of one conductor pair. The balanced logical ports of a network analyzer can be used as the wave source and the receiver.

When measuring the **near-end crosstalk loss** d_{NEXT} (NEXT), the wave source and the receiver are connected to the same end of the cable.

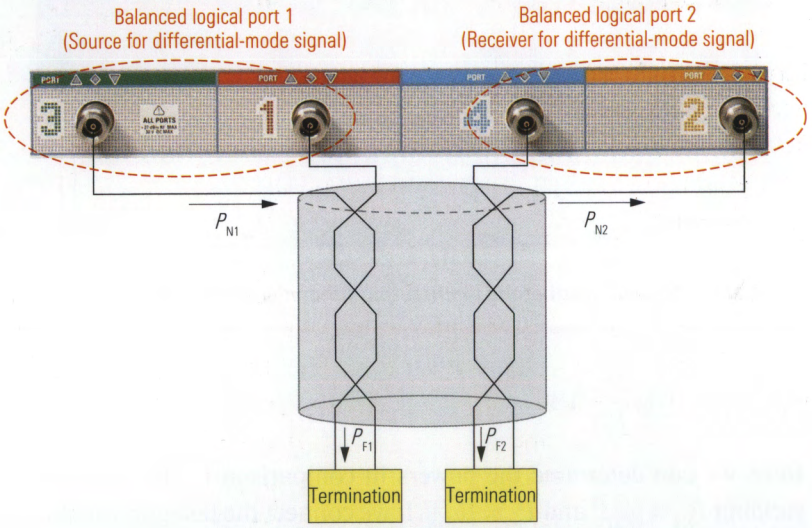


Fig. 4.11.2 Measurement of near-end crosstalk loss.

The source supplies a differential-mode wave with the available power P_{N1}. The receiver detects the cross-coupled power P_{N2} of the signal in differential mode.

$$d_{\mathrm{NEXT}} = 10\lg\frac{P_{\mathrm{N1}}}{P_{\mathrm{N2}}}\,\mathrm{dB} \qquad \textit{(4.11-1)}$$

Due to the termination on all sides and the fact that only differential-mode signals are being considered, the powers P_{N1} and P_{N2} can be determined from the terms $P_{\mathrm{N1}} = |\mathrm{a}_{\mathrm{d1}}|^2$ and $P_{\mathrm{N2}} = |\mathrm{b}_{\mathrm{d2}}|^2$, respectively. If we use the arrangement of the test ports shown in Fig. 4.11.2, we can also calculate d_{NEXT} from the parameter s_{dd21} of the measured values recorded by the network analyzer.

$$d_{\mathrm{NEXT}} = -20\lg|s_{\mathrm{dd\,21}}|\,\mathrm{dB} \qquad \textit{(4.11-2)}$$

The **input/output far-end crosstalk loss** d_{IOFEXT} (IOFEXT) is defined for a similar test setup. The only difference is, that stimulation and detection is made on the opposite ends of the cable. The differential-mode stimulus power is given by P_{N1}. The receiver detects the power P_{F2} (see Fig. 4.11.3).

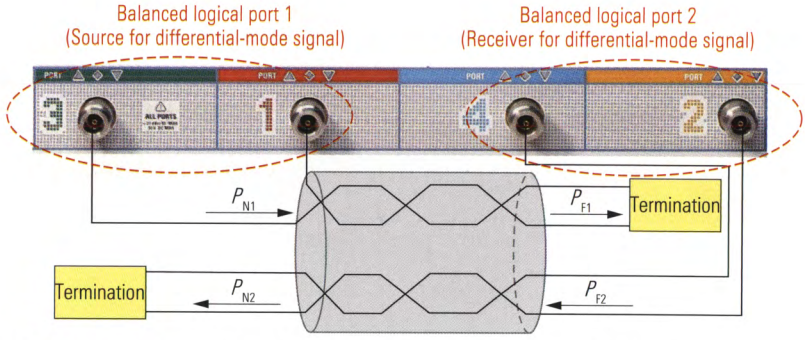


Fig. 4.11.3 Measurement of the input/output far-end crosstalk loss.

$$d_{\mathrm{IOFEXT}} = 10\lg\frac{P_{\mathrm{N1}}}{P_{\mathrm{F\,2}}}\,\mathrm{dB} \qquad \textit{(4.11-3)}$$

Here, we can determine the powers in comparison to the case before, yielding $P_{\mathrm{N1}} = |a_{\mathrm{d1}}|^2$ and $P_{\mathrm{F2}} = |b_{\mathrm{d2}}|^2$. If we connect the test ports as shown in Fig. 4.11.3, then d_{IOFEXT} can also be determined from the parameter s_{dd21} of the measured values recorded by the network analyzer.

$$d_{\mathrm{IOFEXT}} = -20\lg|s_{\mathrm{dd\,21}}|\,\mathrm{dB} \qquad \textit{(4.11-4)}$$

In the case of the **equal-level far-end crosstalk loss** d_{ELFEXT} (ELFEXT), the both powers P_{F1} and P_{F2} are measured on the end of the cable that is opposite to the stimulation end. The power P_{F2} refers to the power that is coupled in as an interference, as in the previous case.

$$d_{\text{ELFEXT}} = 10\lg\frac{P_{\text{F1}}}{P_{\text{F2}}}\,\text{dB} \qquad \textit{(4.11-5)}$$

In principle the power levels P_{F1} and P_{F2} can be determined like before using the appropriate wave quantities. Since the test setup would require a total of six physical test ports (a balanced source and two balanced receivers), we will usually measure d_{ELFEXT} indirectly. To do this, we need to know the **cable loss** d_{TRANS}, which can be measured using the following test setup:

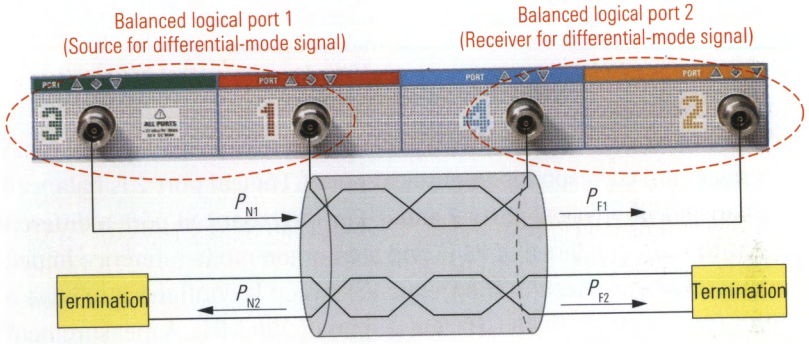


Fig. 4.11.4 Measurement of the cable loss.

$$d_{\text{TRANS}} = 10\lg\frac{P_{\text{N1}}}{P_{\text{F1}}}\,\text{dB} \qquad \textit{(4.11-6)}$$

If we connect the test ports as shown in Fig. 4.11.4, d_{TRANS} can also be calculated from the parameter s_{dd21} of the measured values recorded by the network analyzer.

$$d_{\text{TRANS}} = -20\lg\left|s_{\text{dd}\,21}\right|\text{dB} \qquad \textit{(4.11-7)}$$

The power P_{F2} is caused by the power P_{N1} due to cross-coupling mechanisms. On the other hand, the power P_{F1} is caused by the power P_{N1} due to the cable loss d_{TRANS}. Accordingly, we have the following relationship between the far-end crosstalk loss quantities:

$$d_{\text{ELFEXT}} = d_{\text{IOFEXT}} - d_{\text{TRANS}} \qquad \textit{(4.11-8)}$$

4.12 Filter with balanced and unbalanced port, imbalance and common-mode rejection

In this example, we are interested in a surface acoustic wave filter (SAW-filter) with single-ended port 1 and balanced port 2. To check the functionality of the filter, it must be embedded in a specific matching circuit that is realized using embedding. In theory, we could also implement the matching circuit physically using suitable components. However, by using embedding we are independent of component tolerances in the matching circuitry. This is advantageous, for example, if we need to compare measurement results from several test stations.

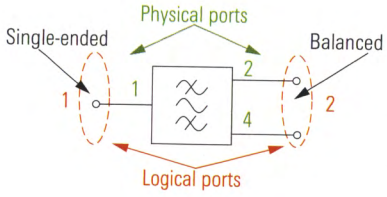


Fig. 4.12.1
Bandpass filter with single-ended port 1 and balanced port 2.

First, we must determine the test port configuration: Logical port 1 is single-ended and corresponds to physical port 1. Logical port 2 is balanced and consists of physical ports 2 and 4. For the balanced port, a differential reference impedance of 75 Ω and a common-mode reference impedance of 25 Ω are selected. The frequency sweep is configured to have a center frequency of 1.9645 GHz and a span of 200 MHz. A measurement bandwidth of 1 kHz and a test port output power of 0 dBm are sufficient. All of these settings are easy to make with the measurement wizard, for example. The dynamic range can be influenced by varying the test port output power and IF bandwidth parameters.

If the largest possible dynamic range is desired, we recommend using the smallest possible IF bandwidth. Increasing the test port output power can also help to improve the dynamic range. Of course, the DUT and the measurement receiver must not exceed their linear operating range. A test port output power of -10 dBm or 0 dBm are frequently used standard values.

A calibration unit is used to simplify the calibration process. Once the DUT is connected, embedding at logical port 2 with a 15 nH parallel inductance is defined.

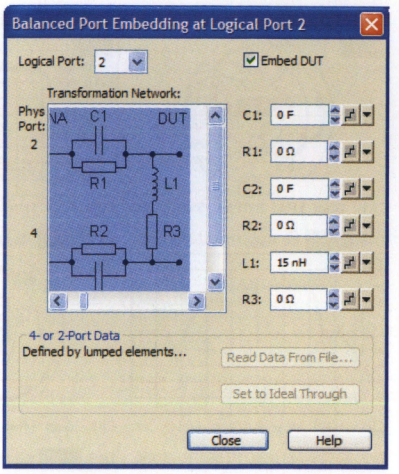


Fig. 4.12.2
Embedding with a 15 nH parallel inductance.

For the center frequency the traces in Fig. 4.12.3 show a reflection coefficient $s_{ss11} \approx 0$ at port 1 and a differential-mode reflection coefficient $s_{dd22} \approx 0$ at port 2. This means the pass band region shows a good input / output match for the unbalanced to balanced configuration in which the filter is used. The common-mode reflection coefficient s_{cc22} at port 2 within the passband region corresponds more or less to total reflection. Since the power is rejected as a result of total reflection at the DUT, it is no longer available for transmission. The transmission coefficients s_{sc12} and s_{cs21} for the signals in common-mode thus have very low values. The transmission coefficient s_{ds21} characterizes the transmission of the single-ended signal from port 1 into the differential-mode signal at port 2. This is the desired main function of the filter. Accordingly, the parameter s_{ds21} exhibits the bandpass characteristic of the filter. The quantities s_{dc22} and s_{cd22} characterize the mode conversion that arises as a result of the reflection at port 2. Both of them have very small values; this is desirable because it simplifies the separation of the common and differential-mode signals in the subsequent parts of the circuit.

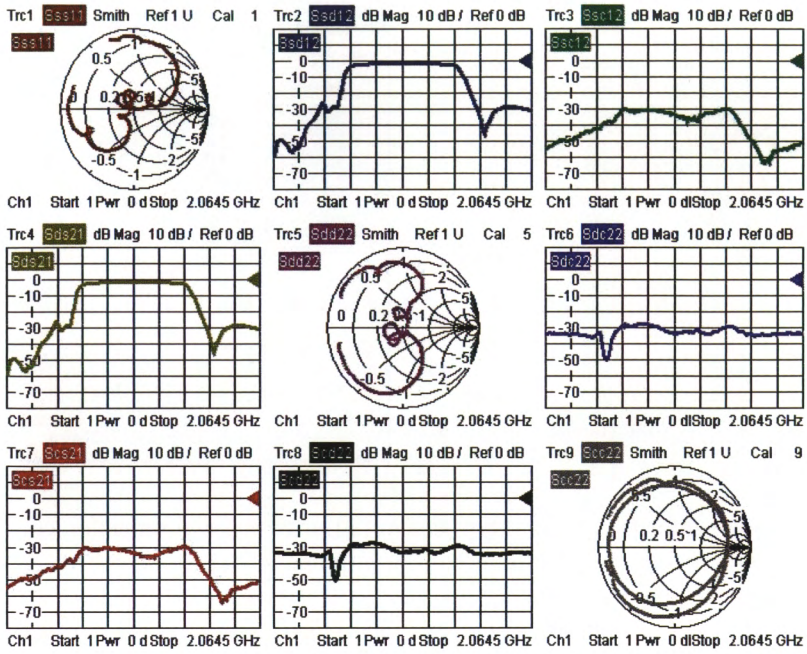


Fig. 4.12.3 Mixed-mode S-parameters for the surface acoustic wave filter.

Now we would like to analyze the parameters s_{ds21} and s_{ss11} in greater detail. Using special marker functions, it is possible to automatically determine the 3 dB cutoff frequencies, the center frequency f_{CF} and the Q factor in the s_{ds21} trace (see Fig. 4.12.4).



The center frequency f_{CF} of a bandpass filter is typically defined as the average of the upper and lower 3 dB edge frequencies. The difference between the upper and the lower band edge is known as the 3 dB bandwidth B. Another important quantity is the quality factor Q. It only fulfills the following equation in the case of a single resonant circuit or a single resonator:

$$Q = \frac{\text{Energy stored in the resonator}}{\text{Energy dissipated during a resonance period}}$$

Nevertheless, the quality factor (Q factor) can always be interpreted as the reciprocal of the relative bandwidth, i.e. $Q = f_{\mathrm{CF}}/B$.

To read off the match at the 3 dB edge frequencies in a simple manner from the s_{ss11} and s_{dd22} traces the coupling marker function can be used.



Coupling the markers makes them also visible on all of the other traces. The position of the markers on the stimulus axis is based on the trace that is selected when coupling is activated.

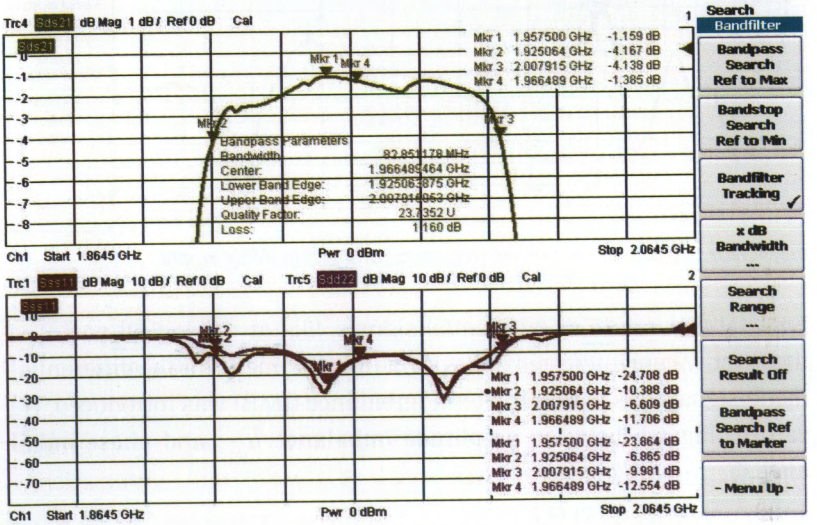


Fig. 4.12.4 Measurement of the bandwidth and center frequency.

Using statistical functions, it is possible to assess, for example, the ripple of a trace. In this process, the maximum and minimum values are calculated along with the average value for a specified section of the trace. In the following, we would like to analyze the ripple of the amplitude and the group delay of s_{cds21}. In this example, the passband of the filter with its 3 dB bandwidth has been specified as an evaluation range for the statistical function (see Fig. 4.12.5). The aperture stepwidth must be determined empirically.

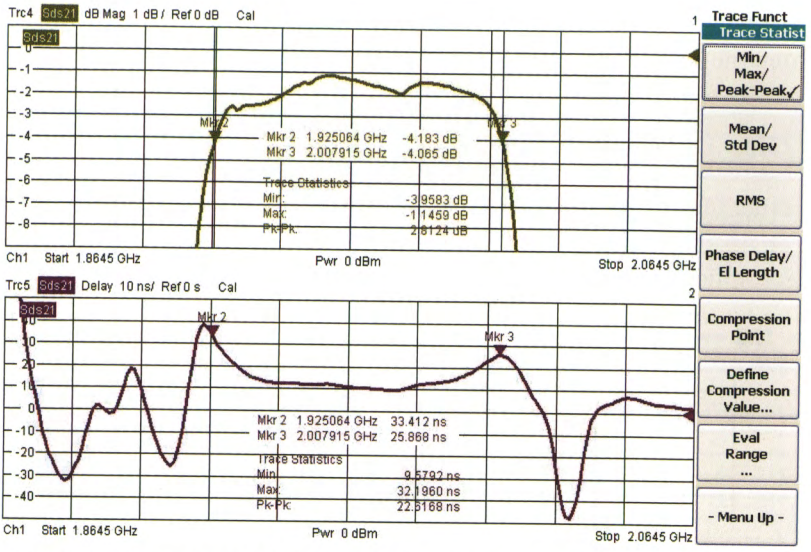


Fig. 4.12.5 Measurement of the amplitude and group delay ripple.

For any generation of differential signals, like at the logical port 2 of the filter, we are interested in having a figure of merit for the differential-mode signal. This is why the term **imbalance** (IMB) was introduced. We can distinguish between **amplitude imbalance** a_{IMB} and **phase imbalance** Φ_{IMB}.

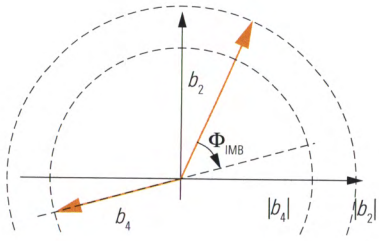


Fig. 4.12.6 Illustration of the imbalance.

The amplitude imbalance a_{IMB} is calculated from the amplitude ratio of the single-ended waves b_2 and b_4 that form the balanced logical port. Ideally, both will have the same amplitude. The values are typically displayed in dB. With the three-port device (Fig. 4.12.1), we can determine a_{IMB} from the single-ended S-parameters s_{21} and s_{41}.

$$a_{\mathrm{IMB}} = 20\lg\frac{|b_2|}{|b_4|}\,\mathrm{dB} = 20\lg\left|\frac{s_{21}}{s_{41}}\right|\mathrm{dB} \qquad \textbf{(4.12-1)}$$

The phase imbalance Φ_{IMB} characterizes the phase deviation from the ideal value $\arg(b_2) - \arg(b_4) = 180°$. For the device shown in Fig. 4.12.1, we can calculate Φ_{IMB} from the single-ended S-parameters s_{21} and s_{41}.

$$\Phi_{IMB} = \arg\left(\frac{b_2}{b_4}\right) - 180° = \arg\left(\frac{s_{21}}{s_{41}}\right) - 180° \qquad \textit{(4.12-2)}$$

In addition, we need a quantity that characterizes the difference between the **common-mode power gain** G_c and the **differential-mode power gain** G_d. This quantity is known as the **common-mode rejection ratio** (CMRR).

$$d_{CMR} = 10 \lg \frac{G_d}{G_c} \text{ dB} \qquad \textit{(4.12-3)}$$

With the network shown in Fig. 4.12.1 the differential-mode power gain G_d is measured between logical port 1 and logical port 2. Only the differential-mode power is considered at the balanced logical port 2. The common-mode power gain is also measured between logical ports 1 and 2, but only the common-mode power is considered at the balanced logical port 2. Based on the measured quantities in Fig. 4.12.3, the common-mode rejection ratio can be calculated as follows:

$$d_{CMR} = 20 \lg \frac{|s_{ds\,21}|}{|s_{cs\,21}|} \text{ dB} \qquad \textit{(4.12-4)}$$

4.13 Measurement of switching times and drift effects

Electromechanical switches (RF relays) are sometimes used to switch RF signals. They generally exhibit bouncing along with a switching time in the range of several milliseconds. Alternatively switches can be based on field effect transistors (e.g. FET switch, GaAs switch). They do not have slow mechanical drive mechanisms and can react faster. But they do require a switching time that is primarily due to charge and recharge effects. In this context, it is common practice to indicate a switching time that is referred to the 10% value and the 90% value of the transmission coefficient. The related switching time values are typically in the range of several 100 ns.

Measurement of the switching time using an external trigger

For characterization of switching time, the transmission a_{s21} or the magnitude of the transmission coefficient $|s_{21}|$ is measured using a time sweep. On the analyzer, **external trigger** is activated. The trigger should be configured so that precisely one sweep is triggered by a trigger event. The trigger signal is tapped as a 3.3 V TTL level and fed to the EXT TRIGGER connector on the network analyzer. Using a resistor network, the voltage can be adapted as required. In case of electromechanical switches, the tap should always be arranged in the digital part of the switch controller. The signal that feeds directly to the electromechanical drive mechanism should never be used since it can exhibit a **self-induced voltage** during switching. Under certain circumstances, this voltage can briefly attain a value of 100 V or greater. It can cause faulty trigger or even damage the network analyzer in the worst case.

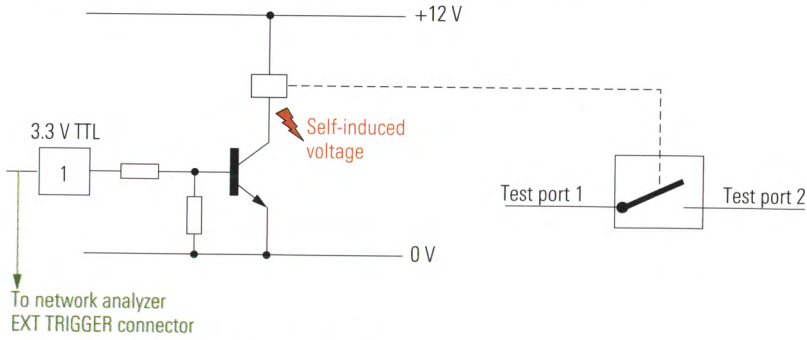


Fig. 4.13.1 Tapping the trigger signal.

Fig. 4.13.2 shows the transmission coefficient for an RF relay after it has been switched. A time sweep was triggered using a **trigger delay** of 11 ms. In this case, the changes in the transmission can be seen very easily using linear display of the amplitude. The bounce of the switch is subject to statistical fluctuations. This makes it necessary to measure multiple switching operations. Based on about 25 measurements, marker 1 was set to the time value after which no more bounce is expected.

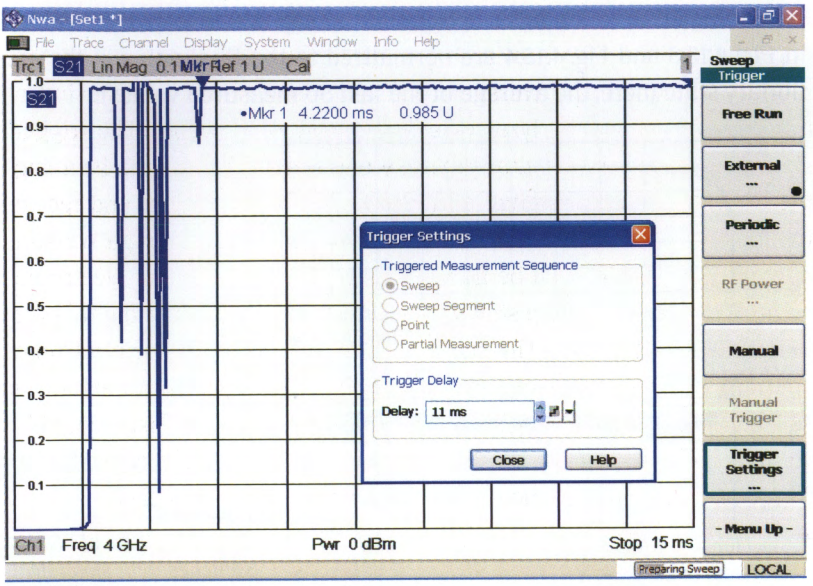


Fig. 4.13.2 *Measurement result for the switch bounce along with the trigger settings.*

Measurement of the drift using an RF trigger

In switches based on field effect transistors, we are also interested in the drift of the switch. We might observe, for example, a lagging of the level from the 90% value to the 99.9% value within 120 ms. The measurement can be triggered directly by the RF level so that no external trigger signal is required. When measuring the transmission s_{21}, the RF level of the wave quantity b_2 should be selected. The trigger threshold must be suitably adapted. With this trigger mode, the network analyzer monitors the RF level and triggers the sweep as soon as the specified trigger threshold is exceeded.

Measurement of the drift using a software trigger

Not all network analyzers have an RF trigger. For Fig. 4.13.3 and Fig. 4.13.4, a time sweep with a 1 s sweep time was configured. Manual trigger mode was selected. Using a program running on an external computer, the sweep is triggered and then the DUT (FET switch) is switched. This is known as an asynchronous trigger. The trace (magnitude and phase) is transferred to the external computer and then the behavior of the RF trigger is simulated, i.e. a search is made in the trace data for the time at which the parameter s_{21} exceeds a specified threshold of, say, –30 dB for the first time. The trace data prior to this time are discarded. The

remaining observation interval is limited to, say, 200 ms. All of the traces in Fig. 4.13.3 and Fig. 4.13.4 are normalized to the final value in the stationary state (here, the average of the last 50 measured values).

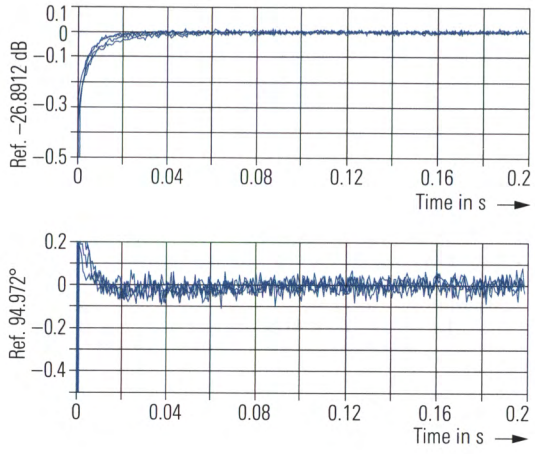


Fig. 4.13.3
Switching time of FET switches with small drift.

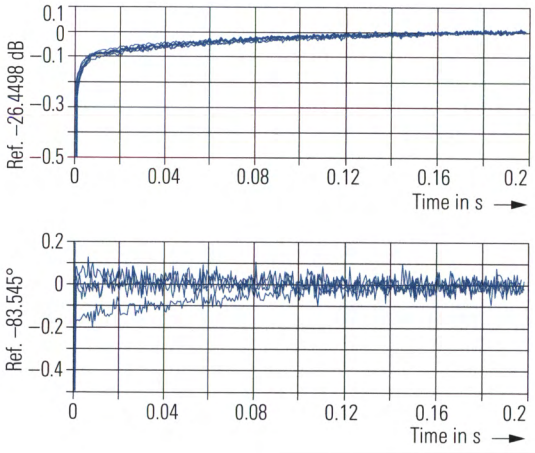


Fig. 4.13.4
Switching time of FET switches with large drift.

One problem that arises when analyzing fast switching operations is that a sweep time of about 1 s is required on the network analyzer to handle the timing fluctuations that occur with an asynchronous trigger. However, the network analyzers used in the context of this book provide a very large maximum number of measurement points (20001 points) so that a resolution of about 0.05 ms between the individual measurement points can be attained in conjunction with a suitable measurement bandwidth. This means that processes with an order of magnitude of 1 ms can be measured with no problem despite the sweep time of 1 s.

The software necessary to make such measurements can be written using the script language Matlab® from The MathWorks. Remote control of test instruments using Matlab® is described in the Application Note [1EF46], for example. That document also describes the installation procedure needed to use the `mexrsib` functions. The following program code represents one possible implementation that is based on the SCPI remote control commands for the network analyzer family used here. The controller (PC) is connected to the test instrument via a LAN. The PC's serial interface (RS-232-C) is used to control the switch. The serial interface generally has a 9-pin Sub-D connector and PIN 4 (Signal DTR) is used to turn the switch on and off. The rise time of this signal is dependent on the PC that is used (800 ns in this case). The voltage of the DTR signal (typically -12 V or +12 V) must be adapted using appropriate circuitry. The RF measurement path lies between test ports 1 and 2 of the network analyzer.

```
% Measurement of the switching behavior with an
% asynchronous trigger
% Language   : Matlab Version 7 (R14)
% Test instrument: R&S (R) ZVA, R&S (R) ZVB, R&S (R) ZVT
% History    : December 16, 2005, M. Hiebel, author

% ** Constants**
ipAdrZvb = '99.999.99.99'; % Enter IP address of ZVB here!
rfTiggerLevel = -30;   % Trigger level
maxTime = 0.2;         % Length of time segment in s
maxLevelScale = 10;    % Maximum scale value for level in dB
minLevelScale = -50;   % Minimum scale value for level in dB
maxPhaseScale = 180;   % Maximum scale value for phase in °
minPhaseScale = -180;  % Minimum scale value for phase in °

% ** Prepare the RS-232-C connection on COM1 port **
instrreset;                 % Close open connection if necessary,
                            % e.g. COM1
s = serial('COM1');         % Set DTR to nominal -12V,
                            % i.e. switch is opened

% ** Set up LAN connection to ZVB **
hdl = mexrsib('ibfind', ipAdrZvb); % Handle for RSIB.DLL

% ** Transmit instrument settings to ZVB **
mexrsib('ibwrt', hdl, '*RST');                  % Preset
mexrsib('ibwrt', hdl, '*OPC?');                 % Wait until...
dummy = mexrsib('ibrd', hdl);                   % instrument ready
mexrsib('ibwrt', hdl, 'TRIG:SOUR MAN');         % Manual trigger
mexrsib('ibwrt', hdl, 'SENS1:BWID 10kHZ');      % IF bandwidth
mexrsib('ibwrt', hdl, 'SOUR:POW 0 dBm');        % Output power
mexrsib('ibwrt', hdl, 'SENS1:SWEEP:TYPE POINT');% CW sweep
mexrsib('ibwrt', hdl, 'SENS1:FREQ:CW 1GHZ');    % Center freq.
mexrsib('ibwrt', hdl, 'SENS1:SWE:POIN 801');    % 801 points
```

```
mexrsib('ibwrt', hdl, 'SENS:SWE:TIME 1S');       % Sweep time 1s
mexrsib('ibwrt', hdl, '*OPC?');                  % Wait until...
dummy = mexrsib('ibrd', hdl);                  % instrument ready

% ** Activate trigger and turn on switch **
pause(1)                                % Matlab waits 1 s
mexrsib('ibwrt', hdl, '*TRG');          % Activate trigger
pause(0.01)
fopen(s)                                % Set DTR to nominal +12V
mexrsib('ibwrt', hdl, '*OPC?');         % Wait until...
dummy = mexrsib('ibrd', hdl);           % measurement complete

% ** Read out amplitude values from ZVB **
mexrsib('ibwrt', hdl, 'CALC1:PAR:SEL „Trc1"');  % Select Trc1
mexrsib('ibwrt', hdl, 'CALC1:FORM MAGN');       % Format dB
mexrsib('ibwrt', hdl, 'CALC1:DATA? FDATA');     % Request...
ascBuffer = mexrsib('ibrd', hdl);               % and read out
ascBuffer = strrep(ascBuffer, ',', ' ');    % Replace all commas
                                            % with spaces
vectLev = sscanf(ascBuffer, '%f');          % Convert into
                                            % numerical vector
% ** Read out phase values from ZVB **
mexrsib('ibwrt', hdl, 'CALC1:FORM PHAS');   % Format phase
mexrsib('ibwrt', hdl, 'CALC1:DATA? FDATA');% Request trace
ascBuffer = mexrsib('ibrd', hdl);           % and read out
ascBuffer = strrep(ascBuffer, ',', ' ');    % Replace commas
                                            % with spaces
vectPhas = sscanf(ascBuffer, '%f');         % Convert into
                                            % numerical vector
% ** Define time axis **
vectTime = 1/800 * [0:800]

% ** Determine exact trigger time point **
aboveIdx = min(find(vectLev>rfTiggerLevel));
belowIdx = max(find(vectLev<rfTiggerLevel));
if isempty(aboveIdx) | isempty(belowIdx)
    disp('The RF level was not exceeded');
    return;
end % if
trigTime = interp1(vectLev(belowIdx:aboveIdx), ...
                   vectTime(belowIdx:aboveIdx), ...
                   rfTiggerLevel, 'linear');
disp(['Trigger time point.....:',num2str(trigTime, ...
      '%3.5f'),' s']);

% ** Calculation of the automatic reference level **
offsLevel = mean([vectLev((length(vectLev)- ...
                  50):length(vectLev))]);
offsPhase = mean([vectPhas((length(vectPhas)- ...
                  50):length(vectPhas))]);
disp(['Automatic reference level..:',...
      num2str(offsLevel, '%3.5f'),' dB']);
disp(['Automatic reference phase..:',...
      num2str(offsPhase, '%3.5f'),' °']);
```

```
% ** Include trigger time point and reference level in computa-
tion **
timeAxis = vectTime - trigTime;
level = vectLev - offsLevel;
phase = vectPhas - offsPhase;

% ** Graphical output by Matlab **
subplot(2,1,1);
    hold on; grid on;
    title(['Software trigger']);
    plot(timeAxis, level, '-');
    handAxes1 = gca; % Handle for current axis
    set(handAxes1, 'XLimMode','manual',...
        'XLim',[0 maxTime],...
        'YLimMode','manual',...
        'YLim',[minLevelScale maxLevelScale],...
        'XGrid', 'on', 'YGrid', 'on');
    ylabel(['Ref.',num2str(offsLevel,'%2.4f'),'dB']);
subplot(2,1,2);
    hold on; grid on;
    plot(timeAxis, level, '-');
    handAxes1 = gca; % Handle for current axis
    set(handAxes1, 'XLimMode','manual',...
        'XLim',[0 maxTime],...
        'YLimMode','manual',...
        'YLim',[minPhaseScale maxPhaseScale],...
        'XGrid', 'on', 'YGrid', 'on');
    ylabel(['Ref.',num2str(offsPhase,'%3.3f'),'°']);
mexrsib('ibloc', hdl); % ZVB returns to local operation
```

4.14 Measurements on amplifiers in pulsed operating mode

In the field of mobile communications, output amplifiers are often operated in what is known as **pulsed operating mode**. In the GSM mobile phone standard (Global System for Mobile Communications), the amplifiers operate in a **timeslot** with a length of 577 µs which is repeated with a periodicity of 8 timeslots (4.616 ms). These operating conditions cannot be changed during measurements using a network analyzer for the following reasons:

- The amplifiers are usually not designed for continuous operation. They can be damaged if their power-on time is significantly exceeded.
- We would like to gain insight into the behavior of the component under realistic operating conditions.

Signals from a function generator switch the DUT on and off. The SYNC output of the function generator is connected to the EXT TRIGGER input of the network analyzer. External trigger mode is activated on the network analyzer. It must be configured so that exactly one measurement point is recorded per trigger event. This is known as a **point trigger**. The actual measurement is performed as a transmission measurement using a frequency sweep. In the present case, the sweep range is adjusted to the GSM uplink frequencies, e.g. 1710 MHz to 1850 MHz (GSM 1800) or 880 MHz to 960 MHz (GSM 900).

The **trigger delay** of 100 µs prevents the measurement from being made while the DUT is settling into its steady state. The analyzers IF bandwidth must be selected so that the settling time of the IF filter is short enough to be performed in the remaining time interval of the duty cycle. (In the present case we calculate a remaining time of 577 µs −100 µs = 477 µs.)

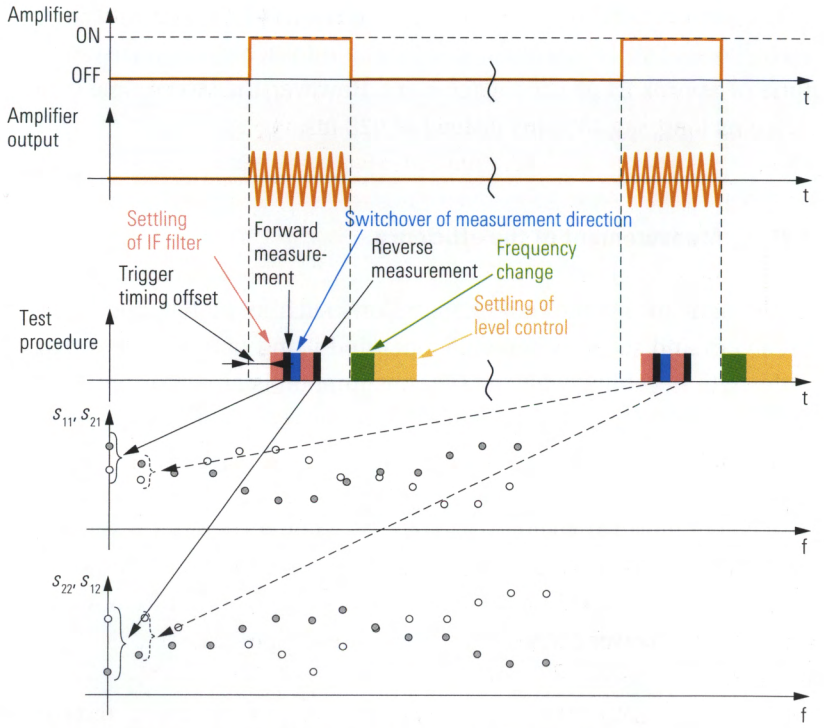


Fig. 4.14.1 Measurement procedure with two-port calibration.

Since one measurement point is recorded every 4.616 ms, a sweep containing 201 points will require a sweep time of about 928 ms.

A complete two-port calibration is usually activated for this measurement. Regardless of the selected S-parameter all two-port S-parameters have to be measured by the analyzer. Meaning within a time interval of 477 µs, the analyzer must perform the forward and reverse measurement. The necessary time is the sum of the time required for the change of direction plus the settling time of the IF filter for each direction. Typical measurement times are then around 82 µs if automatic level control is switched off and a bandwidth of 100 kHz is used. In contrast to this a measurement time of 31 µs can be achieved without a complete two-port calibration.

Some network analyzers also provide a **partial trigger** mode. If this mode is selected a trigger event will only trigger a partial measurement. For definition of partial measurement please refer to section 2.7.1. The change in measurement direction can then take place in the trigger

pauses (amplifier "OFF" state in Fig. 4.14.1). In this case, even if system error correction is activated, it is possible to achieve a measurement time of approx. 31 µs per trigger event. However, the sweep time is then twice as long, say, 1855 ms instead of 928 ms.

4.15 Measurement of the efficiency

Equipment for mobile radio applications such as mobile phones, GPS receivers and so on is generally operated using batteries or rechargeable cells. Accordingly low power consumption is an important requirement. A major factor, for this is a RF frontend using amplifiers with high **efficiency**.

The **power added efficiency** (PAE) tells us what portion of the power fed to an amplifier at its DC inputs is actually converted into RF power. The converted RF power is the difference between the output RF power and the input RF power of the DUT (typically an amplifier).

$$P_{\mathrm{RF}} = P_{\mathrm{RF\,out}} - P_{\mathrm{RF\,in}} \qquad \textit{(4.15-1)}$$

Power which is reflected by the amplifier input due to mismatch is considered to contribute to the power loss. For the RF input power, we can thus assume that $P_{\mathrm{RF\,in}} = |a_1|^2$. The output power is the power $P_{\mathrm{RF\,out}} = |b_2|^2$ which is transferred to a matched load.

$$P_{\mathrm{RF}} = |b_2|^2 - |a_1|^2 \qquad \textit{(4.15-2)}$$

From the powers involved, we can calculate the efficiency of the amplifier as follows:

$$\eta_{\mathrm{PAE}}(f, P_{\mathrm{RF\,in}}) = \frac{|b_2|^2 - |a_1|^2}{P_{\mathrm{DC}}} \qquad \textit{(4.15-3)}$$

In amplifiers, the efficiency η_{PAE} is typically dependent on the input power, operating frequency and the temperature of the circuit's components. Using a frequency sweep or a power sweep, it is possible to analyze the efficiency as a function of several different parameters. As a general rule, the efficiency will be at its maximum in the vicinity of the

1 dB compression point, while it usually decreases in the upper operating frequency range of an amplifier.

Modern network analyzers often have active test sets (see also section 2.4.3). This makes it possible to arrange the DC supply via the PORT BIAS input of the network analyzer, assuming the DUT itself does not have its own DC bias network. If the network analyzer has a passive test set, then DC decoupling must be ensured between the biasing of the DUT and the test ports of the analyzer.

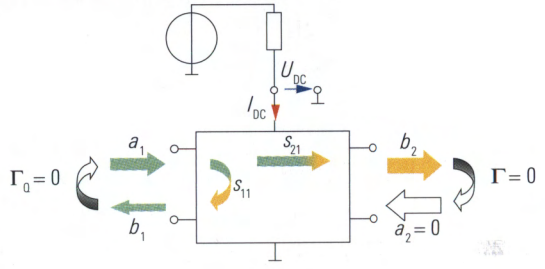


Fig. 4.15.1
Quantities for determining the efficiency.

The network analyzer has special **DC measurement inputs** (DC MEAS) for measuring the DC power consumption of the DUT. The voltage U_{DC} can usually be applied directly to these inputs. However, the current I_{DC} must be measured indirectly as a voltage drop across a **precision resistor**. On the other hand some amplifiers are supplied using a constant current source. The numeric value of the current is then entered on the network analyzer, so there is no need to measure the current. A similar procedure can be used if the amplifier is supplied with a constant voltage. The following table illustrates a few selected power supply concepts along with their associated measured quantities. Calculation of the constant k required by the network analyzer is also included.

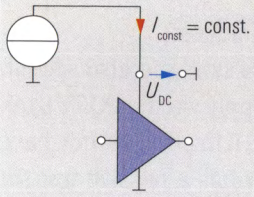
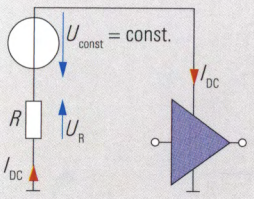
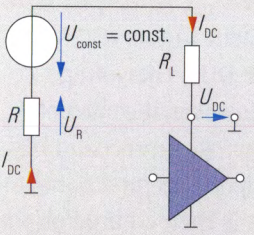
DC measurement circuit	Description and formula
	**Configuration 1: Constant current supply** $P_{DC} = I_{const} \cdot U_{DC} = k \cdot U_{DC}$ Exact computation of the DC power only requires measurement of the voltage U_{DC} on the DUT. The network analyzer is informed of the current I_{const} in the form of the constant k in A or W/V.
	**Configuration 2: Constant voltage supply** $P_{DC} = I_{DC} \cdot U_{const} = k \cdot U_R$ The power consumption in the resistor R is neglected. The voltage U_R is tapped differentially across the resistor R and used for an indirect current measurement. We know that $I_{DC} = U_R/R$ and thus $k = U_{const}/R$.
	**Configuration 3: Load resistor R_L** $P_{DC} = I_{DC} \cdot U_{DC} = k \cdot U_R U_{DC}$ The voltage on the DUT is measured as U_{DC}. As before the current measurement is handled indirectly on the resistor R. We know that $I_{DC} = U_R/R$ and thus $k = 1/R$.

Table 4.15.1 *Measurement circuits for determining P_{DC}.*

The following figure illustrates the PAE measurement on a low-amplitude signal amplifier. Configuration 2 was selected for use as our DC measurement circuit. The frequency sweep covers the range from 300 MHz to 3.3 GHz. Using a power calibration, the wave quantities a_1 and b_2 that are relevant in this measurement were calibrated. Due to the relatively low output power of the DUT (1 mW) and the DC power consumption of approx. 1.5 W (126 mA at 12 V), an efficiency of less than 0.1% was measured in its operating frequency range (300 MHz to 1 GHz).

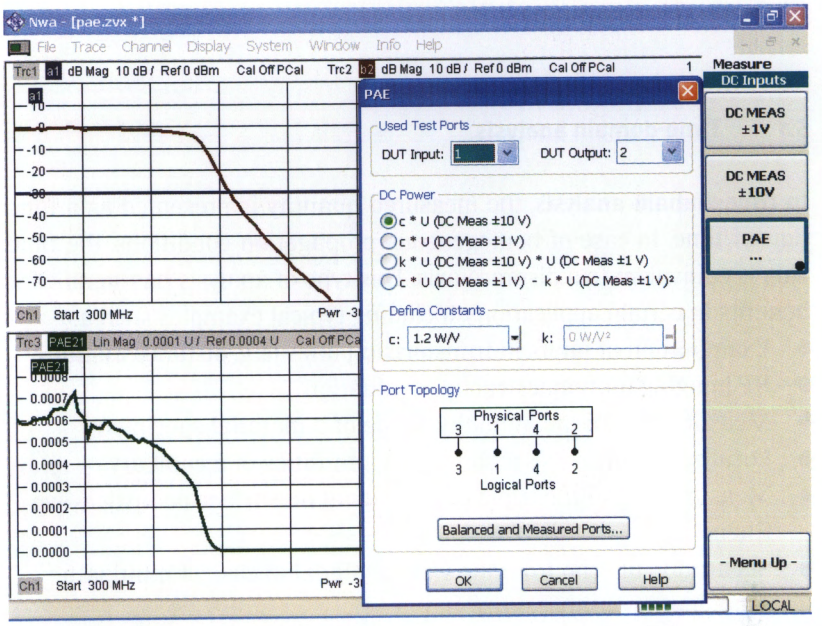


Fig. 4.15.2 Example for determining the efficiency.

5 Time-domain measurements

5.1 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 multipath propagation
- Moving the reference plane across unknown irregularities
- Calibration optimization using time-domain measurements

This chapter 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 t_d. 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 $u(t)$, wave quantities $a(t)$ and $b(t)$ and all of the time-domain quantities derived from them.

5.1.1 Impulse and step response

Networks can directly be analyzed and described in time domain. A **Dirac impulse** $\delta(t)$ 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.

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

The **response** of a linear network to a Dirac impulse $\delta(t)$ 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.

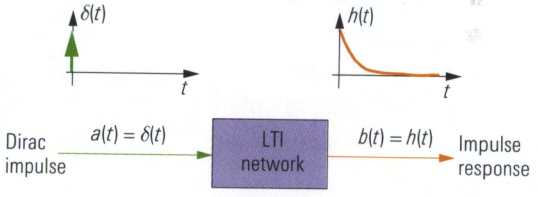


Fig. 5.1.1 Linear time-invariant network with Dirac impulse as stimulus.

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 off-

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.

set 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].

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

The generation of an ideal Dirac impulse is impossible, even its approximation is problematic. It is much simpler to generate a **unit step**[2] $\sigma(t)$. 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** $\Theta(t)$.

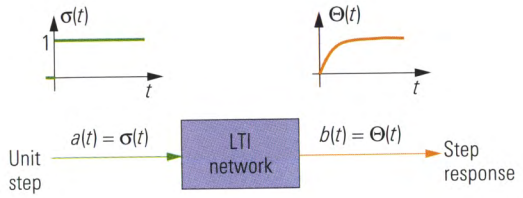


Fig. 5.1.2: Linear time-invariant network with an unit step as stimulus.

The step response $\Theta(t)$ can be calculated by integrating the impulse response $h(t)$ with respect to time[3]

$$\Theta(t) = \int_{0}^{t} h(\tau) \mathrm{d}\tau \qquad \textbf{(5.1-3)}$$

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

$$h(t) = \frac{d}{\mathrm{d}t} \Theta(t) \qquad \textbf{(5.1-4)}$$

1) In technical literature, e.g. in [Di87] on page 759, this integral is often referred to as a convolution product. The notation $b(t) = a(t) * h(t)$ has been introduced and should not be confused with the product $a(t) \cdot h(t)$!

2) Also called Heaviside function in mathematical literature.

3) Derivation for (5.1-3): We can rewrite the integral (5.1-2) in an equivalent fashion using the substitution $\tau \rightarrow t - \tau$. We thus obtain the integral $\int_{-\infty}^{\infty} h(\tau) a(t-\tau) \mathrm{d}\tau$. If we insert the unit step $\sigma(t)$ as our stimulus $a(t)$, the components for $\tau > t$ disappear since the function $\sigma(t-\tau)$ has a value of 0 in this range. In the remainder of the integration range, the function $\sigma(t-\tau)$ has a value of 1. Accordingly, we can simplify the integral as $\int_{-\infty}^{t} h(\tau) \mathrm{d}\tau$. Networks consisting of physical components always exhibit causal behavior. Accordingly, for $t < 0$ their impulse response fulfills the condition $h(t) = 0$. We can thus limit the integration to the range $\tau = 0$ to t.

5.1.2 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 $a(t) = \delta(t);\ b(t) = \Gamma_h(t)$ **(5.1-5)**

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

$\Gamma_\Theta(t)$ as the step response from $a(t) = \sigma(t);\ b(t) = \Gamma_\Theta(t)$ **(5.1-6)**

It is recommended to use the step response $\Gamma_\Theta(t)$ 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. The next section discusses the step response traces for some important one-port devices.

Analysis of a network based on the quantities $\Gamma_\Theta(t)$ and $\Gamma_h(t)$ 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.

5.1.3 Time domain reflectometry using an oscilloscope

Using a **high-speed oscilloscope** and a **step generator**, it is possible to make measurements directly in the time domain. This type of measurement provides a very intuitive starting point for interpreting the step response.

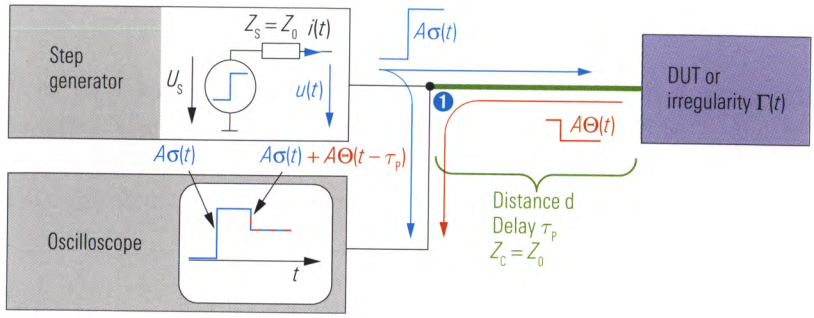


Fig. 5.1.3 Test setup with a step generator and a high-speed oscilloscope.

Using the step generator shown in Fig. 5.1.3, a voltage step $U_S \cdot \sigma(t)$ is generated at time $t = 0$. The transmission line to the DUT is assumed to be homogeneous and lossless. The probe of the oscilloscope is also assumed to have an ideal high impedance. If the characteristic impedance Z_C and the source impedance Z_S are equal to Z_0, then Z_0 can be interpreted as the reference impedance. Based on the terminal voltage $u(t)$ and the current $i(t)$, we can calculate the stimulus $a(t)$ and the response $b(t)$ in terms of wave quantities.

$$a(t) = \frac{u(t) + Z_0 i(t)}{2\sqrt{Z_0}} \qquad \textbf{(5.1-7)}$$

$$b(t) = \frac{u(t) - Z_0 i(t)}{2\sqrt{Z_0}} \qquad \textbf{(5.1-8)}$$

The terminal voltage $u(t)$ for $t > 0$ can be determined from the source voltage U_S and the voltage drop $Z_0 i(t)$ occurring across the impedance Z_0.

$$u(t) = U_S - Z_0 i(t) \qquad \textbf{(5.1-9)}$$

Next we calculate the stimulus $a(t)$ and response $b(t)$ as a function of the voltage $u(t)$ and the constant quantities U_S, Z_0. Therefore we must solve formula (5.1-9) for $i(t)$ and insert it into formulas (5.1-7) to (5.1-8).

$$a(t) = \frac{U_S}{2\sqrt{Z_0}} \qquad \textbf{(5.1-10)}$$

$$b(t) = \frac{2u(t) - U_S}{2\sqrt{Z_0}} \qquad \textbf{(5.1-11)}$$

With equation (5.1-6) we obtain from the above the step response $\Gamma_\Theta(t)$ for $t > 0$:

$$\Gamma_\Theta(t) = \frac{b(t)}{a(t)} = \frac{2u(t) - U_S}{U_S} \qquad \textit{(5.1-12)}$$

Solving formula (5.1-12) for $u(t)$ yields the voltage that can be measured with the oscilloscope at test point ❶ for $t > 0$:

$$u(t) = \frac{1}{2} U_S \left(1 + \Gamma_\Theta(t)\right) = \sqrt{Z_0}\left(a(t) + b(t)\right) \qquad \textit{(5.1-13)}$$

Apart from its constant scaling factor, the voltage $u(t)$ corresponds to the additive superposition $a(t) + b(t)$. Separation of the stimulus $a(t)$ and response $b(t)$ is possible using the delay τ_p. This delay arises due to the transmission line between the test point ❶ and the DUT. Since the signal propagates through the transmission line twice, once in the forward and once in the reverse direction, we can calculate the delay from the distance d and the velocity of propagation v_p, as follows:

$$\tau_p = \frac{2d}{v_p} \qquad \textit{(5.1-14)}$$

Case 1: One irregularity and source impedance equal to characteristic impedance ($Z_C = Z_S = Z_0$)

Based on the assumptions mentioned above we can distinguish a few special cases for the DUT. The following list contains the most important cases and clarifies what is happening electrically. As an alternative, a formal interpretation using formula (5.1-13) is also possible. The following diagrams use two coordinate systems so that it is possible to read off the voltage $u(t)$ as well as the step response $\Gamma_\Theta(t)$ with respect to a time axis. The coordinate system shown with a dashed line refers to the step response $\Gamma_\Theta(t)$.

- If the impedance of the DUT is equal to the characteristic impedance of the transmission line, then no wave is reflected: $\Gamma_\Theta(t) = 0$. The oscilloscope registers only the stimulus function $u(t) = \frac{1}{2} U_S \sigma(t)$.

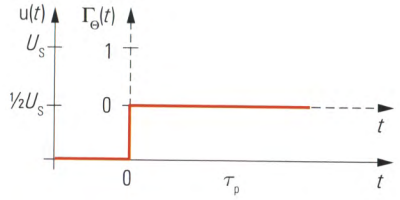


Fig. 5.1.4
u(t) and $\Gamma_\Theta(t)$ at test point ❶ for a match used as DUT.

- If the DUT is an open, the voltage is fully reflected at the transmission line's open end and the two voltages add up. We will see a staircase with two steps of equal height on the oscilloscope.

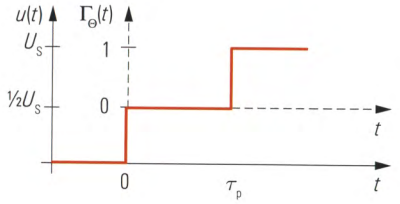


Fig. 5.1.5
u(t) and $\Gamma_\Theta(t)$ at test point ❶ for an open used as DUT.

- If the DUT is a short-circuit, the entire voltage is also reflected, but this time with the opposite sign since a voltage of 0 V must be present at the short-circuit. On the oscilloscope, the two voltages are then subtracted from one another.

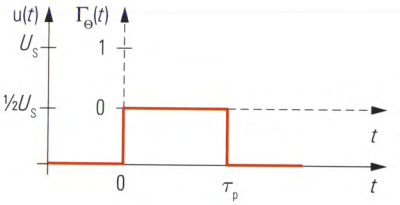


Fig. 5.1.6
u(t) and $\Gamma_\Theta(t)$ at test point ❶ for a short used as DUT.

- If a resistive load causes a mismatch $\Gamma \neq 0$, a portion of the stimulus $a(t)$ is reflected and will appear in the signal $u(t) = ½U_S(1 + \Gamma \cdot \sigma(t - \tau_p))$ on the oscilloscope.
- The transmission line can also be replaced by a line with an irregularity. Based on the delay τ_p and the velocity of propagation v_p, we can use formula (5.1-14) to calculate the distance d to the irregularity (also called distance of fault).
- A capacitor at the end of the transmission line will act as a short when the voltage step reaches it. As charge increases, the capacitor becomes like an open. The charging current returns to $i = 0$. For $t > \tau_p$, the voltage during the charging process increases according to $u(t) = U_S(1 - e^{-(t-\tau_p)/\tau})$ with the **time constant** $\tau = Z_0 \cdot C$.

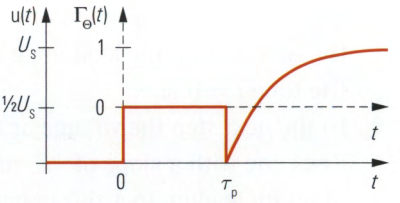


Fig. 5.1.7
u(t) and $\Gamma_\Theta(t)$ at test point ❶ for a capacitor as DUT.

- In case of an inductor at the end of the transmission line, the self-induced voltage at the beginning of the pulse prevents the flow of current. The inductor initially behaves like an open. For $t > \tau_p$ the self-induced voltage fades away, the voltage decreases according to $u(t) = U_S e^{-(t-\tau_p)/\tau}$ with the time constant $\tau = L/Z_0$.

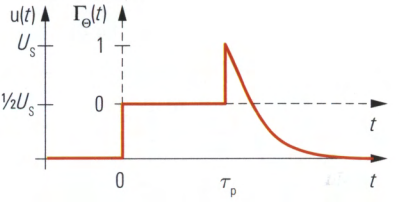


Fig. 5.1.8
u(t) and $\Gamma_\Theta(t)$ at test point ❶ for an inductor as DUT.

Case 2: Multiple reflections

Depending on the properties of the components, it is possible to observe **multiple reflections** in the system. The following gives an example of a simple test setup causing multiple reflections.

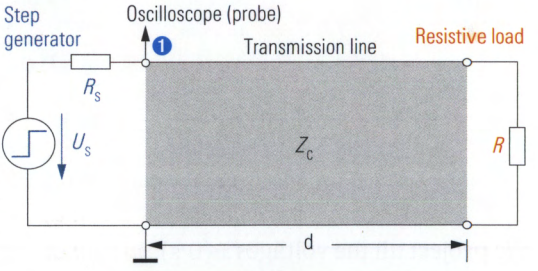


Fig. 5.1.9
Test setup with multiple reflection.

The real impedances $Z_C < R_S < R$ designate the characteristic impedance of the transmission line, the source impedance and the load impedance. The transmission line is assumed homogeneous and lossless. A graphical technique (Bergeron diagram) is used for the following discussion. We draw the lines of constant resistance $-R_S$ and R into the u/i coordinate system of Fig. 5.1.10.

- At first we determine the voltage $u(0\ \mathrm{s})$ for $t = 0$ s. A line with constant resistance Z_C starting at the u/i origin (0 A, 0 V) has to be drawn. At its intersection with the $-R_S$ line we read off the voltage

$u(0\ \mathrm{s})$. It can be calculated by $u(0\ \mathrm{s}) = U_S Z_C/(Z_C + R_S)$ alternatively. Note that the load impedance R does not has any impact at the time $t = 0$ s.

- In the next step the voltage at the time $t = \tau_P$ is of interest. Two lines one with a slope of $-Z_C$ and another with Z_C are added to the diagram leading to a zigzag pattern starting at the former intersection and yielding a second intersection with the $-R_S$ line. This intersection denotes the voltage $u(\tau_P)$ we are looking for. Note that the voltage $u(\tau_P)$ is a superposition of the stimulus and a portion that has been reflected at impedances R and R_S so far.
- Further steps can be carried out leading to voltages $u(2\tau_P)$, $u(3\tau_P)$, ... up to the **equilibrium state** $u(\infty)$.
- The voltage $u(\infty)$ can be read off from the intersection of the R and $-R_S$ line or be calculated by $u(\infty) = U_S R/(R_S + R)$. Note the characteristic impedance Z_C does not have any impact on the equilibrium state $\mathrm{t} \rightarrow \infty$.

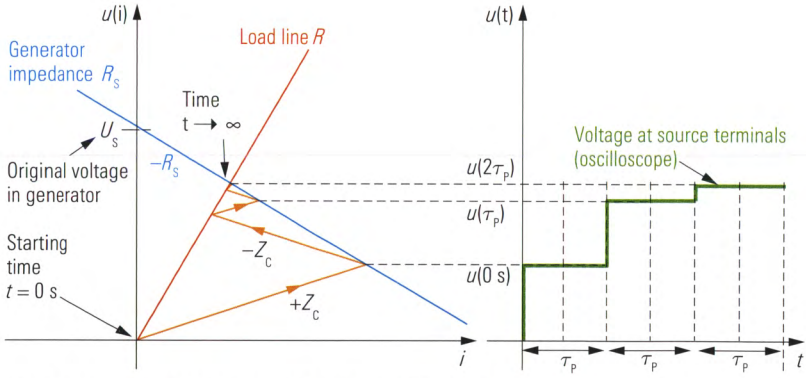


Fig. 5.1.10 Graphical analysis of the multiple reflection.

We project all the voltages $u(0\ \mathrm{s})$, $u(\tau_P)$, ... $u(\infty)$ to the u/t coordinate system shown in the right part of Fig. 5.1.10. The result is the $u(t)$ trace known from the oscilloscope. The cascaded steps of $u(t)$ with their intervals τ_P are very typical for multiple reflections within simple test setups like the one shown in Fig. 5.1.9. We can avoid multiple reflections by matching the source to the transmission line by $R_S = Z_C$; meaning that the best possible source match is a desirable aim of both a vector network analyzer and the test setup in Fig. 5.1.3.

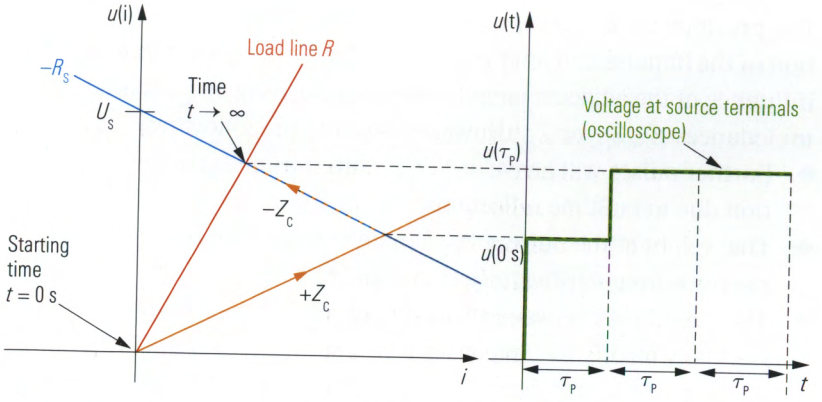


Fig. 5.1.11 *Like Fig. 5.1.10 but without multiple reflections.*

Case 3: Multiple irregularities

If we feed a step impulse or a Dirac impulse into a DUT with several distributed irregularities, a portion of the signal will be reflected at each irregularity. Furthermore multiple reflections will occur between irregularities and result in some cascaded step pattern or some virtual impulses. The following table shows the situation when a single Dirac impulse arrives at $t = t_0$ at the irregularity.

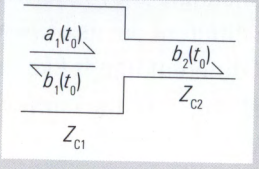
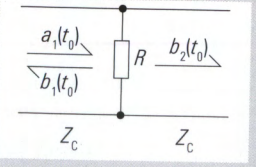
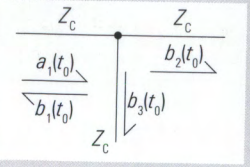
Irregularity	Formulas
Impedance step 	$a_1(t_0) = \delta(t - t_0)$ $b_1(t_0) = \frac{Z_{C2} - Z_{C1}}{Z_{C2} + Z_{C1}} \delta(t - t_0)$ $b_2(t_0) = \frac{2\sqrt{Z_{C2} \cdot Z_{C1}}}{Z_{C2} + Z_{C1}} \delta(t - t_0)$
Shunt resistor 	$a_1(t_0) = \delta(t - t_0)$ $b_1(t_0) = \frac{-Z_C}{2R + Z_C} \delta(t - t_0)$ $b_2(t_0) = \frac{2R}{Z_C + 2R} \delta(t - t_0)$
T-junction 	$a_1(t_0) = \delta(t - t_0)$ $b_1(t_0) = -\frac{1}{3} \cdot \delta(t - t_0)$ $b_2(t_0) = b_3(t_0) = \frac{2}{3} \cdot \delta(t - t_0)$

Table 5.1.1 *Typical irregularities stimulated by a single Dirac impulse observed at the time $t = t_0$.*

The previous table 5.1.1 represents only a snap-shot for the first interaction of the impulse and an irregularity at time $t = t_0$. At this moment the influence of the adjacent circuit is given exclusively by its characteristic impedances Z_C, Z_{C1} or Z_{C2}. However consider the following:

- Further pulses will arrive at t_1, t_2, ... even from the opposite direction due to multiple reflections.
- The weight of the pulse $a_1(t_0)$ can be smaller than one because of previous irregularities, reflections and line loss.
- The circuit part between the test point ❶ and the irregularity will normally modify the responses $b_1(t_0)$, $b_1(t_1)$, ... even an interchange of their sign is possible.

To determine the individual parameter (Z_{C2}, Z_{C1}, R or Z_C) of each irregularity all of the multiple reflections, line losses and irregularities have to be located by a special algorithm. A description of this algorithm would exceed the scope of this book.

5.1.4 Fourier transform

Any periodic time domain function can be rewritten as the sum of a series of harmonics, like the black trace of Fig. 5.1.13 which yields $0.5\ \text{V} + 0.636\ \text{V}\cdot\cos(2\pi t\cdot 1\ \text{GHz}) + 0.2122\ \text{V}\cdot\cos(2\pi t\cdot 3\ \text{GHz} + \pi)$. Note that both representations are equivalent, meaning they result in the same signal. It is like looking at a house from different points of view, the house remains the same but we will discover different things like its front door or its back door. The weighs and initial angles of the harmonics are known as amplitude and phase respectively.

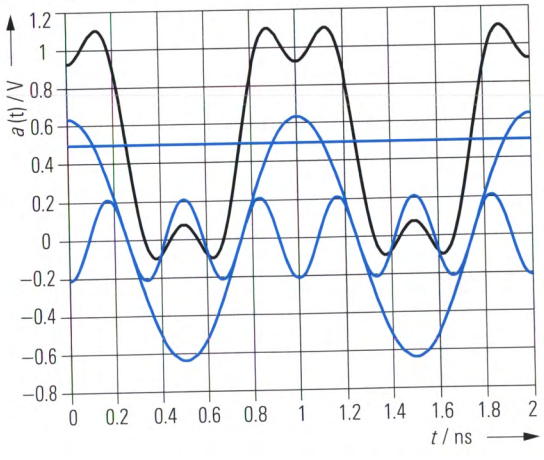


Fig. 5.1.13
Composite oscillation.

Using the **continuous Fourier transform**[1] (FTC), we can also calculate the decomposition of nonperiodic functions such as a step response into sinusoidal components. The calculated spectrum is continuous instead of containing only discrete harmonic frequencies. The reverse of this procedure is known as the **inverse continuous Fourier transform** (IFTC).

Impulse response and frequency response

The continuous Fourier transform of the impulse response $\Gamma_h(t)$ yields a spectrum $\Gamma_H(f)$ which is known as the frequency response. The reverse of this procedure is also possible using the inverse continuous Fourier transform[2].

$$\Gamma_h(t) \circ\!\!-\!\!\bullet\, \Gamma_H(f) \qquad \textbf{(5.1-15)}$$

Formulas (5.1-16) and (5.1-17) define the continuous Fourier transform and its inverse. Note, however, that tables are normally used to obtain a Fourier correspondence.

FTC:
$$\Gamma_H(f) = \int_{-\infty}^{\infty} \Gamma_h(t) e^{-j2\pi ft}\,\mathrm{d}t \qquad \textbf{(5.1-16)}$$

IFTC:
$$\Gamma_h(t) = \int_{-\infty}^{\infty} \Gamma_H(f) e^{j2\pi ft}\,\mathrm{d}f \qquad \textbf{(5.1-17)}$$

Similar to formula (5.1-16), we can define the spectral representation $a_s(f)$ and $b_s(f)$ for the stimulus $a(t)$ and the response $b(t)$. The spectra $a_s(f)$, $b_s(f)$ and $\Gamma_H(f)$ are related by the following linear equation:

$$b_s(f) = \Gamma_H(f) \cdot a_s(f) \qquad \textbf{(5.1-18)}$$

This relationship corresponds to formula (1.2-1) from section 1.2. Of course, the phasor Γ that is used in formula (1.2-1) was considered only for frequencies $f \geq 0$. If we want to be able to use the definition of the inverse Fourier transform from formula (5.1-17), the reflection coefficient $\Gamma(f)$ must be continued towards the negative frequency range. As a boundary condition, we require the impulse response $\Gamma_h(t)$ calculated using (5.1-17) to have a real value. This is precisely the case if the con-

1) In technical literature, it is common to omit the term "continuous". Accordingly, the abbreviation FT is customary instead of FTC and the same with IFTC.

2) The symbol "$\circ\!\!-\!\!\bullet$" will be used hereafter to represent the correspondence between quantities in the time and frequency domains.

version formula between the quantities $\Gamma(f)$ and $\Gamma_H(f)$ is specified as follows.[1)]

$$\Gamma_H(f) = \begin{cases} \Gamma(f) & \text{for } f \geq 0 \\ \Gamma^*(-f) & \text{for } f < 0 \end{cases} \qquad \textit{(5.1-19)}$$

Similar observations[2)] can also be made for the spectra $a_s(f)$, $b_s(f)$ and the quantities a, b. The following figure summarizes the relationships given in this section which form the theoretical background for the time-domain option of a network analyzer.

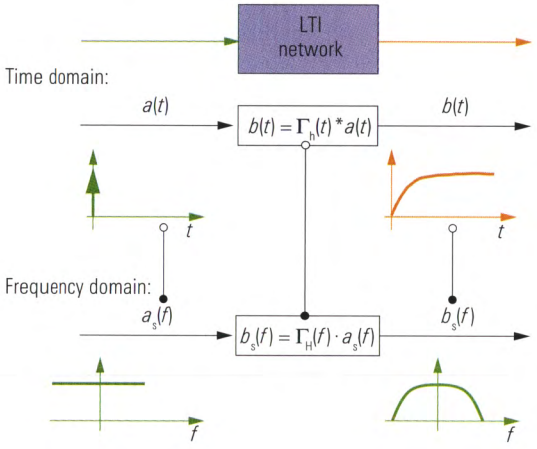


Fig. 5.1.14
Comparison of time- and frequency-domain representation.

5.2 Numerical inverse Fourier transform

This section discusses the theoretical relationship between the inverse Fourier transform in accordance with formula (5.1-17) and its numerical implementation in the network analyzer. It is intended for readers who are interested in a detailed discussion of the relationships introduced in

1) $\Gamma^*(f)$ is used to designate the conjugate complex value of $\Gamma(f)$. For more details, see Appendix A1.

2) Remark: The quantities $\Gamma(f)$ and $\Gamma_H(f)$ are quotients which have not assigned a physical unit. In contrast to this the quantities $a(f)$, $b(f)$, $a_s(f)$ and $b_s(f)$ have a unit of $\sqrt{W}$. The integral $\int_0^\infty |a(f)|^2 - |b(f)|^2 \mathrm{d}f$ leads to the energy that is absorbed by a one port device. The value calculated this way should be the same as given by the integral $\int_{-\infty}^{\infty} |a_s(f)|^2 - |b_s(f)|^2 \mathrm{d}f$. To ensure this, formula (5.1-19) has to be modified when it is used with quantities $a(f)$ and $a_s(f)$. Meaning the terms $\Gamma(f)$ and $\Gamma^*(-f)$ have to be replaced by $\frac{1}{2}a(f)$ and $\frac{1}{2}a^*(-f)$ for all frequencies $f \neq 0$. The same applies to the quantities $b(f)$ and $b_s(f)$.

section 5.3. The discussion in section 5.3 briefly revisits the main results from the previous section. Readers who so desire can skip the current section and concentrate on section 5.3.

In a vector network analyzer, the magnitude and phase of the measured quantities are recorded over a discrete frequency axis. Due to the discrete frequency axis and the finite frequency range $0 \leq f < \infty$, an inverse continuous Fourier transform according the formula (5.1-17) cannot be performed. Instead, we use an **inverse discrete Fourier transform (IDFT)**. We will have a look at the peculiarities of IDFT using the following example with $N = 5$ points. In practice, it is typical to have a total number of points $N >> 5$. Typical values are $N = 100$ to several times 1000.

5.2.1 Inverse discrete Fourier transform

Let us assume a complex spectrum with the constant **frequency step size** Δf. The frequency axis begins at $f_0 = 0$ Hz and includes N **frequency samples**. The following frequency vector is defined:

$$f = \begin{bmatrix} f_0 & f_1 & f_2 & \dots & f_n & \dots & f_{N-1} \end{bmatrix} \qquad \textit{(5.2-1)}$$

where $f_n = \Delta f \cdot n$ and $n = \{0 \dots N-1\}$ *(5.2-2)*

Since the frequencies f_1, f_2, … f_{N-1} are all integer multiples (harmonics) of the frequency $f_1 = \Delta f$, we speak of a **harmonic grid**. The impulse response that is computed must have time-discrete values so that they can be processed numerically. These values have the **sample time** Δt. We can process only a finite number M of **time-domain samples**.

$$t = \begin{bmatrix} t_0 & t_1 & t_2 & \dots & t_m & \dots & t_{M-1} \end{bmatrix} \qquad \textit{(5.2-3)}$$

where $t_m = \Delta t \cdot m$ and $m = \{0 \dots M-1\}$ *(5.2-4)*

The following pages give a detailed insight of an inverse discrete Fourier transform. The data set in Fig. 5.2.1 does not contain any values for $f < 0$.

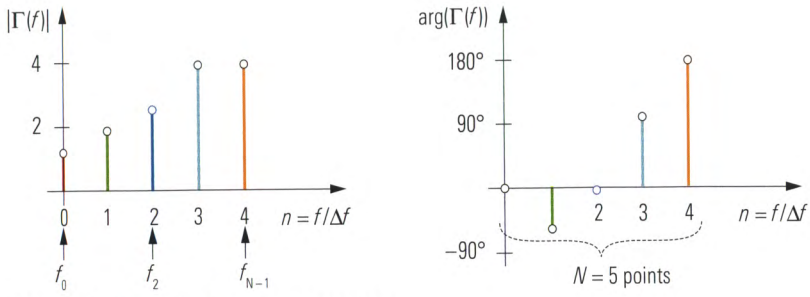


Fig. 5.2.1 Discrete frequency response.

Using the relationship from formula (5.1-19), we can add the values for the negative frequency axis. What we obtain is known as the discrete frequency response $\Gamma_H(m\Delta f)$. It contains a total of M values:

$$M = 2N - 1. \qquad \textit{(5.2-5)}$$

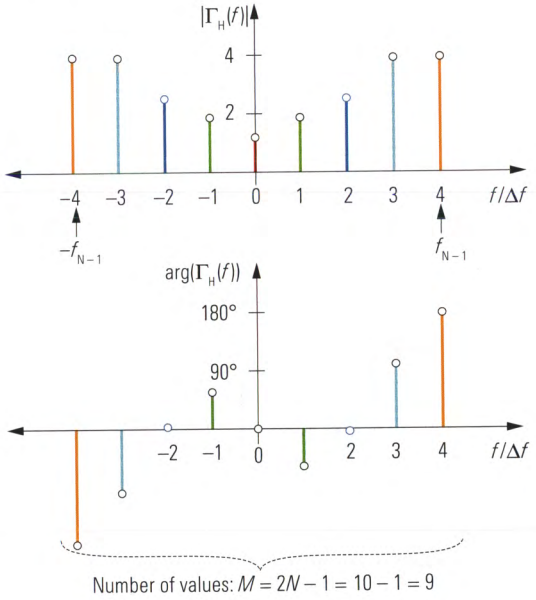


Fig. 5.2.2 Discrete frequency response.

The expected transformation result should contain time-discrete values with the sample time Δt. However, **sampling** of the time axis using Δt requires a **spectrum** that is periodical with $F = 1/\Delta t$. If we choose a periodic continuation that is too narrow, the frequency response $\Gamma_H(m\Delta f)$ will have an overlap with its periodic continuation (see the ranges with the gray background in Fig. 5.2.3). This would deform the computed time-domain trace. Ideal is a periodic continuation of the

spectrum where $F = \Delta f \cdot M$ (see Fig. 5.2.4). Together with $F = 1/\Delta t$ we yield the following time step size:

$$\Delta t = \frac{1}{F} = \frac{1}{\Delta f \cdot M} \qquad \textit{(5.2-6)}$$

The time step size according to formula (5.2-6) is also known as the **time resolution**. The reason for this will become clear in section 5.2.2. The spectrum $\Gamma_p(m\Delta f)$ shown in Fig. 5.2.4 is called periodically continued.

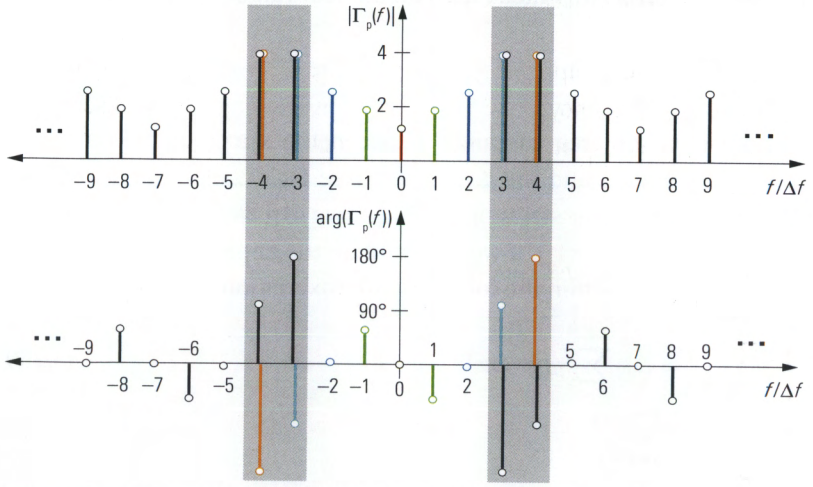


Fig. 5.2.3 Spectrum with excessively narrow periodic continuation.

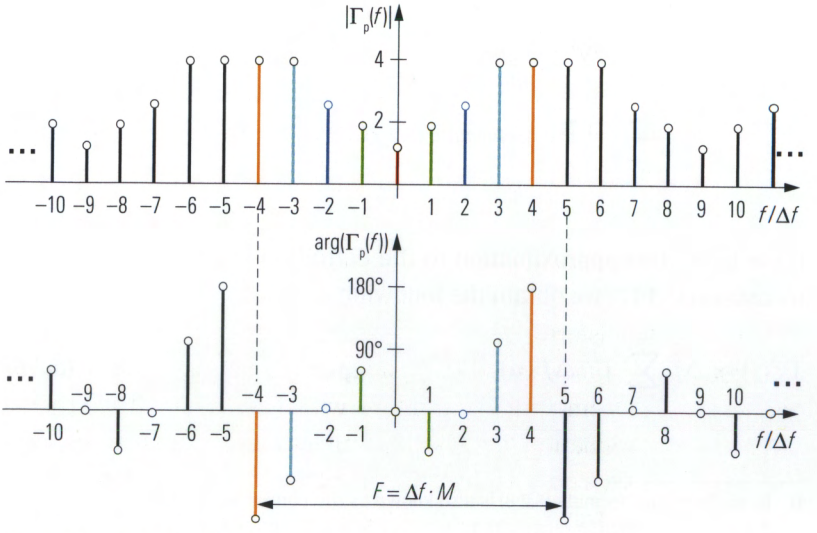


Fig. 5.2.4 Periodically continued spectrum.

Now it is necessary to derive a discrete equivalent of the inverse continuous Fourier transform (IFTC). Therefore formula (5.1-17) will be adapted in steps a) to c) below to meet the specific requirements of a numerical transformation.

a) Approximation of the integral

First, the integral from formula (5.1-17) is replaced by a sum.

The conversion is illustrated based on the real-valued integral as follows:

$$\int_{-\infty}^{\infty} y(x)\,dx$$

It represents the area contained between the x axis of the Cartesian coordinate system and the graph $y(x)$ (see Fig. 5.2.5). If we divide this area into rectangular strips of width Δx, we can calculate the area of a single strip as $\Delta x \cdot y(x_0)$. If we then add up the areas of each of the individual strips, we obtain the following approximation:

$$\int_{-\infty}^{\infty} y(x)\,\mathrm{d}x \approx \Delta x \sum_{k=-\infty}^{\infty} y(k \cdot \Delta x) \qquad \textit{(5.2-7)}$$

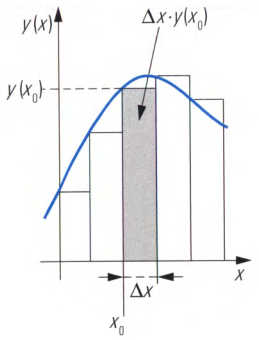


Fig. 5.2.5
Illustration of the approximation according to formula (5.2-7).

If we apply this approximation to the definition of the inverse Fourier transform (5.1-17), we obtain the following sum:[1)]

$$\Gamma_{\mathrm{h}}(t) \approx \Delta f \sum_{n=-\infty}^{\infty} \Gamma_{\mathrm{p}}(\Delta f \cdot n) e^{j2\pi n \Delta f \cdot t} \qquad \textit{(5.2-8)}$$

1) In mathematics, formula (5.2-8) is also known as the Fourier series. The terms $\Delta f \cdot \Gamma_{\mathrm{p}}(\Delta f \cdot n)$ where n ranges from $-\infty \ldots +\infty$ would correspond to what is known as the Fourier coefficients.

b) Reduction of the sum to a finite interval

For further evaluation, the sum (5.2-8) must be reduced to the finite interval from $-(N-1)$ to $N-1$. The spectrum outside this interval is masked using a **window function** $W(f)$. In the simplest case, this involves the use of the following **rectangle window**[1]:

$$W(n) = \begin{cases} 1/F & \text{for } |n| \leq N-1 \\ 0 & \text{otherwise} \end{cases} \qquad \textit{(5.2-9)}$$

The bandwidth of the window $W(n)$ is determined by the periodicity of the spectrum; it is equal to F. To avoid any unwanted scaling effects the area contained under the rectangle window has to be one. This is the reason for $1/F$ in formula (5.2-9). We can now replace the periodic spectrum $\Gamma_{\mathrm{p}}(\Delta f \cdot n)$ in formula (5.2-8) by the discrete frequency response $\Gamma_{\mathrm{H}}(\Delta f \cdot n)$ from Fig. 5.2.2.

$$\Gamma_{\mathrm{h}}(\Delta t \cdot m) = \frac{\Delta f}{F} \sum_{n=-(N-1)}^{N-1} \Gamma_{\mathrm{H}}(\Delta f \cdot n) e^{j2\pi \cdot nm/M} \qquad \textit{(5.2-10)}$$

As we know from formula (5.2-6), the factor $\Delta f/F$ is equal to the inverse number of points $1/M$. We thus obtain the conventional definition of the inverse discrete Fourier transform.

$$\Gamma_{\mathrm{h}}(\Delta t \cdot m) = \frac{1}{M} \sum_{n=-(N-1)}^{N-1} \Gamma_{\mathrm{H}}(\Delta f \cdot n) e^{j2\pi \cdot nm/M} \qquad \textit{(5.2-11)}$$

c) Simplification for systems with real-valued time-domain quantities

The frequency response $\Gamma_{\mathrm{H}}(f)$ can be written in terms of magnitude $|\Gamma_{\mathrm{H}}(f)|$ and phase $\arg(\Gamma_{\mathrm{H}}(f))$. The phase values can be accommodated in the complex exponential function $e^{j(2\pi nm/M + \arg(\Gamma_{\mathrm{H}}(f)))}$. It can be represented as $\cos(2\pi nm/M + arg(\Gamma_{\mathrm{H}}(f))) + j\sin(2\pi nm/M + arg(\Gamma_{\mathrm{H}}(f)))$. In the current application, a real-valued time-domain signal $\Gamma_{\mathrm{h}}(t)$ can be assumed so that the term $j\sin(2\pi nm/M + arg(\Gamma_{\mathrm{H}}(f)))$ is unnecessary and we have the following:

$$\Gamma_{\mathrm{h}}(\Delta t \cdot m) = \frac{1}{M} \sum_{n=-(N-1)}^{N-1} |\Gamma_{\mathrm{H}}(\Delta f \cdot n)| \cos\left(\frac{2\pi nm}{M} + \arg(\Gamma_{\mathrm{H}}(\Delta f \cdot n))\right) \qquad \textit{(5.2-12)}$$

1) In some technical literature (e.g. [1E244]) the rectangle window is called no profiling.

The cos term provides the same value for symmetric indices n and $-n$. The frequency response fulfills the condition $\arg(\Gamma_H(\Delta f \cdot n)) = -\arg(\Gamma_H(\Delta f \cdot n))$. The terms for negative and positive n can be combined in a pair-by-pair fashion as $2|\Gamma_H(\Delta f \cdot n)|\cos(2\pi nm/M + \arg(\Gamma_H(\Delta f \cdot n))$. No combination is possible for $n = 0$. All other terms yield a sum of $n = 1 \ldots N-1$. It is thus possible, to directly use the measured values $\Gamma(\Delta f \cdot n)$ defined at frequencies $f \geq 0$ (i.e. $n = 0 \ldots N-1$), instead of the frequency response $\Gamma_H(\Delta f \cdot n)$.

$$\Gamma_h(\Delta t \cdot m) = \frac{\Gamma(\Delta f \cdot 0)}{M} + \frac{2}{M}\sum_{n=1}^{N-1} |\Gamma(\Delta f \cdot n)| \cos\left(\frac{2\pi nm}{M} + \arg(\Gamma(\Delta f \cdot n))\right)$$

(5.2-13)

Under the assumption of a real-valued time-domain signal, we can apply formula (5.2-13) directly to our measured phasor values and calculate their time-domain representation. We may even solve this formula on paper, provided the number of frequency samples is small enough. To give a better understanding the following paragraph describes this paper work step by step using graphically means:

To construct Fig. 5.2.6, we apply formula (5.2-13) stepwise to the data set from Fig. 5.2.1. The measured phasors $\Gamma(f_1)$, $\Gamma(f_2)$, $\Gamma(f_3)$ and $\Gamma(f_4)$ are scaled with the factor $2/M \approx 0.22$. They are drawn in as rotating phasors in the complex coordinate system in the top part of Fig. 5.2.6. For the measured value $\Gamma(f_0)$, scaling with $1/M \approx 0.11$ is necessary. Corresponding to their frequencies $f_0 = 0$, $f_1 = \Delta f$, $f_2 = 2\Delta f$, $f_3 = 3\Delta f$, the phasors rotate at different angular speeds. The phasor associated with $n = 0$ has the frequency $f = 0$ and thus does not rotate. For each phasor, a real time-domain signal can be plotted that arises through projection of the rotating phasor on the real axis $\mathrm{Re}\{\Gamma(f)\}$. At discrete time steps defined by the formulas (5.2-4) and (5.2-6), the sum of all time-domain components belonging to a time step is calculated and drawn in as a discrete result (black circles). We thus obtain the discrete time-domain signal $\Gamma_h(\Delta t \cdot m)$. In Fig. 5.2.6, the time-domain function repeats with a period T. This is due to the harmonic frequency vector cf. formulas (5.2-1) and (5.2-2). All of the frequencies $f > 0$ of this vector are integral multiples of the frequency Δf. The lowest frequency $f_1 = \Delta f$ thus determines the period T of the time-domain signal.

$$T = 1 / \Delta f \qquad \textbf{(5.2-14)}$$

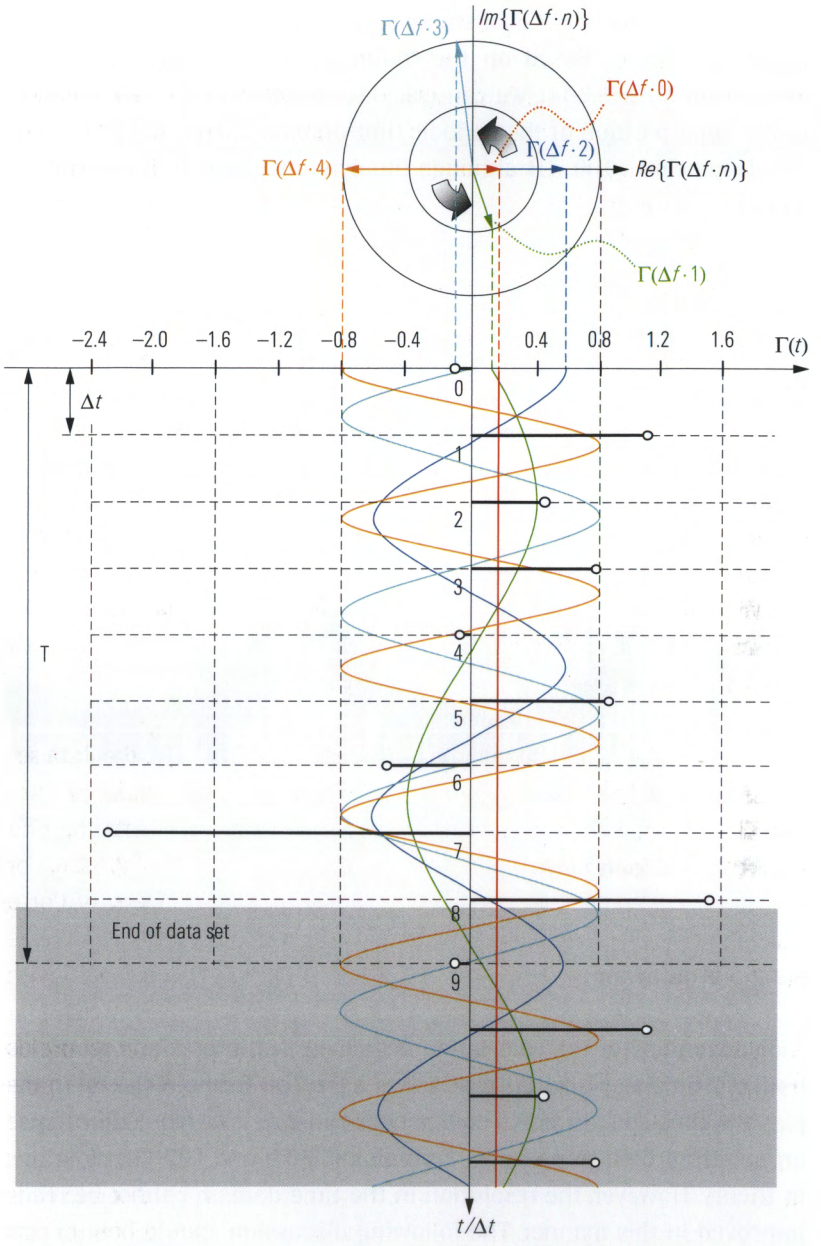


Fig. 5.2.6 Construction of the time-domain signal corresponding to Fig. 5.2.1.

The interval from $-T/2$ to $T/2$ is known as the **ambiguity range**. Only within this range is it possible to uniquely assign the time axis. If the time function decreases very slowly and the ambiguity range is too narrow, the impulse response is distorted as a result of the overlap of adia-

cent periodic curves. Fig. 5.2.7 shows a time-domain trace with $N = 800$ measured values. Based on the frequency step size $\Delta f = 10$ MHz and on formula (5.2-14), we can calculate the period $T = 100$ ns. Due to the superposition of the periodic time-domain curves, the trace does not drop below -100 dB although the dynamic range of the hardware would be better.

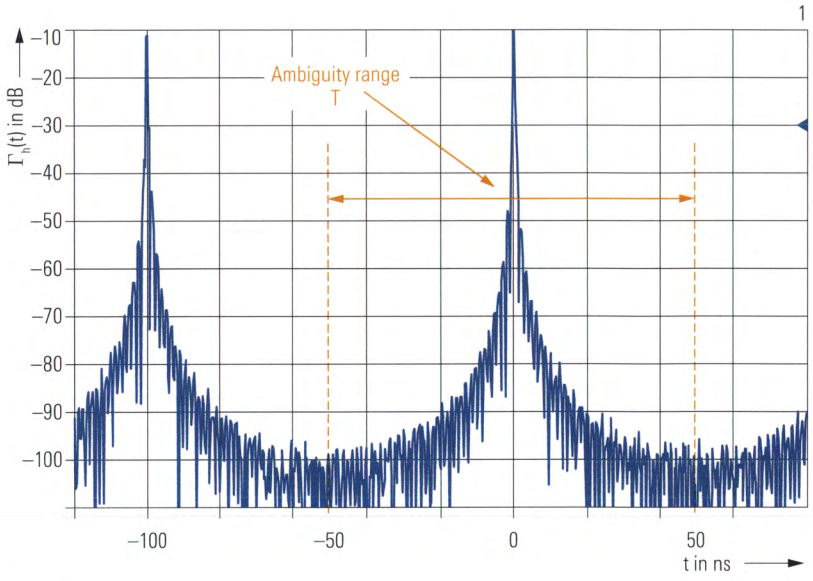


Fig. 5.2.7 Ambiguity range.

5.2.2 Windowing

A disadvantage of the calculation described in the previous section is its fixed time step size Δt that arises based on formula (5.2-6). To display the time-domain signal on a network analyzer, we would like to use an arbitrary time step size t_d. Formulas (5.2-11) and (5.2-13) allow this in theory. However, the resolution in the time domain cannot be really improved in this manner. The following discussion should help to clarify why this is the case.

Using the rectangle window, the periodic spectrum from Fig. 5.2.4 is reduced to the range $-N$ to N yielding an occupied bandwidth F. If we want a time step size different from $\Delta t = 1/F$, then there are further considerations necessary. We have to investigate the effects of win-

dowing on the time-domain signal: Windowing with any $W(f)$ corresponds to multiplication of the frequency response $\Gamma_H(n\Delta f)$ by the window $W(f)$. The equivalent in the time domain is the convolution $\Gamma_h(n\Delta t)*w(t)$. Due to the discrete axes, a sum is necessary to do the convolution instead of an integral. Like the integral, the sum extends over the range $n = -\infty$ to ∞. The time-domain function $\Gamma_h(n\Delta t)$ thus must be continued periodically for the convolution outside of $|n| > N - 1$.

$$\sum_{n=\infty}^{n=-\infty} \Gamma_h(n\Delta t)\, w(t - n\Delta T) \quad \circ\!\!-\!\!\bullet \quad W(f)\Gamma_H(f) \qquad \textbf{(5.2-15)}$$

Applying the inverse Fourier transform to $W(f)$ yields the time-domain representation $w(t)$ of the window. In the case of a rectangle window defined by formula (5.2-9), we obtain the so-called si function:[1)]

$$w(t) = \mathrm{si}(\pi \cdot t / T) = \frac{\sin(\pi \cdot t / T)}{\pi \cdot t / T} \qquad \textbf{(5.2-16)}$$

The rectangle window and its equivalent in the time domain, the si function, are shown in the following figure. Besides the main pulse, the si function has significant side lobes. The greatest (negative) **side lobe** is located to the right and left of the main pulse. It has an amplitude of -0.212 or -13 dB.

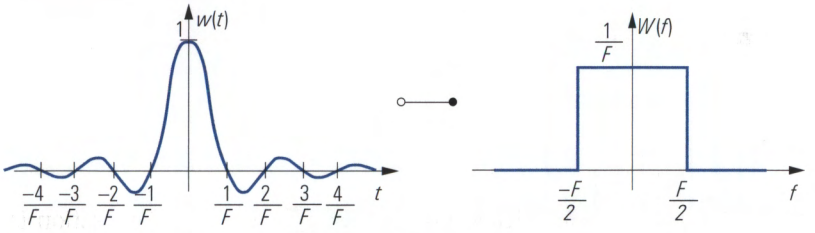


Fig. 5.2.8 si function and rectangle window.

We insert the si function as $w(t)$ into formula (5.2-15). The convolution is done in a similar way, like explained in context of formula (5.1-2). We get the dashed curves of Fig. 5.2.9, which have to be summed up for each individual frequency point. The result (the colored circles) represents our time-domain signal. This signal is familiar from Fig. 5.2.6. But

1) The function si(x) should not be confused with the integral sine Si(x), e.g. [BS91], page 14 which has a different meaning.

the method used here it is more advantageous in context of the following discussion.

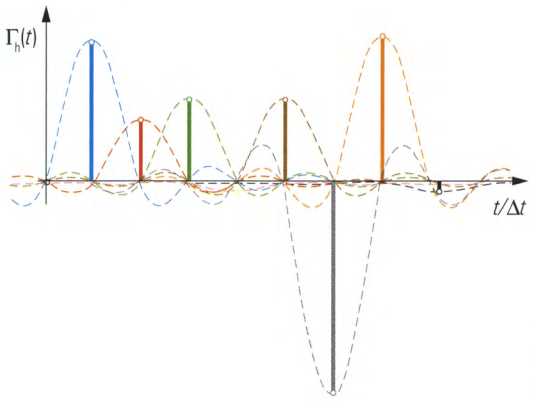


Fig. 5.2.9
No influence of adjacent values for step size Δt.

The zeros of the si function are located at a distances of $1/F$, $2/F$, $3/F$, ... to the right and left of the main pulse. This explains why the si function does not have any influence on the computation of the time-domain signal according to formulas (5.2-10) to (5.2-13). The discrete time steps 0, Δt, $2\Delta t$, $3\Delta t$, ... are chosen so that in each sample exactly one of the si functions makes a non-zero contribution at the time step $m\Delta t$. If we choose a time step size t_d, half of the value Δt the situation becomes different.

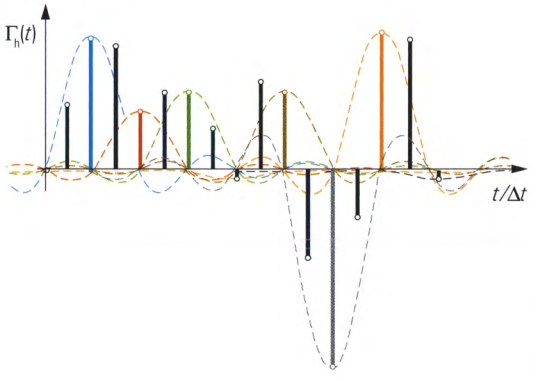


Fig. 5.2.10
Influence of adjacent values for $t_d < \Delta t$.

The newly added time samples exhibit a superposition of multiple si functions (see Fig. above). The values Γ_h belonging to these samples are not only influenced by the si functions next to them. They are also influenced by si functions whose main pulse is located far away, particularly if these si functions have a large weight. This leakage effect explains

why an improved resolution is not possible for $t_d < \Delta t$. Hence it is obvious to designate Δt as the time domain resolution.

The following figure illustrates an extreme case of pulse leakage. It contains two pulses with very different weights. The dashed traces (red and green) represent the individual pulses at 0 ps and 300 ps. Their main pulses differ by a factor of 1/20. Normally, they are not accessible separately.

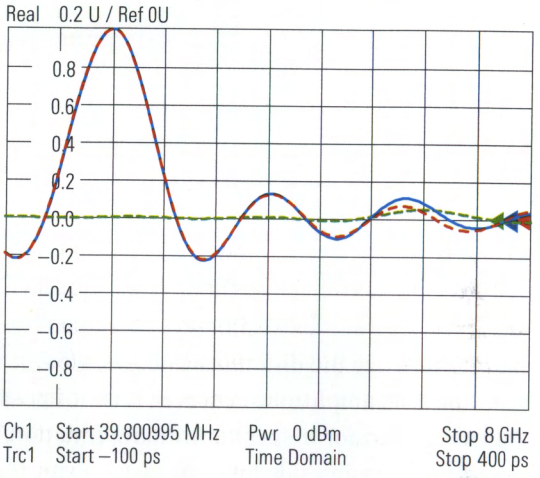


Fig. 5.2.11
Two pulses seperated by 300 ps and with a 26 dB level difference with rectangle window.

The smaller main pulse at $t = 300$ ps is covered by the side lobe of the pulse centered at $t = 0$ ps. Normally the small deviation in the (blue) measurement result (summation curve) cannot be seen, because the red curve necessary for comparison purpose is missing. The problem of pulse leakage is a well known phenomenon of numerical (inverse) Fourier transform. A typical technique to keep up with this problem is to use other window functions $W(f)$ different from the rectangle window. Any function fulfilling the following requirements can be used as $W(f)$:

- Real-valued function
- Function must exhibit a value of 0 outside $f = -F/2$ to $F/2$
- The overall integral of the function must have a value which avoids any unwanted numerical scaling of the time-domain values

The obvious thing to do is to shape the window so that its impulse response $w(t)$ has smaller side lobes than the si function. For example, take the Hann window; its side lobes are smaller by a factor of about 10 compared to the rectangle window. The following figure illustrates the

same situation with a Hann window. At the time sample $t = 300$ ps, the red trace has a magnitude that is less than 0.005. As a result, the impulses can be distinguished directly on the blue summation trace.

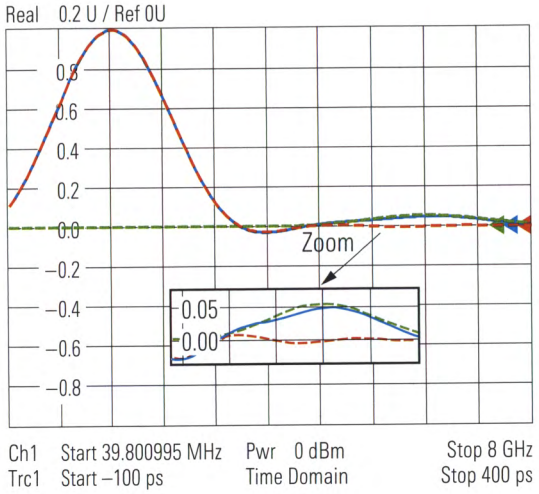


Fig. 5.2.12
Two pulses separated by 300 ps and with a 26 dB level difference with Hann window.

To characterize the different windows, we can use the value of the **highest side lobe amplitude**. However, it is not possible to give a general recommendation for a specific window function. A window function with rapidly decreasing side lobes always has a wide main pulse. Accordingly, two closely spaced pulses with a similar level can no longer be distinguished from one another.

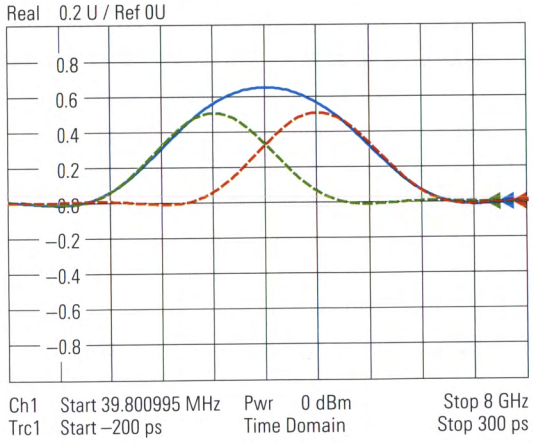


Fig. 5.2.13
Two pulses separated by 100 ps and of equal level with Hann window.

In Fig. 5.2.13 we used the Hann window. The two impulses at $t = 0$ ps and $t = 100$ ps cannot be distinguished from one another although

they have the same amplitude. The dashed traces are plotted for illustration purposes only. If, instead of the Hann window we use a rectangle window we obtain the time-domain trace shown below. The narrow main pulse of the si function makes it possible to distinguish the two pulses.

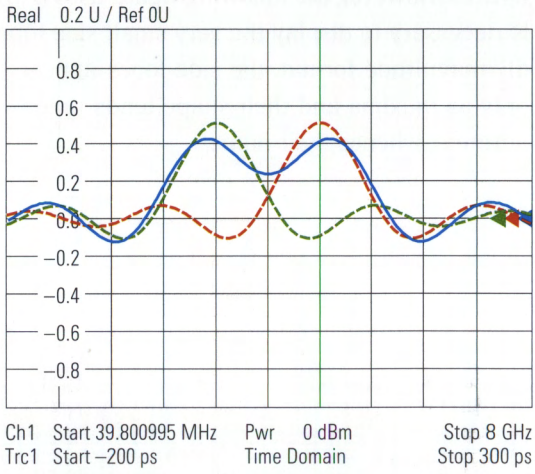


Fig. 5.2.14
Two pulses separated by 100 ps and with a 0 dB level difference with rectangle window.

Fig. 5.2.14 also shows clearly that the signs are important for superpositioned pulses (green and red traces). If an additional pulse is located at a side lobe, with a different sign and the size of the side lobe is half of the additional pulse then the only change at the location will be the sign; the magnitude of the superimposed side lobe will remain the same. In magnitude or dB display format, the additional pulse would be entirely hidden!

The time-domain response of the rectangle window has a 3 dB pulse width of $0.89 \cdot 1/F$. The 3 dB pulse width of the other window functions is also inversely proportional to the window width F, but they have factors that are different than 0.89. This makes it useful to compare the pulse widths of all of the window functions with that of the rectangle window. For this purpose, we can use the ratio formed in equation (5.2-17). It is known as the **relative impulse width** B_{IBRR} and takes no physical unit.

$$B_{\text{IBRR}} = \frac{\text{3 dB impulse width of the window of interest}}{\text{3 dB impulse width of the rectangle window}} \qquad \textit{(5.2-17)}$$

In section 5.5.2, the relative impulse width and the level difference with respect to the side lobe with the highest magnitude are listed for all of the common window functions. Direct comparison is possible with the following figure. With the previous figures the format "Real" part was used so that the sign information of the samples could be seen from the figures. However, the following figure uses a dB magnitude format. This is necessary to display the very small side lobes. Using a magnitude or dB magnitude format, the side lobes appear (regardless of their sign) only as maxima and their shape tends to look like the side lobes of an antenna, whence their name.

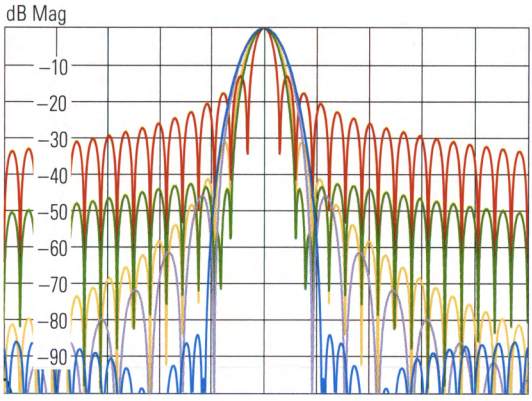


Fig. 5.2.15
Direct comparison of the different window functions. ***Red:*** *Rectangle window (no profiling)* ***Green:*** *Hamming window (low first side lobe)* ***Yellow:*** *Hann window (normal profile)* ***Violet:*** *Bohman window (steep falloff) and* ***Blue:*** *Dolph-Chebyshev window with user-selectable side lobe suppression (here, 80 dB).*

5.2.3 Bandpass mode

We are not always able to start a measurement at 0 Hz. A narrowband DUT such as the one shown in Fig. 5.2.16 has a non-zero frequency response $\Gamma_H(f)$ only in a limited range around its center frequency.

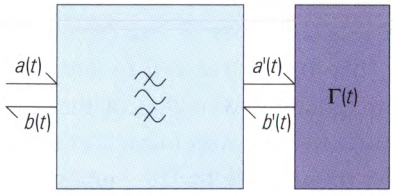


Fig. 5.2.16
DUT with a bandpass filter at the input.

To get a useful time-domain result, it is important to place the frequency samples inside the passband and its vicinity. As can be seen in the following figure, this is generally impossible using the frequency grid from formula (5.2-2).

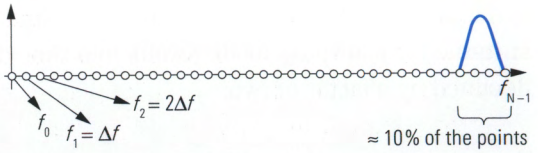


Fig. 5.2.17 Bandpass filter with harmonic grid.

In fact, we need a transformation that allows an arbitrary lower frequency limit. The corresponding transformation is known as **bandpass mode**, while the transformation described in sections 5.2.1 and 5.2.2 is known as **lowpass mode**. In bandpass mode, the frequency vector has a lower frequency limit $f_0 > 0$.

$$f = \begin{bmatrix} f_0 & f_1 & f_2 & \dots & f_n & \dots & f_{N-1} \end{bmatrix} \qquad \textbf{(5.2-18)}$$

where $f_n = f_0 + n\Delta f$ and $n = \{0 \dots N-1\}$ and $F = f_{N-1} - f_0$ **(5.2-19)**

In contrast to formula (5.2-1), the frequencies f_2, f_3, f_4, ... are not multiples of the frequency f_1 any more. Accordingly, we do not have a harmonic grid. The following figure shows an example similar to section 5.2.1. For the sake of simplicity it has only $N = 5$ points.

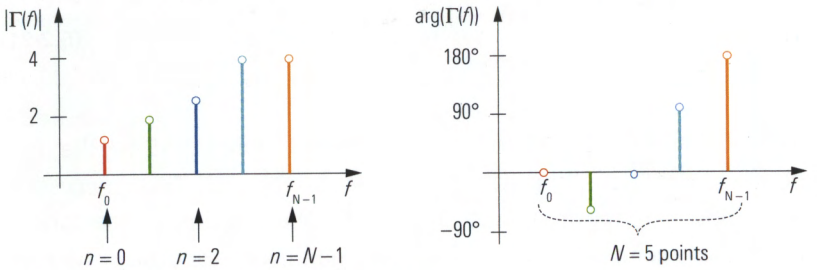


Fig. 5.2.18 Example of a bandpass spectrum.

In contrast to the lowpass mode (cf. Fig. 5.2.2) it is not possible to extend the frequency axis in the direction of $f < 0$. However it is necessary to arrange the spectrum centered around the frequency $f = 0$ Hz, which is done by a frequency shift of $f_C = (f_{N-1} + f_0)/2$. Within this operation the total number of frequency points remains the same, meaning $M = N$. With formula (5.2-6) we thus have the following time domain resolution:

$$\Delta t = \frac{1}{F} = \frac{1}{\Delta f \cdot N} \qquad \textbf{(5.2-20)}$$

Comparing formulas (5.2-20) and (5.2-6) we can draw the following conclusion: Applying the bandpass mode to a DUT that exhibits significant

transmission down to low frequencies and that is therefore much more suitable for a lowpass mode results in a time domain resolution that is declined by a factor of two.

To go on with our example, the measured and shifted values are periodically continued on both sides like shown in the following figure.

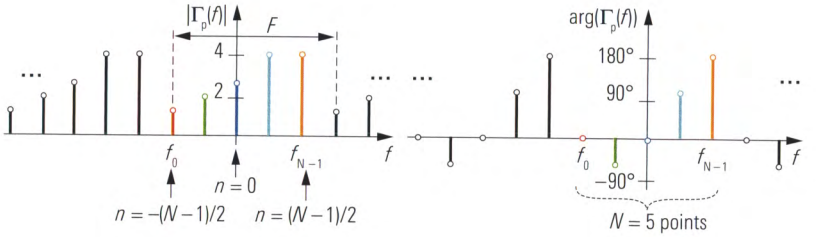


Fig. 5.2.19 Example of a periodically continued bandpass spectrum.

Similar to lowpass mode, windowing is also applied. The window function must be shaped to cover the range $-F/2 \ldots F/2$. Therefore, the rectangle window known from formula (5.2-9) has to be modified.

$$W(n) = \begin{cases} 1/F & \text{for} \left|n\right| \leq (N-1)/2 \\ 0 & \text{otherwise} \end{cases} \qquad \textbf{(5.2-21)}$$

After the numerical transformation based on formula (5.2-11), the time-domain result is multiplied by the complex factor $e^{j2\pi f_c t}$. This is necessary to compensate the frequency shift $f - f_C$. The computed time-domain response is complex and is known as the bandpass response $\Gamma_{bp}(\Delta t \cdot m)$. The magnitude $|\Gamma_{bp}(\Delta t \cdot m)|$ of the bandpass response is interpreted like an impulse response $\Gamma_h(t)$. However, further evaluation such as computation of the step response fails due to the complex values of $\Gamma_{bp}(\Delta t \cdot m)$. Regardless of whether we use bandpass or lowpass mode for the numerical transformation, the time-domain signal has the same period T in accordance with formula (5.2-14). In both cases, the ambiguity range is given by $-T/2$ to $T/2$.

Formula (5.1-5) defines the impulse response $\Gamma_h(t)$ as the response of a network stimulated with a Dirac impulse $\delta(t)$. We will try to find a similar definition for the bandpass response $\Gamma_{bp}(\Delta t \cdot m)$. If we want to determine the stimulus function $\zeta(t)$ associated with the bandpass response $\Gamma_{bp}(\Delta t \cdot m)$, it is necessary to retrace the computational steps "Window-

ing with $W(f)$" and "Frequency shift by $f - f_C$" in the time domain. They are given by the functions $w(t)$ and $e^{j2\pi f_C t}$ respectively. The stimulus $\zeta(t)$ associated with the bandpass response $\Gamma_{bp}(\Delta t \cdot m)$ consists of a **burst signal**. When using a rectangle window as $W(f)$, we obtain the burst signal as follows:

$$\zeta(t) = w(t)\, e^{j2\pi f_C t} = \frac{\sin(\pi \cdot F \cdot t)}{\pi \cdot F \cdot t}\, e^{j2\pi f_C t} \qquad \textit{(5.2-22)}$$

Fig. 5.2.20 shows an example of the real and imaginary parts of this burst signal $\zeta(t)$. The envelope of the signal $\zeta(t)$ is given by the function $w(t)$. The complex carrier oscillation has a frequency of f_C. Since $|e^{j2\pi f_C t}| = 1$, the envelope can be determined from the magnitude $|\zeta(t)| = |w(t)|$.

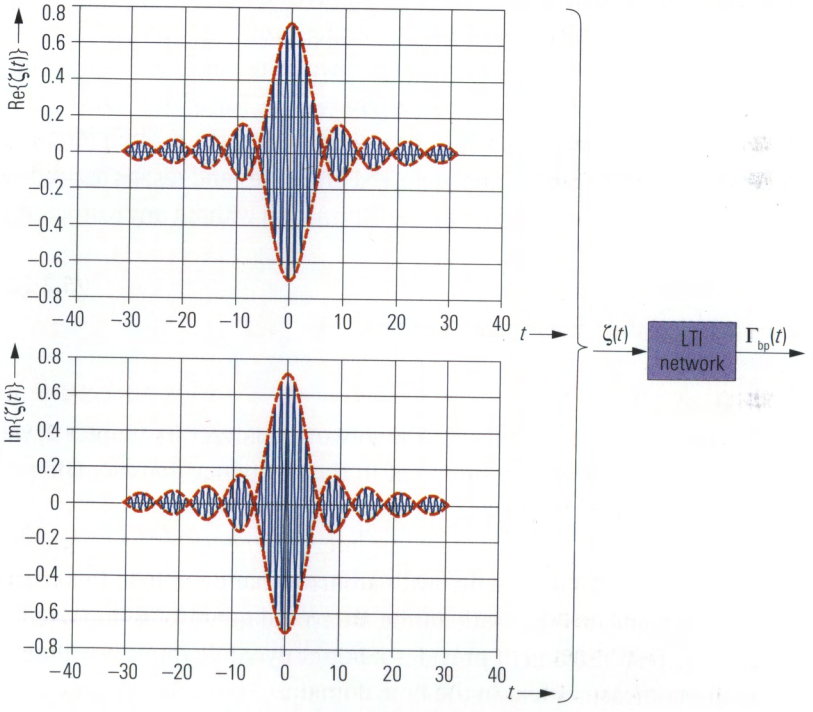


Fig. 5.2.20 Linear network with bandpass stimulus.

5.2.4 Transformations optimized for computing time

If we perform the computation of the time-domain quantity $\Gamma_h(t)$ in the order implied by the sum (5.2-11), then we will need to perform M^2 complex multiplications. Since multiplications are the most time-consuming operations, they serve as a good measure for the total computing time. If we choose the time step size Δt in accordance with formula (5.2-6) and the quantity M fulfills the requirement that $M = 2^x$ where $x \in \{1, 2, 3, \ldots\}$, then the term $e^{j2\pi nm/M}$ will exhibit certain symmetries and periodicities. This fact is exploited by the **inverse fast Fourier transform** (IFFT) to reduce the number of multiplication operations to $N \cdot \mathrm{ld}(N)$.[1] The best-known implementation of this is the Radix 2 algorithm described by Cooley and Tukey [CT65]. The requirement that $M = 2^x$ can be circumvented by filling up the data set with zero values. For the arbitrary time step size t_d presented in section 5.2.2, however, another implementation known as the **inverse chirp z-transform (CZT)** [RS69] is required. It consists of two IFFT operations and a convolution and thus requires slightly more computing time than an FFT. It is preferred for use in network analyzers. The properties and results derived in the preceding sections apply to this transformation without any restrictions.

5.3 Using the time-domain option

Motivation

In discussing the basic concept of a network analyzer in chapter 2, we focused exclusively on measurement in the frequency domain. The reasons are as follows:

- Unlike time-domain measurements, measurements in the frequency domain can be implemented with a narrow bandwidth. In this manner, the signal-to-noise ratio of the measured quantity is improved and the test instrument provides a better dynamic range compared to direct measurement in the time domain.
- For technical reasons analog to digital conversion is possible up to a frequency limit of a few GHz only. In a test instrument based on the heterodyne principle, this limit is irrelevant since the analog/digital converter (see section 2.6, Fig. 2.6.1) only has to satisfy the requirements of the intermediate frequency f_{IF}.

1) Here, $\mathrm{ld}(x)$ stands for the base 2 logarithm (logarithm dualis). It can be calculated, for example, as $\mathrm{ld}(x) = \lg(x)/\lg(2)$.

A vector network analyzer measures, both the amplitude and the phase of the S-parameters. Accordingly, the requirements for unambiguous transformation into the time domain are satisfied. Most manufacturers offer a **time-domain option** for their vector network analyzers.

Illustrative description of the time-domain transformation

The measured values recorded by the network analyzer describe a set of phasors. For the sake of simplicity, Fig. 5.2.6 shows a set of only five rotating phasors. They rotate at different frequencies and have their own amplitude and starting phase. From their projection on the real axis, four sinusoidal time-domain signals and one constant component are created. Based on the superposition of the time-domain signals, the signal $\Gamma(t)$ arises which is evaluated at the discrete-time points $n\Delta t$.

Numerical inverse Fourier transform

The measured values on the network analyzer are acquired over a discrete frequency axis with the frequency step size Δf. As a result, a time-domain signal with the period $T = 1/\Delta f$ can be computed. The values in the interval $-T/2 \ldots T/2$ encompass what is known as the ambiguity range, while the values outside this interval are irrelevant. However, if we do not choose a suitable period T, the adjacent values can distort the values in the ambiguity range. Any measurement using the network analyzer always has an upper frequency limit. This results in a discrete time axis with the limited resolution Δt. The same applies to the modified transform known as bandpass mode which can be used if measured values are available in a limited frequency segment only. In the latter case, the resolution Δt is determined by the bandwidth of the frequency segment. The limited time resolution can cause pulses to be superimposed and mask one another. If this happens, it is no longer possible to distinguish them. By choosing an appropriate window function, it is possible to vary the shape of the pulses and their superposition. For an example of this, see Fig. 5.2.11 through Fig. 5.2.14 in section 5.2.2.

5.3.1 Operation in lowpass mode

The linear frequency sweep that is set on the network analyzer covers the range f_{Start} to f_{Stop} with the frequency step size f_{Step} and a number of points M. First, we will assume that the frequencies f_{Start} and f_{Stop} are integer multiples of the step size f_{Step}. The start frequency f_{Start} always

has a minimum value $f_{min} > 0$ depending on the analyzer model. The DC value for $f = 0$ Hz cannot be measured with the heterodyne principle. It must be supplied to the analyzer or the analyzer has to estimate the value by extrapolation. It always has a phase value of 0°. Furthermore, it is often possible to derive the DC value by inspection. The following table provides some examples for simple circuit models:

DUT	DC value
Open or capacitance	$\Gamma(0) = 1$
Short or inductance	$\Gamma(0) = -1$
Resistor R or resistor R with parallel capacitor or resistor R with series inductance	$\Gamma(0) = \dfrac{R - Z_0}{R + Z_0}$

Table 5.3.1 *Typical DC Values*

For the lowpass transformation, a **harmonic grid** in accordance with formula (5.2-2) is required. In other words, all of the frequencies in the grid must be integer multiples of the frequency step size Δf. The first frequency sample in the grid is located at $f = 0$ and the second is located at $f_1 = \Delta f$. If $f_1 < f_{min}$, the frequency sweep is supplemented by K frequency samples in the range $f = 0$ to f_{min}. These additional frequency samples are shown in red in the following figure.

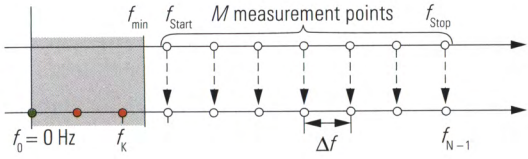


Fig. 5.3.1
Expansion of the harmonic grid.

The associated measured values in the range f_1 to f_K must be determined through interpolation. The frequency vector now encompasses a total of $N = M + K$ points. For frequencies $f < 0$, it is continued in accordance with formula (5.1-19) from section 5.1.4. It thus encompasses a total of $2N - 1$ frequency samples in the range from $-f_{N-1}$ to f_{N-1} with the frequency step size $\Delta f = f_{Step}$ (see Fig. 5.3.2). The frequency step size f_{Step} is usually not accessible as a parameter on the network analyzer. However, it can be calculated from the number of points M and the start and stop frequencies f_{Start} and f_{Stop} of the frequency sweep.

$$f_{Step} = \frac{f_{Stop} - f_{Start}}{M - 1} \qquad \textbf{(5.3-1)}$$

The following figure illustrates the relationships between the parameters on the network analyzer and the frequency axis used for the transformation.

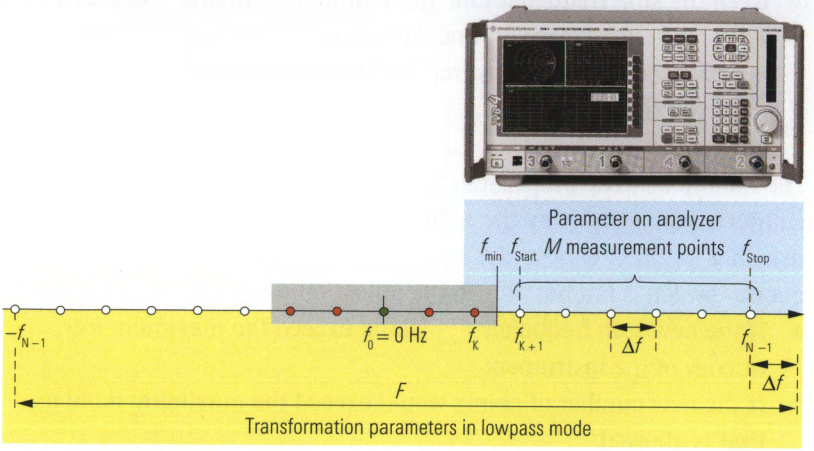


Fig. 5.3.2 Comparison of the parameters for the analyzer and the transformation in lowpass mode.

When using the lowpass transformation, the ambiguity range $-\frac{1}{2}T \ldots \frac{1}{2}T$ and the time resolution Δt are very important quantities. Based on the periodicity F of the spectrum in accordance with Fig. 5.3.2 and formula (5.2-6), we determine the time resolution as follows:

$$\Delta t = \frac{1}{F} = \frac{1}{2f_{Stop} + f_{Step}} \approx \frac{1}{2f_{Stop}} \qquad \textbf{(5.3-2)}$$

By replacing Δf of formula (5.2-14) with f_{Step} of formula (5.3-1), we obtain the quantity T.

$$T = \frac{1}{\Delta f} = \frac{1}{f_{Step}} = \frac{M-1}{f_{Stop} - f_{Start}} \approx \frac{M}{f_{Stop} - f_{Start}} \qquad \textbf{(5.3-3)}$$

If the start and stop frequencies are not integer multiples of the frequency step size f_{Step}, the frequency sweep can be modified using one of the following three rules:

a) **Keep stop frequency and number of points**
The number of points and the stop frequency is left unchanged the frequency step size is modified.

b) **Keep step width and number of points**
The number of points and the frequency step size is left unchanged, the stop frequency is reduced.

c) **Keep the stop frequency and approximate frequency step size**
The stop frequency is left unchanged and the number of points is increased in such a way that the frequency step size remains approximately the same

The modified sweep properties (f_{Step}, f_{Stop} and numbers of points) are calculated automatically by the network analyzer in this example. All the user has to do is choose the desired procedure. Adaptation of the frequency sweep is not possible in the following cases:

- If the new stop frequency f_{Stop} would exceed the maximum frequency of the instrument
- If the new number of points would exceed the maximum number that is allowed

A change in sweep properties can also affect the time resolution and the ambiguity range. We can differ three approaches "Rule a", "Rule b" and "Rule c". These rules can be compared in terms of the sweep time, time resolution and ambiguity range:

Criterion	Rule a	Rule b	Rule c
Sweep time	Unchanged	Unchanged	Increases
Time resolution	Unchanged	Decreases	Unchanged
Ambiguity range	Decreases	Unchanged	Increases

Table 5.3.2 *Influence of the three different rules*

Mathematically speaking, a reduction in the number of points automatically leads to a smaller ambiguity range. In most network analyzers, the number of points for the time-domain representation is linked to the number of points used in the frequency sweep. A reduction in the number of points in the frequency sweep thus also results in a reduced point grid in the time-domain representation. Accordingly, we obtain for $N < 100$ a time-domain trace with visible kinks. This loss of resolution is not related to formula (5.3-2)! The example traces shown in Fig. 5.3.3 are based on 201 and 21 points.

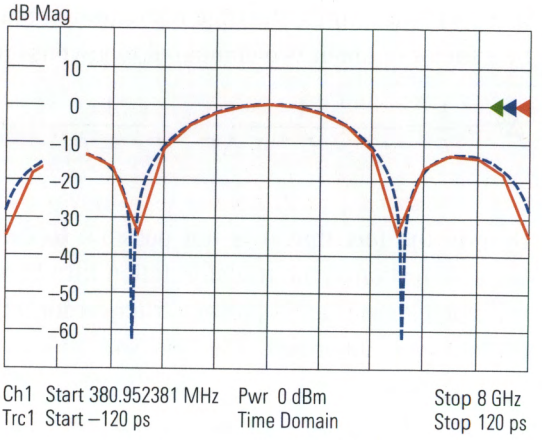


Fig. 5.3.3 Comparison of traces for 201 points (blue) and 21 points (red).

5.3.2 Operation in bandpass mode

Bandpass mode does not pose any particular requirements on the start and stop frequencies. The frequency grid only has to have a constant step size. In cases where the step response and the sign of the impulse response are not necessary and maximum resolution is not required, bandpass mode can be a convenient alternative to lowpass mode.

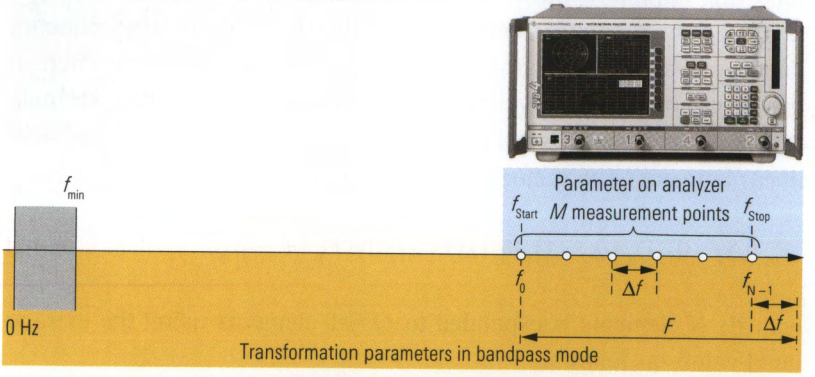


Fig. 5.3.4 Comparison of parameters for the analyzer and the transformation in bandpass mode.

As we can see in Fig. 5.3.4, the relationship between the parameters on the network analyzer and the frequency axis are much simpler than in lowpass mode. For the same frequency step size Δf, both modes have the same ambiguity range; see formula (5.3-3). Compared to lowpass mode, the quantity Δt is roughly doubled assuming comparable condi-

tions. In other words, the time resolution in bandpass mode is reduced by a factor of about two compared to lowpass mode.

$$\Delta t = \frac{1}{F} = \frac{1}{f_{\text{Stop}} - f_{\text{Start}} + \Delta f} \approx \frac{1}{f_{\text{Stop}} - f_{\text{Start}}} \qquad \textit{(5.3-4)}$$

Despite the fact that it is not possible to evaluate the phase of the impulse response, the transformation into bandpass mode is uniquely reversible. This is an important criterion for its use in conjunction with the time gate described in section 5.4.

5.3.3 Benefits of extrapolation

To obtain the best possible time resolution, we will generally use lowpass mode. The highest possible stop frequency should be selected, but this is limited by the maximum frequency of the network analyzer or the connector type that is used. Nevertheless, the time resolution obtained in this manner is insufficient in many cases. For example, using the N-type connector which is specified for a maximum frequency of 18 GHz, a **distance resolution** of only ≈ 8 mm is possible. We recommend the following approach: Based on the measured data recorded up to f_{Stop}, it is possible to predict the behavior of the DUT beyond this frequency limit in most cases. Linear prediction is particularly appropriate here. It makes use of all of the available measured data to derive the extrapolation. It is particularly useful for extrapolating the frequency response of linear networks. The vector of the measured values

$$\left[\Gamma_{\text{H}}(\Delta f \cdot 0), \Gamma_{\text{H}}(\Delta f \cdot 1), \Gamma_{\text{H}}(\Delta f \cdot 2), \ldots, \Gamma_{\text{H}}\left(f_{\text{Stop}}\right)\right] \qquad \textit{(5.3-5)}$$

with its N elements is expanded to $Q - 1$ elements using the extrapolated values as follows:

$$\left[\Gamma_{\text{H}}\left(f_{\text{Stop}} + \Delta f\right), \Gamma_{\text{H}}\left(f_{\text{Stop}} + 2\Delta f\right) \ \ldots \ \Gamma_{\text{H}}\left(\Delta f \cdot (Q-1)\right)\right] \qquad \textit{(5.3-6)}$$

The expansion factor $V = Q/N$ characterizes this expansion. In a calculation without this expansion, it has a value of $V = 1$. At the right edge of the extrapolation range, the accuracy of the extrapolation naturally decreases. On the other hand, the influence of the values at the edge is highly limited by the window function which tends towards 0 there.

Experiments with this technique have shown that a value of $V = 2 \ldots 8$ represents the optimum choice. It results in an improvement in the distance resolution by the same factor of 2 to 8. Theoretically, extrapolation is also possible on both sides of the measurement frequency vector. Accordingly, this technique can also be used to improve bandpass mode. The example in the following figure illustrates the efficiency of this technique. Here, a segmented transmission line with sections of 50 Ω, 34 Ω, 40 Ω and 50 Ω was created and measured in the frequency range 50 MHz to 24 GHz. The steps of the characteristic impedance along the transmission line are located at the electrical distances $d_1 = 14.25$ mm, $d_2 = 27.73$ mm and $d_3 = 35.67$ mm. They have their equivalent in the step response of the reflection coefficient.

The following calculation yields the deltas of the step response $\Delta\Gamma_\Theta(d)$ which we expect when measuring $\Gamma_\Theta(d)$ with reference to $d = 0$ and considering multiple reflections as far as they influence the size of the deltas (formulas see table 5.1.1).

$$\Delta\Gamma_\Theta(d_1) = \frac{34\Omega - 50\Omega}{34\Omega + 50\Omega} = -0.1905$$

$$\Delta\Gamma_\Theta(d_2) = \frac{2\sqrt{34\Omega \cdot 50\Omega}}{34\Omega + 50\Omega} \cdot \frac{40\Omega - 34\Omega}{40\Omega + 34\Omega} \cdot \frac{2\sqrt{34\Omega \cdot 50\Omega}}{34\Omega + 50\Omega} = 0.0781$$

$$\Delta\Gamma_\Theta(d_3) = \frac{50\Omega - 40\Omega}{50\Omega + 40\Omega} \cdot \frac{4\left(40\Omega \cdot 34\Omega\right)}{\left(40\Omega + 34\Omega\right)^2} \cdot \frac{4\left(34\Omega \cdot 50\Omega\right)}{\left(34\Omega + 50\Omega\right)^2} = 0.106$$

The step response in case of a non extrapolated transformation in low-pass mode $(V = 1)$ is shown in red in Fig. 5.3.5. The blue curve shows the result with linear prediction $(V = 8)$. Even a trained expert would not be able to guess the second step based on the red curve or even predict a definite value for this step. The measured steps of the blue curve agree well with the theoretical calculated deltas $\Delta\Gamma_\Theta(d)$ above.

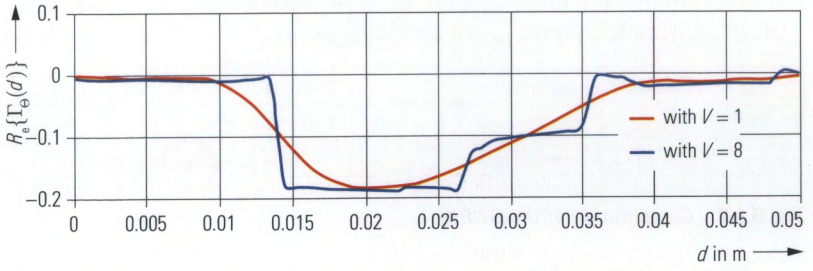


Fig. 5.3.5 Step response with (blue) and without (red) the prediction algorithm.

5.3.4 Processing sequence

In section 2.7.2, Fig. 2.7.3, we discussed the network analyzer's processing chain for measured data. The location of the time-domain transformation within the processing chain must be chosen carefully when designing a vector network analyzer. This becomes clear within the following examples. The paths shown using dashed lines cannot lead to any meaningful time-domain result for practical applications, so they are not supported.

Examples of formatting

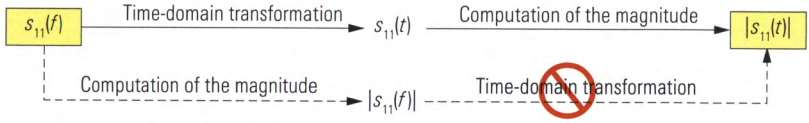


Fig. 5.3.6 Computation of the magnitude

Like all formatting operations, the magnitude is computed after the transformation. It would make no sense to compute the magnitude prior to the transformation since this would result in the loss of the phase information needed for the transformation. The following formatting options are important for the time-domain representation:

- The real part format since it provides information about the sign of the time-domain quantity. However, it is useful only in conjunction with lowpass mode.
- For bandpass mode, the magnitude or dB magnitude format is best. The latter is also used in lowpass mode when a larger value range is required than offered by the real part or magnitude.

Since the measured quantities in time-domain have real values, the phase in lowpass mode can only take on the values 0 and ±180°; in bandpass mode, the phase cannot be evaluated at all. The group delay τ_G derived from the phase is not available either.

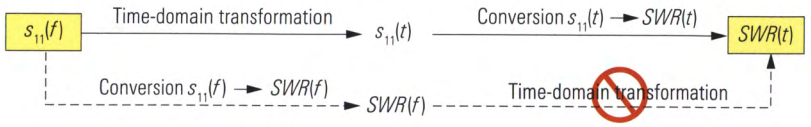


Fig. 5.3.7 Computation of the SWR(t).

The standing wave ratio (SWR) is also defined with respect to a time or distance axis. Similar to formula (4.3-2) from section 4.3, it is defined as:

$$SWR(t) = \frac{1+\left|\Gamma_h(t)\right|}{1-\left|\Gamma_h(t)\right|} \qquad \textit{(5.3-7)}$$

A time-domain transformation of the magnitude quantity $SWR(f)$ would lead to a different (incorrect) result. Computation of the magnitude and other nonlinear operations cannot be interchanged with the Fourier transform!

Example of trace mathematics and selection of the measured quantity

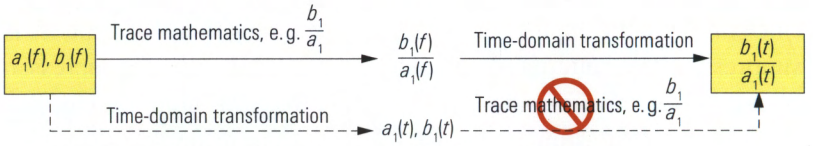


Fig. 5.3.8 Computation with trace mathematics.

5.4 Time gate

One-port devices often have a two-port network in front of their input. Using the deembedding technique described in section 4.9, it is possible to remove components within a cascaded system numerically. In this manner, the two-port network S_p in Fig. 5.4.1 can be eliminated. Of course, this requires information about either its topology and the component values or its S-parameters. Such information is simply unavailable in many cases.

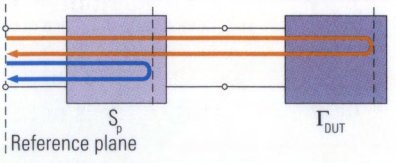


Fig. 5.4.1 Cascaded DUTs during a reflection measurement.

Practically all components including the two-port network S_p or the one-port device Γ_{DUT} have an electrical length that is not equal to 0. Due to the different distances to the reference plane of the network analyzer, the associated reflections (shown in Fig. 5.4.1 and Fig. 5.4.2 in blue and

orange) arrive at the test port at different times and can thus be distinguished from one an other.

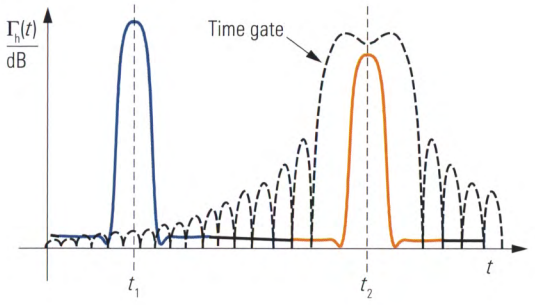


Fig. 5.4.2
Impulse response for the cascaded DUTs.

In time-domain representation, we can use a time gate to select a specific portion of the impulse response. The remaining impulse response is suppressed. As an alternative, the selected time range can be suppressed. In most cases, the time gate is configured in the time-domain since this provides the best estimation of its selectivity. For example, we can specifically select the impulse associated with the DUT Γ_{DUT} at $t = t_2$.

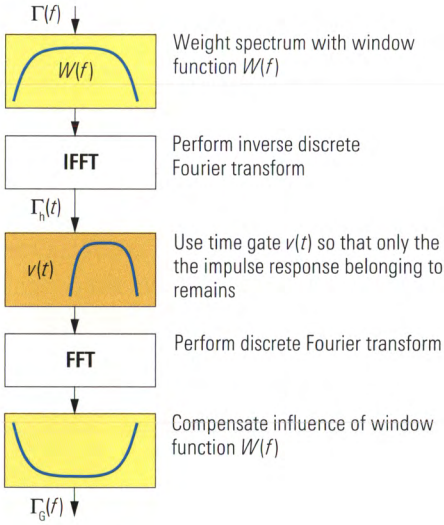


Fig. 5.4.3
Transformation sequence for the time gate.

For further evaluation, the signal weighted with the time gate is transformed back into the frequency domain. The re-transformed spectrum is distorted due to the window function $W(f)$ used at the start of the transformation sequence. This distortion is reversed using a suitable mathematical operation on the spectrum (see Fig. 5.4.3).

Usage of the time gate corresponds to multiplication of the impulse response $\Gamma_h(t)$ with the gate function $v(t)$. In the frequency domain, this corresponds to convolution of the measured spectrum $\Gamma(f)$ with the gate spectrum $V(f)$. The problems involved with the convolution are familiar from Fig. 5.2.9 ff. If we want to avoid corruption of the spectrum due to side lobes, the function $V(f)$ must not contain any side lobes. To meet this condition, however we need to accept side lobes (in the time domain) for the time gate $v(t)$. They determine the selectivity of the gate for signal components lying outside of the gate.

In Fig. 5.4.2, the gate with the steepest edges is shown. Due to its steep edges, pulses that are very close to one another can be clearly distinguished. However, the selectivity for signal components that are farther away is comparably worse. We can define different gate functions $v(t)$. The most important of these functions are listed in section 5.5.3 along with the relevant parameters. The following figure allows direct comparison of the gate functions $v(t)$.

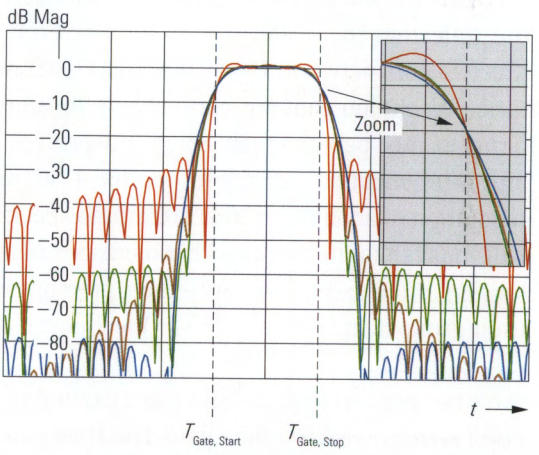


Fig. 5.4.4 *Comparison of different gate functions.* ***Red:*** *Steepest edges gate;* ***Green:*** *Steep edges gate;* ***Brown:*** *Normal gate;* ***Blue:*** *Gate with user-selectable side lobe suppression (here, 70 dB).*

The parameters $T_{\text{Gate,Start}}$ and $T_{\text{Gate,Stop}}$ designate the 6 dB points of the gate function. The definition of the quantities **passband ripple** and **side lobe suppression** can be found in the following figure. Similar to the case of the window functions, it is not possible to make a general recommendation for a specific time gate function.

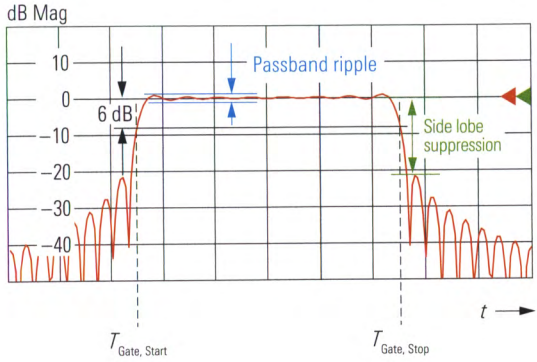


Fig. 5.4.5
Steepest edges.

Let us recall Fig. 5.4.1 from the start of section 5.4. Assume are that we would like to determine the reflection coefficient Γ_{DUT} as precisely as possible and that the two-port network S_p at the input is narrow banded. In such a case, the sweep range is chosen to be approx. 1.5 to 2 times the bandwidth of the two-port network. This ensures that the frequency range in which the reflection coefficient Γ_{DUT} is observed is not inadvertently limited by the sweep range. The useful frequency range F of the reflection coefficient Γ_{DUT} is thus determined by the bandwidth of the two-port network. Using formula (5.3-2), we can calculate the distance resolution that we can achieve to analyze the network components located behind the two-port network. For a bandwidth of 100 MHz and a propagation velocity of approx. $3 \cdot 10^8$ m/s, we obtain a distance resolution of about 1.5 m. Our attempt to specifically apply the time gate will thus fail for networks with dimensions of only a few centimeters. Filters are often exceptions since their propagation time is determined by their group delay instead of their mechanical dimensions. If the filter is made up of resonators with a high Q factor or it has a high group delay for other reasons (e.g. measurement example in section 6.2), it can make good sense to use the time gate. The time gate can also be used advantageously for transmission measurements. It serves as follows:

- As an alternative to hardgating (section 9.1) for distinguishing **multipath propagation** (see Fig. 5.4.6)
- For distinguishing the triple transit signal (measurement example in section 6.2).

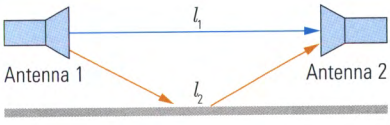


Fig. 5.4.6
Usage of the time gate in case of multipath propagation.

5.5 Tables and diagrams

5.5.1 Impulse and step responses for important reflection coefficients

Numerical computation of the Fourier transform results in a broadening of the impulses and smoothing of discontinuities. In addition, we obtain a discrete time axis with the step size t_d or Δt. The impulse and step response curves in Table 5.5.2 (see next page) must be interpreted with the aid of Table 5.5.1 when using the time-domain option. Table 5.5.1 provides a qualitative explanation of the difference between continuous and numerical transformation.

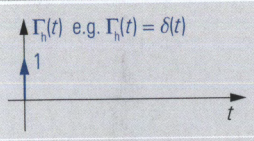
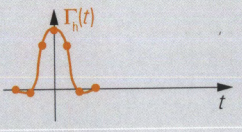
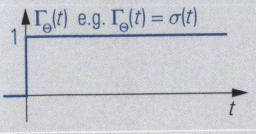
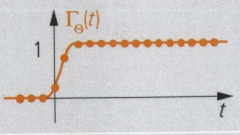
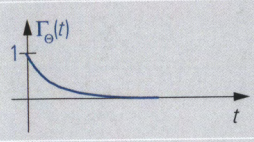
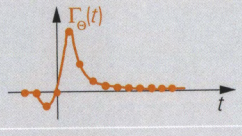
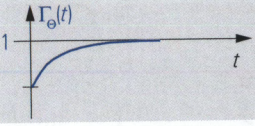
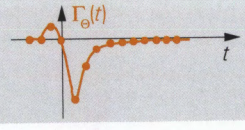
Continuous Fourier transformation	Numerical transformation	
		The time curve $w(t)$ of the window function takes the place of the Dirac impulse $\delta(t)$.
		The step function $\sigma(t)$ is smoothed by the window function. The step already begins at $t < 0$.
		The impulse response $\Gamma_\Theta(t) = e^{-t/\tau}$ is preceded by a virtual response for $t < 0$.
		The impulse response $\Gamma_\Theta(t) = -e^{-t/\tau}$ is preceded by a virtual response for $t < 0$.

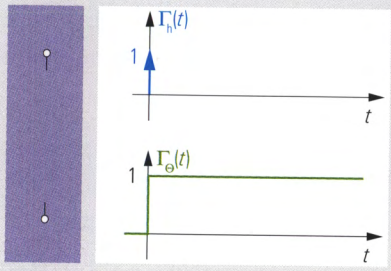
Table 5.5.1 Influence of the window function on the impulse and step response.

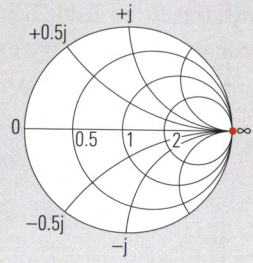
In the section below, some Fourier correspondences have been gathered together for use in time domain reflectometry. The associated complex reflection coefficients for $f > 0$ are also shown as locus curves in the Smith chart.

$\Gamma(t) \circ\!\!-\!\!\bullet\ \Gamma(f)$

1. Open (O)

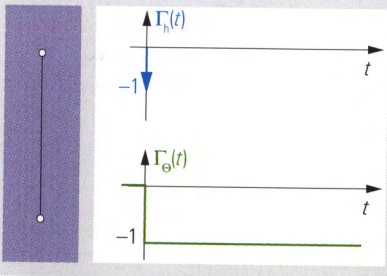
$\Gamma_h(t) = \delta(t)$
$\Gamma_\Theta(t) = \sigma(t)$ $\quad \circ\!\!-\!\!\bullet \quad$ $\Gamma(f) = 1$

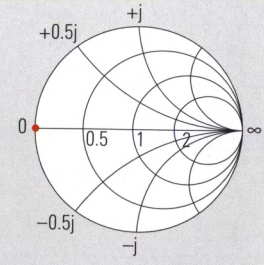




2. Short (S)

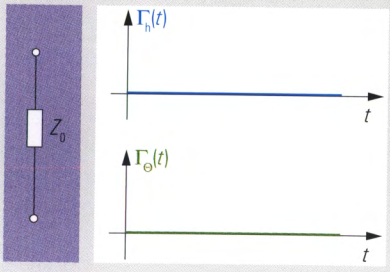
$\Gamma_h(t) = -\delta(t)$
$\Gamma_\Theta(t) = -\sigma(t)$ $\quad \circ\!\!-\!\!\bullet \quad$ $\Gamma(f) = -1$

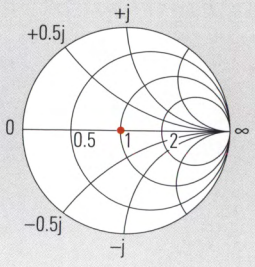




3. Match

$\Gamma_h(t) = 0$
$\Gamma_\Theta(t) = 0$ $\quad \circ\!\!-\!\!\bullet \quad$ $\Gamma(f) = 0$



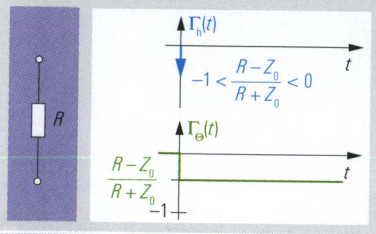


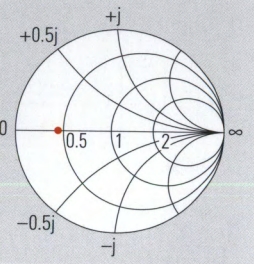
$\Gamma(t) \circ\!\!-\!\!\bullet\ \Gamma(f)$

4. Resistor (R) where $0 < R < Z_0 < \infty$

$$\Gamma_h(t) = \frac{R - Z_0}{R + Z_0}\delta(t) \qquad \circ\!\!-\!\!\bullet \qquad \Gamma(f) = \frac{R - Z_0}{R + Z_0} < 0$$

$$\Gamma_\Theta(t) = \frac{R - Z_0}{R + Z_0}\sigma(t)$$

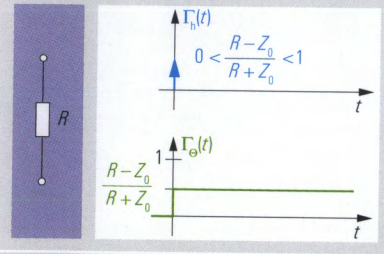


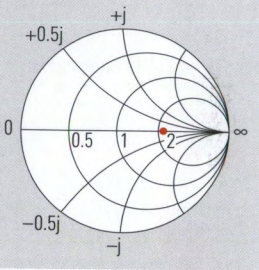


5. Resistor (R) where $\infty > R > Z_0 > 0$

$$\Gamma_h(t) = \frac{R - Z_0}{R + Z_0}\delta(t) \qquad \circ\!\!-\!\!\bullet \qquad \Gamma(f) = \frac{R - Z_0}{R + Z_0} > 0$$

$$\Gamma_\Theta(t) = \frac{R - Z_0}{R + Z_0}\sigma(t)$$

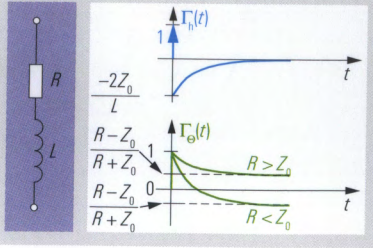


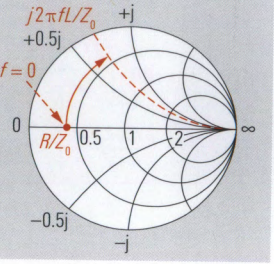


6. RL series connection

$$\Gamma_h(t) = -\frac{2Z_0}{L}e^{-t/\tau}\sigma(t) + \delta(t) \qquad \circ\!\!-\!\!\bullet \qquad \Gamma(f) = \frac{R + j2\pi fL - Z_0}{R + j2\pi fL + Z_0}$$

$$\Gamma_\Theta(t) = \left[\frac{R - Z_0}{R + Z_0} + \left(1 - \frac{R - Z_0}{R + Z_0}\right)e^{-t/\tau}\right]\sigma(t) \quad \text{where } \tau = \frac{L}{R + Z_0}$$



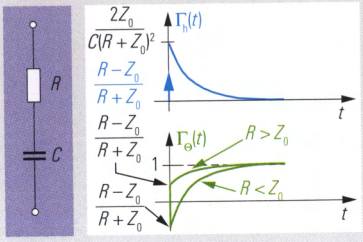


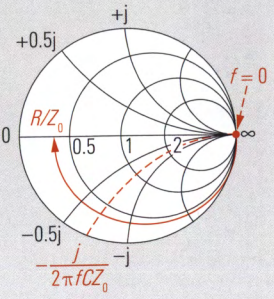
$\Gamma(t) \circ\!\!-\!\!\bullet\ \Gamma(f)$

7. RC series connection

$$\Gamma_h(t) = \frac{2Z_0}{C(R+Z_0)^2} e^{-t/\tau}\sigma(t) + \frac{R-Z_0}{R+Z_0}\delta(t) \quad \circ\!\!-\!\!\bullet \quad \Gamma(f) = \frac{j2\pi fC(R-Z_0)+1}{j2\pi fC(R+Z_0)+1}$$

$$\Gamma_\Theta(t) = \left[1-\left(1-\frac{R-Z_0}{R+Z_0}\right)e^{-t/\tau}\right]\sigma(t) \text{ where } \tau = (R+Z_0)C$$

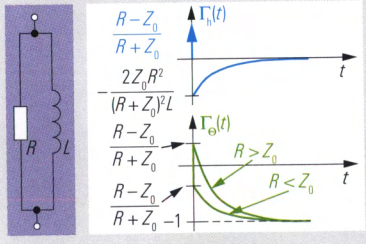


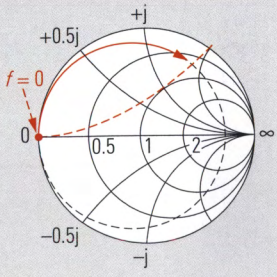


8. RL parallel connection

$$\Gamma_h(t) = -\frac{2Z_0R^2}{(R+Z_0)^2L} e^{-t/\tau}\sigma(t) + \frac{R-Z_0}{R+Z_0}\delta(t) \quad \circ\!\!-\!\!\bullet \quad \Gamma(f) = \frac{(R-Z_0)\,j2\pi fL - Z_0R}{(R+Z_0)\,j2\pi fL + Z_0R}$$

$$\Gamma_\Theta(t) = \left[\left(\frac{R-Z_0}{R+Z_0}+1\right)e^{-t/\tau} - 1\right]\sigma(t) \text{ where } \tau = \frac{(R+Z_0)L}{Z_0R}$$



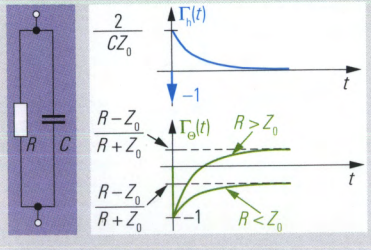


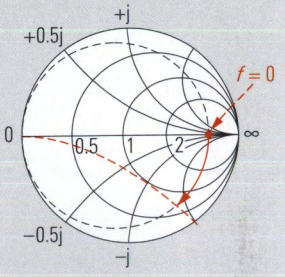
$\Gamma(t) \circ\!\!-\!\!\bullet\ \Gamma(f)$

9. RC parallel connection

$$\Gamma_h(t) = \frac{2}{CZ_0} e^{-t/\tau}\sigma(t) - \delta(t) \quad \circ\!\!-\!\!\bullet \quad \Gamma(f) = \frac{R - Z_0 - j2\pi fCRZ_0}{R + Z_0 + j2\pi fCRZ_0}$$

$$\Gamma_\Theta(t) = \left[\frac{R-Z_0}{R+Z_0} - \left(\frac{R-Z_0}{R+Z_0}+1\right)e^{-t/\tau}\right]\sigma(t) \quad \text{where } \tau = \frac{RZ_0C}{R+Z_0}$$





10. Transmission line, terminated with R

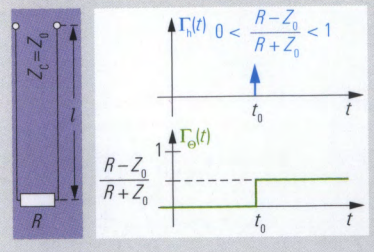
where $R > Z_0$ and $Z_C = Z_0$, all impedances real

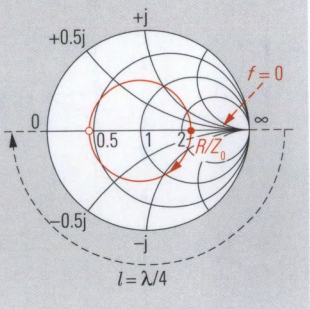
$$\Gamma_h(t) = \frac{R - Z_0}{R + Z_0}\delta(t - t_0) \quad \circ\!\!-\!\!\bullet \quad \Gamma(f) = \frac{R - Z_0}{R + Z_0} e^{-j4\pi l/\lambda}$$

$$\text{where } \lambda = \frac{c_0}{f\sqrt{\varepsilon_{r,\text{eff}}}}$$

$$\Gamma_\Theta(t) = \frac{R - Z_0}{R + Z_0}\sigma(t - t_0)$$

$$\text{where } t_0 = \frac{2l}{\lambda} = \frac{2l \cdot \sqrt{\varepsilon_{r,\text{eff}}}}{c_0}$$





$\Gamma(t)$ o—• $\Gamma(f)$

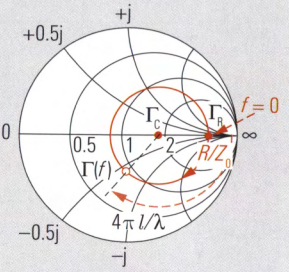
11. Mismatched transmission line o—•
where $Z_0 < R < Z_C$
and R, Z_C, Z_0 real

The Smith chart is uniformly referred to Z_0 which is why we use Γ_R and Γ_C.

$$\Gamma_R = \frac{R - Z_0}{R + Z_0};\ \Gamma_C = \frac{Z_C - Z_0}{Z_C + Z_0}$$

Here, the reflection coefficients Γ_1 and Γ_2 are referred to the impedance Z_0 or Z_C located prior to the reflection point in the direction of propagation.

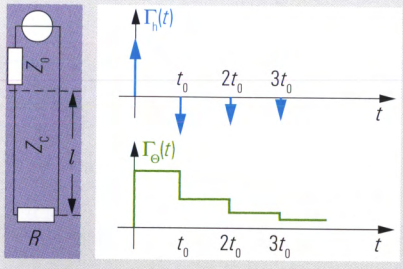
$$\Gamma_1 = \frac{Z_C - Z_0}{Z_C + Z_0};\ \Gamma_2 = \frac{R - Z_C}{R + Z_C}$$



where λ, t_0, see no. 10

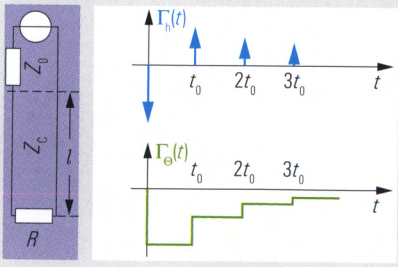
$$\Gamma_h(t) = \Gamma_1\delta(t) + \\ + (1 + \Gamma_1)\Gamma_2(1 - \Gamma_1)\delta(t - t_0) + \\ + (1 + \Gamma_1)\Gamma_2^2(-\Gamma_1)(1 - \Gamma_1)\delta(t - 2t_0) + \ldots$$

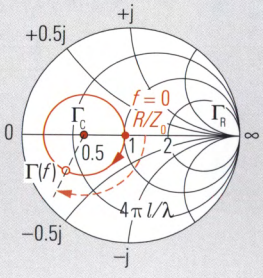
$$\Gamma_\Theta(t) = \Gamma_1\sigma(t) + \\ + (1 + \Gamma_1)\Gamma_2(1 - \Gamma_1)\sigma(t - t_0) + \\ + (1 + \Gamma_1)\Gamma_2^2(-\Gamma_1)(1 - \Gamma_1)\sigma(t - 2t_0) + \ldots$$



12. Mismatched transmission line, matched for $t \to \infty$ o—•
$Z_0 > Z_C$; $Z_0 = R$ and R, Z_C, Z_0 real
Formulas like no. 11!

The Smith chart is uniformly referred to Z_0, which is why we use Γ_R and Γ_C
Formulas like no. 11!

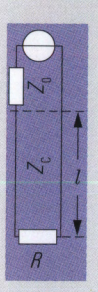


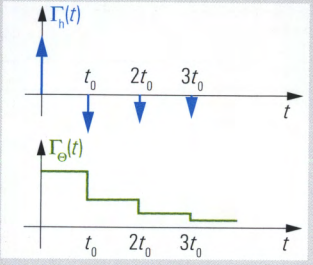


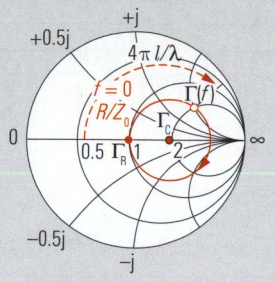
$\Gamma(t)$ o—• $\Gamma(f)$

13. Mismatched transmission line o—•
but matched for $t \to \infty$
$Z_0 < Z_C$; $Z_0 = R$ and R, Z_C, Z_0 real
Formulas like no. 11!

The Smith chart is uniformly referred to Z_0, which is why we use Γ_R and Γ_C.
Formulas like no. 11!



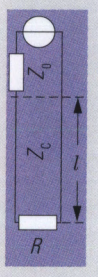


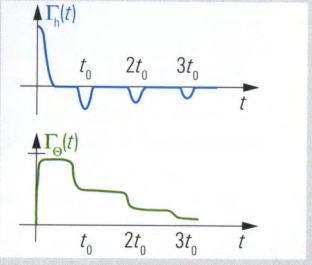


14. Lossy transmission line o—•
$Z_0 < Z_C$; $Z_0 = R$ and R, Z_0 real

Deformation of curve shapes for $\Gamma_h(t)$ and $\Gamma_\Theta(t)$ results from frequency-dependent losses. This represents a typical case of linear distortion.

The Smith chart is uniformly referred to Z_0, which is why we use Γ_C.





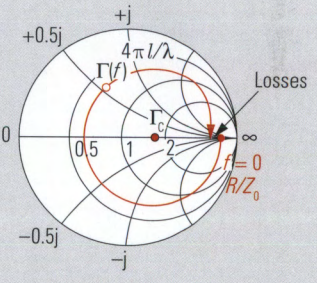
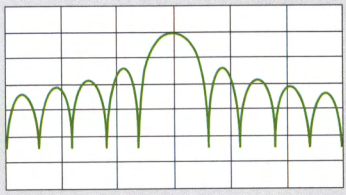
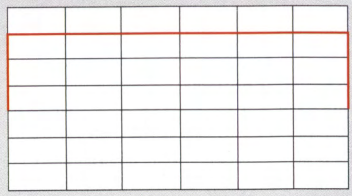
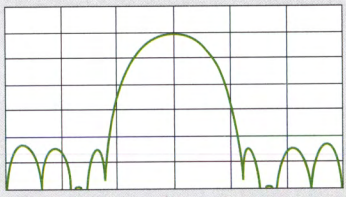
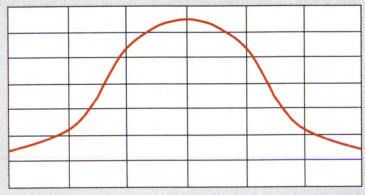
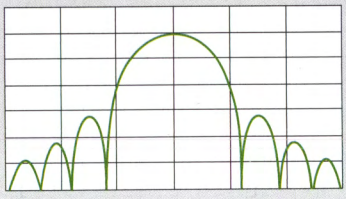
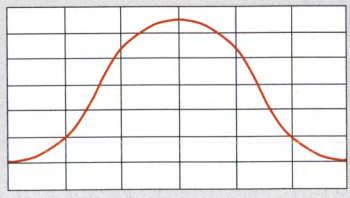


Table 5.5.2 Impulse and step responses of the reflection coefficient for important network types.

5.5.2 Comparison of important window functions

Amplitude shape (dB) in the time domain Characteristics in the time domain	**Name of window** Shape in frequency domain
 Highest side lobe: –13 dB Relative width of main pulse: 1.0 *Well suited for distinguishing close-spaced pulses with about the same amplitude.*	**Rectangle window, no profiling** 
 Highest side lobe: –43 dB Relative width of main pulse: 1.4	**Hamming window, low first side lobe** 
 Highest side lobe: –32 dB Relative width of main pulse: 1.6 *Good compromise between rectangle window, and Bohman window.*	**Hann window, normal profile** 

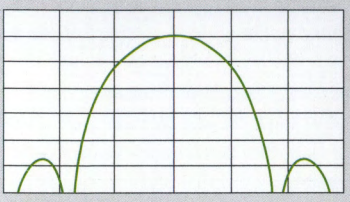
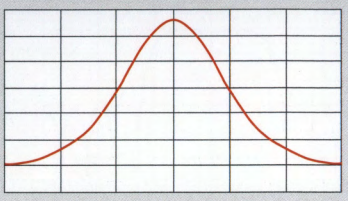
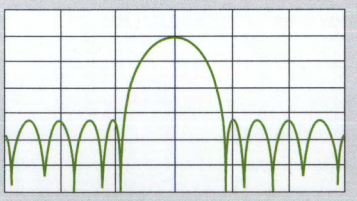
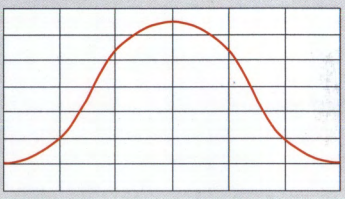
Amplitude shape (dB) in the time domain Characteristics in the time domain	**Name of window** Shape in frequency domain
 Highest side lobe: –46 dB Relative width of main pulse: 1.9 *Good distinction between wide-spaced pulses with very different amplitudes.*	**Bohman window, steep falloff** 
 Highest side lobe: –32 dB Relative width of main pulse: 1.2 *The values cited above are only examples since the side lobe suppression can be specified as a parameter.*	**Dolph-Chebyshev window, arbitrary side lobes** 

Table 5.5.3 *Comparison of important window functions.*

5.5.3 Comparison of important time gates

In analogy with the window functions $w(t)$, it is possible to define the gate functions $v(t)$. In the case of the window functions, we have sought for a main pulse that is as narrow as possible. In contrast, the width and center time of a gate function must be freely configurable. The time curves $v(t)$ shown below are only intended as examples. The values for the side lobe suppression are only minimum values. The resulting spectra $V(t)$ are complex and are thus not included in the table.

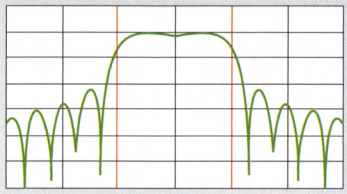
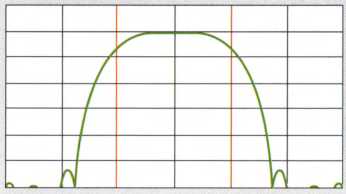
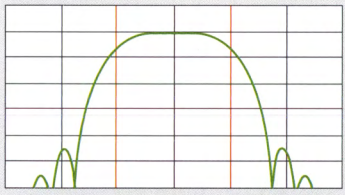
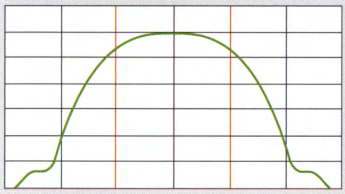
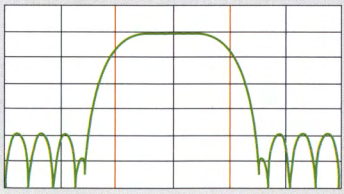
Name of gate function Shape in the time domain (dB scale)	Characteristics in the time domain *Information about usage*
Steepest edges gate 	Highest side lobe: –13 dB Passband ripple: 0.547 dB *Suppression of irregularities that are located relatively close to the gate boundaries and which have a lower level than the signal present in the gate region.*
Steep edges gate 	Highest side lobe: –43 dB Passband ripple: 0.019 dB *Good compromise between edge steepness (selection in the near range) and side lobe suppression (selection in the far range).*
Normal gate 	Highest side lobe: –32 dB Passband ripple: 0.032 dB *Good compromise between edge steepness (selection in the near range) and side lobe suppression (selection in the far range).*
Maximum flatness gate 	Highest side lobe: –46 dB Passband ripple: 0 dB *Maximum suppression outside the gate range, provided there is sufficient distance to the gate boundaries.*
Arbitrary gate shape 	Highest side lobe: –120 dB to –10 dB Passband ripple: 0.071 dB *Individual configuration of the side lobe suppression is possible.*

Table 5.5.4 *Comparison of important gate functions.*

5.5.4 Diagram for determination of the ambiguity range

The following diagram can be used to determine, based on the frequency step size Δf, the quantity $T = 1/\Delta f$ of the ambiguity range $-\frac{1}{2}T$ to $\frac{1}{2}T$, e.g. $\Delta f = 100$ MHz → $T = 10$ ns. Alternatively, for the transformation in bandpass mode, the time resolution $\Delta t = 1/F$ can be read off from the frequency interval F which corresponds more or less to the span, e.g. $F = 1000$ MHz → $\Delta t = 1$ ns.

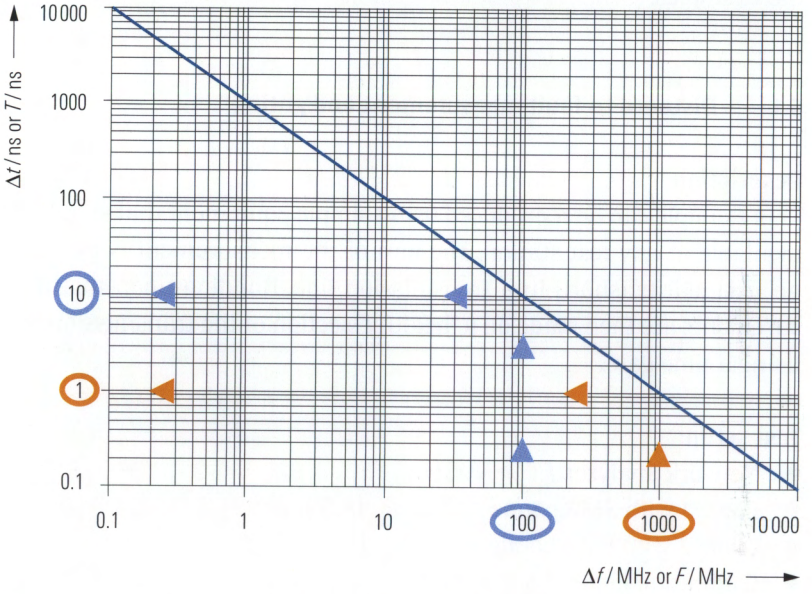


Fig. 5.5.1 Double-logarithmic display of the function 1/f.

6 Examples of time-domain measurements

Now, we will have a look at some typical measurements that are done using the time-domain transformation discussed in chapter 5. 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.

6.1 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. 6.1.3)
- 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:
   - Stop frequency: $f_{Stop} = 8$ GHz (4 GHz if necessary)
   - Start frequency: $f_{Start} = 20$ MHz (10 MHz if $f_{Stop} = 4$ GHz)

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).

- Number of points: $N = 400$ points
- Test port output power: -10 dBm
- Measurement bandwidth: 1 kHz
- Measured quantity: s_{11}
- 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. 6.1.1).



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 like described in section 5.3.1 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).

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: $l_{\text{mech}} = l_1 = 1.02$ m
6. Connect the cable to test port 1. Leave the other end of the cable open (do not install an open standard there).
7. Configure the time axis as follows:
   - Start time: $t_{\text{Start}} = -2$ ns
   - Stop time: $t_{\text{Stop}} = 18$ ns
8. The measurement result you obtain should be similar to that shown in Fig. 6.1.1. 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.

$$\frac{v_p}{c_0} = \frac{2l_{mech}}{t_p c_0} = \frac{2 \cdot 1.02\ \text{m}}{9.679 \cdot 10^{-9}\,\text{s} \cdot 2.99792458 \cdot 10^{8}\,\text{m/s}} = 0.703 \qquad \textbf{(6.1-1)}$$

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 $\tau_p = 2d/v_p$. When displaying reflection quantities with respect to a distance axis, most analyzers take into account the relationship $d = t \cdot v_p/2$, whereas they compute $d = t \cdot v_p$ for transmission quantities.

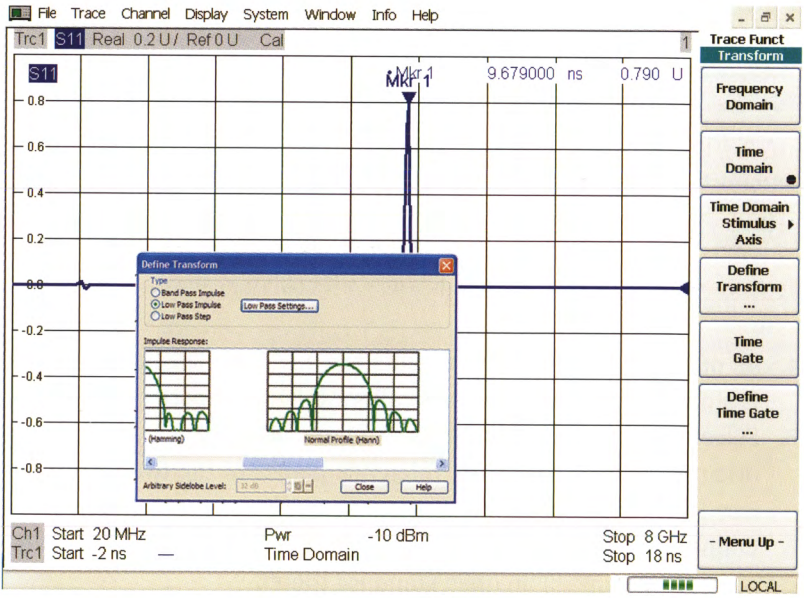


Fig. 6.1.1 Delay for an open transmission line.

9. Enter the calculated velocity factor on the network analyzer (see Fig. 6.1.2) and switch to distance display.

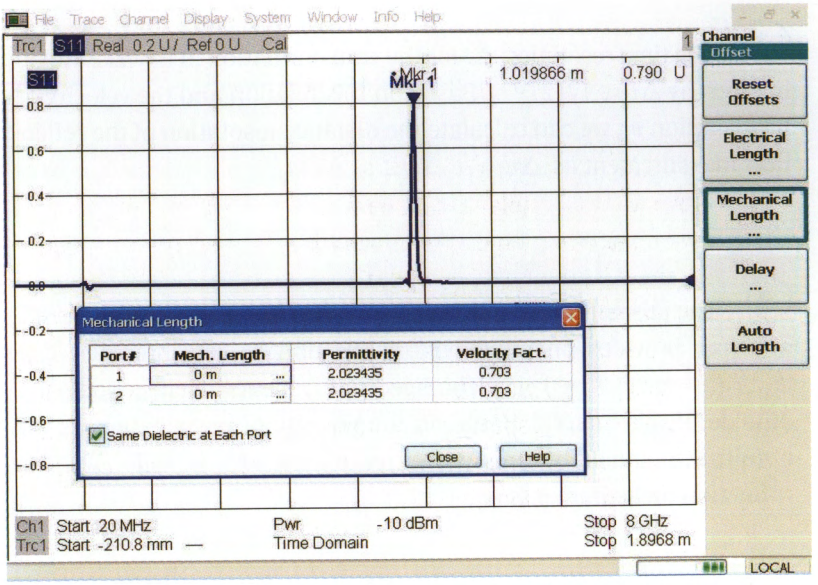


Fig. 6.1.2 Verification of the length measurement.

10. The displayed marker value (Fig. 6.1.2) should correspond to the mechanical length l_{mech} measured in step 5 of part 1.

Part 2 Locating and masking irregularities

A short-circuited T-junction (see Fig. 6.1.3) is used as our irregularity. It is located between two transmission lines with mechanical lengths l_1 and l_2.



Fig. 6.1.3 Test setup: Transmission line with a short-circuited 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. 6.1.3.

The time resolution Δt for the transformation in lowpass mode is given by $\Delta t \approx 1/(2f_{Stop})$. Based on this relation and the velocity of propagation v_p, we can calculate the distance resolution of the reflection measurement as $\Delta d = v_p \Delta t/2 \approx v_p/(4f_{Stop})$.

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

- Check the stop frequency
  $\Delta t = 1/(2 \cdot 8\ \text{GHz}) = 62.5\ \text{ps}$, i.e. $\Delta d = v_p \Delta t/2 = 0.66\ \text{cm}$
  $\Delta t = 1/(2 \cdot 4\ \text{GHz}) = 125\ \text{ps}$, i.e. $\Delta d = v_p \Delta t/2 = 1.31\ \text{cm}$
  Distance resolution here 0.66 cm at $f_{Stop} = 8$ GHz
  or 1.31 cm at $f_{Stop} = 4$ GHz
- Check the frequency step size
  Frequency range: 0 Hz to 8 GHz
  400 measurement points from 20 MHz to 8 GHz $\Rightarrow N = 401$
  $T = 1/\Delta f = (N-1)\ /\ f_{Stop} = 400/8\ \text{GHz} = 50\ \text{ns}$
  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.
- 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. 6.1.3 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. 6.1.4.
4. Consider whether you could optimize the time-domain transformation in the present case (Fig. 6.1.4) using a different window function.

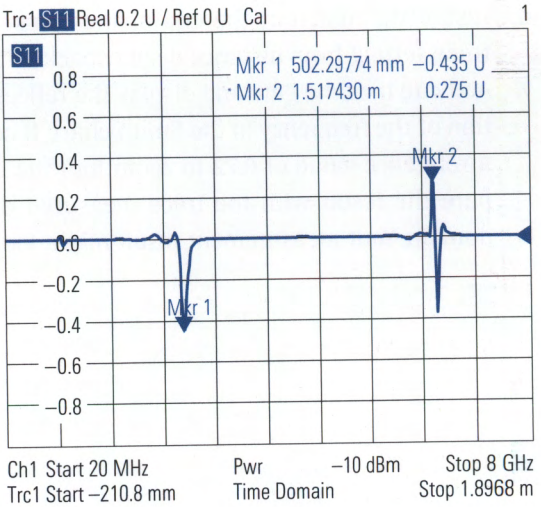


Fig. 6.1.4
Searching for the irregularity and the open end of the transformation line.

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 side-lobe suppression of, say, 80 dB.

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.
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. 6.1.5.

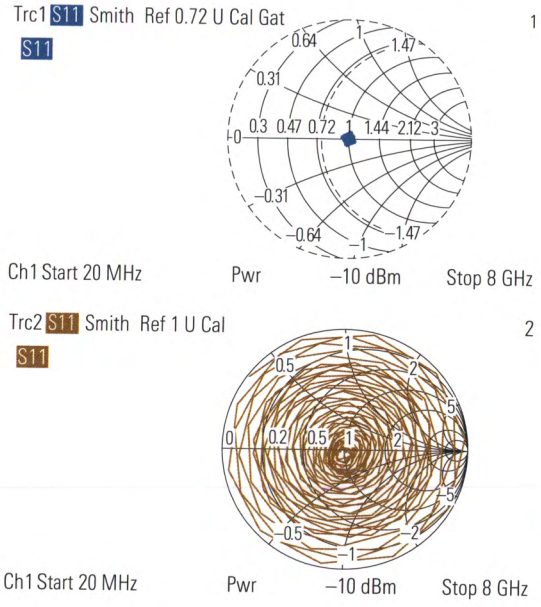


Fig. 6.1.5
Complex reflection coefficient for Fig. 6.1.3 with and without the time gate.

6.2 Measurements on a SAW filter in the time domain

The first time we used a SAW filter was in section 4.14. Now we will have a closer look at this type of filter and their functional principle because time-domain transformation is an ideal tool to optimize these filters.

Surface acoustic wave (SAW) filters consist of two interdigital transducers (IDIs) that are both arranged on a piezoelectric substrate. Such filters are offered in a center-frequency range from $f_0 = 10$ MHz to 2.5 GHz. Bandwidths from 200 kHz to $0.6\, f_0$ with an insertion loss of 2.5 dB to 30 dB are conventional. Adjustment of the filters is not required due to their good reproducibility. They can be manufactured in large quantities at low cost per unit. As noted in the functional description below, the input and output are DC-isolated.

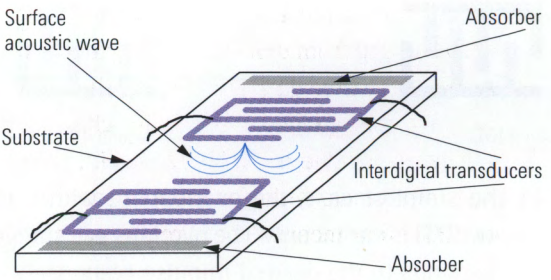


Fig. 6.2.1
Basic principle of a surface acoustic wave filter.

In the first interdigital transducer, an electrical signal launches a surface acoustic wave. This acoustic wave mode propagates on the surface between the substrate and the surrounding medium. It is also known as a Rayleigh wave and from the field of geophysics (seismic research). In the second interdigital transducer, the incident surface acoustic wave generates a piezoelectric output signal. It can be tapped as a balanced signal, for example. The propagation speed of the surface acoustic wave is equal to approx. 3000 m/s, which is slower than an electromagnetic wave by a factor of 10^{-5}. This means that these filters have a relatively high group delay compared to other RF components!

Surface acoustic wave filters with a center frequency $f_0 < 500$ MHz have a relatively high group delay of $\tau_G \approx 0.5$ µs to 4 µs. Accordingly, it is important to have a larger ambiguity range (factor 6 to 8 greater than τ_G). This is achieved by a large number of points and a frequency span that is aligned to the passband of the filter, requiring a bandpass mode transformation. In bandpass mode, only the magnitude information for the impulse response is available. Due to the required dynamic range, it should be displayed on a dB scale.

The desired frequency response of the filter is achieved through variation of the overlapping widths of the transducer fingers (see Fig. 6.2.2).

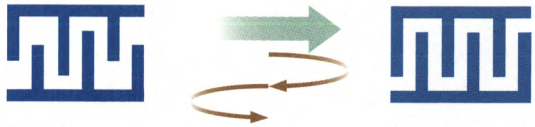


Fig. 6.2.2
Triple transit signal and geometry of the IDTs.

In the simplest case, the overlapping width of one interdigital transducer (IDT) is varied only. The overlaps are staggered so that their shape corresponds to the desired impulse response $s_{21h}(t)$ of the filter. As discussed in section 5.1.4, the impulse response can be calculated from the given frequency response $s_{21H}(f)$. In reality, however, this greatly simplified design description is problematic due to various non-ideal conditions. Most of these conditions can be assessed in the time domain:

- The first IDT launches a surface acoustic wave that travels to the second IDT and generates a piezoelectric output voltage. This voltage launches another surface acoustic wave that travels backwards to the first IDT. There, the same regenerative effect occurs so that the wave is sent back to the second IDT again (see Fig. 6.2.2). After passing between the IDTs three times, the wave is finally tapped at the filters output. Therefore, the impulse response $s_{21h}(t)$ shows a phantom signal at the triple group delay $3\tau_G$. It is called **triple transit signal**.
- Besides acoustic wave propagation, **direct electromagnetic coupling** is unavoidable due to the close arrangement of the IDTs. This direct electromagnetic coupling is typically attenuated by 40 dB to 60 dB compared to the passband. Due to the significantly higher electromagnetic propagation speed, it occurs in the impulse response at the time $t \approx 0$ s.
- To some extent, the surface acoustic wave launched at the first IDT is mode converted into a **bulk propagation mode**. This propagation mode can be coupled in a reverse procedure into the second IDT and impact the filters output signal. In most cases, multiple bulk modes can propagate. The delay of the bulk propagation mode exceeds that of the useful signal by a factor of 1 to about 3.

Based on the impulse response $s_{21h}(t)$, we can distinguish the following:

1. The group delay τ_G of the filter
2. The triple transit signal at $t = 3\tau_G$
3. The direct electromagnetic coupling at $t \approx 0$ s
4. The coupling due to bulk propagation modes at $t \approx \tau_G$ to $3\tau_G$

This is due to the fact that they arrive at the filter output at different time t (see also Fig. 6.2.3).

Test setup

- Network analyzer
- Two cables with SMA connectors for the DUT
- Calibration kit, PC 3.5 system
- DUT (surface acoustic wave filter), $f_0 = 160$ MHz, B = 30 MHz with unbalanced input/output 50 Ω

Procedure

1. Connect the test cables to the analyzer.
2. Configure the network analyzer as follows:
   - Center frequency: f_0, here: 160 MHz
   - Span: $^5/_3$ B, here: 50 MHz
   - Number of points: N = 801 (more if required!)
   - Test port output power: 10 dBm
   - Measurement bandwidth: 100 Hz
   - Measured quantity: s_{21}
   - Format: dB Mag
3. Connect the DUT and first check whether the passband is correctly centered around f_0.
4. Activate the time-domain transformation using the following settings:
   - Bandpass mode and Hamming window
   - Start time: $t_{Start} = -1$ µs, Stop time: $t_{Stop} \approx 4\tau_G$, here: 4 µs
5. Perform a two-port calibration (see section 4.1) and then reconnect the DUT.
6. Check the time resolution and ambiguity range based on the analyzer settings:
   - Time resolution: $\Delta t \approx 1/50$ MHz = 20 ns
     This value is equal to t_{Stop} divided by a factor of 200 and results in good display properties.

- Ambiguity range $T = 1/\Delta f = 801/50\ \text{MHz} \approx 16\ \mu\text{s}$
  The ambiguity range from $-8\ \mu\text{s}$ to $8\ \mu\text{s}$ is sufficient since $8\ \mu\text{s} \approx 6.5\tau_G$.

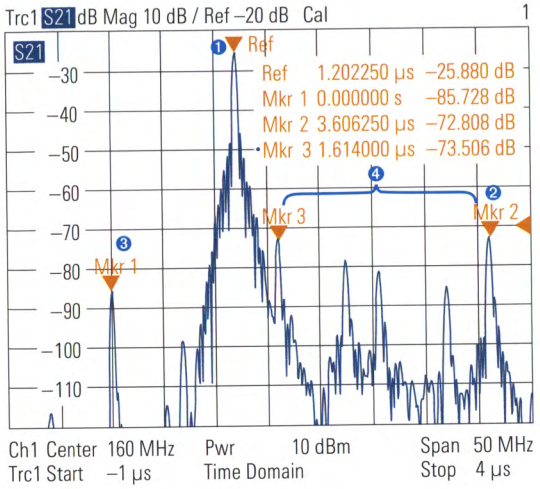


Fig. 6.2.3
Time-domain display for the surface acoustic wave filter.

7. Use the delta markers to identify the secondary maxima of the transmission. The result you obtain might look something like Fig. 6.2.3. The blue numbers are in reference to the information box (i) found above.
8. We would like to get an impression of how the filter would behave without the non-ideal properties referenced above (numbers ❷ to ❹). First, configure the time gate in the time domain representation. This time gate is easily configured in the time domain representation where its adjustment can be controlled visually. You will need to adjust the gate so that the main pulse and its side lobes are within the gate. Here, we made the following settings:
   - Left edge of the gate: $\text{T}_{\text{Gate, Start}} = 700\ \text{ns}$
   - Right edge of the gate: $\text{T}_{\text{Gate, Stop}} = 1.57\ \mu\text{s}$
   - Gate shape: Hann.

   Once you have completed the adjustment process, switch to frequency-domain display (see Fig. 6.2.5 upper part).
9. To assess the influence of the non-idealities within the frequency domain a comparison measurement without gating and time domain transformation is helpful. So, create an additional diagram area and measure the transmission s_{21} directly (see Fig. 6.2.5 lower part). By comparing Trc1 and Trc2 you will find out that Trc2 has an additional ripple on it (select a span of 5 MHz or see the zoomed trace that has been added by layout means).

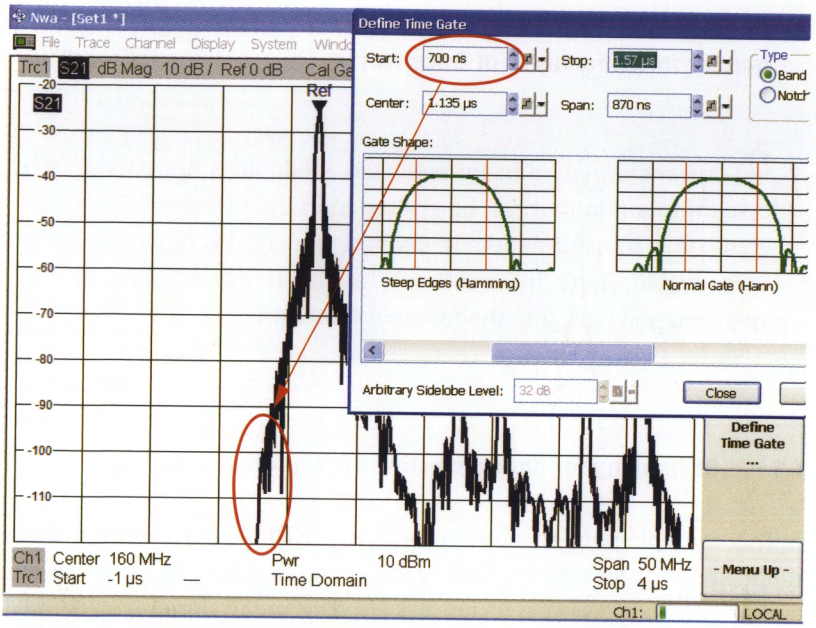


Fig. 6.2.4 Adjustment of the time gate.

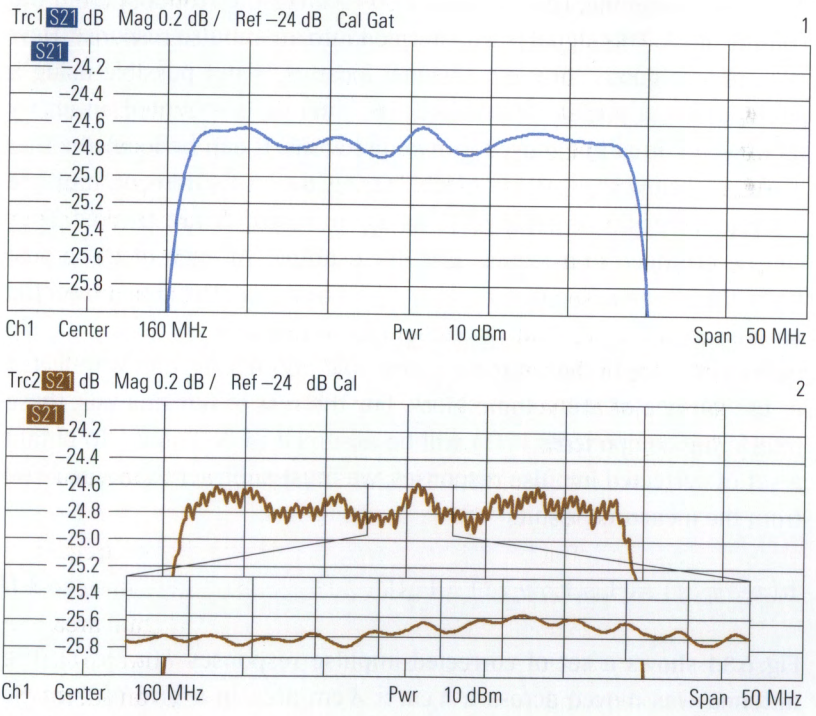


Fig. 6.2.5 Comparison of the frequency response with and without non-ideal properties.

10. Does it make any sense in Fig. 6.2.5 to activate the phase display format for further analysis of trace Trc1?



A transformation using bandpass mode does not provide any phase information that can be evaluated in the time domain. Nevertheless, this transformation is unambiguously reversible. In other words, usable phase information is available again following the transformation back into the frequency domain.

6.3 RF imaging for nondestructive evaluation

In this example nondestructive evaluation (NDE) is done using a small radar system. The target under examination is a, largely homogenous cuboid of about 25 cm × 25 cm × 8 cm. The radar signal is generated by the network analyzer as a frequency sweep and transmitted using a broadband antenna. This antenna also receives the response from the volume block. This signal is transformed into an impulse response. However, unambiguous three-dimensional location is not possible using a single antenna position. Therefore, the antenna is mounted on an x-y table where it can be moved in two directions. It can be located either above or below the volume block. During the measurement, multiple discrete x-y positions referred to as x_a, y_a hereafter are tracked. They can be arranged in a regular grid, for example. At each of these positions, the impulse response $\Gamma_h(x_a, y_a, t)$ is recorded. The signal from the antenna propagates three-dimensionally in space. A portion of the signal is already reflected in the antenna and another portion is reflected at the surface of the volume block. For the rest of our analysis, these interfering components $\Gamma_m(t)$ will be assumed to be known. To obtain a set of corrected impulse responses, we must subtract the value $\Gamma_m(t)$ from the measured results

$$h(x_a, y_a, t) = \Gamma_h(x_a, y_a, t) - \Gamma_m(t) \qquad \textit{(6.3-1)}$$

Fig. 6.3.1 shows a set of corrected impulse responses $h(x, y, t)$. The antenna was moved across a 4 cm × 4 cm area in a grid of 1 cm. So, $5 \times 5 = 25$ antenna positions were recorded. The volume block was found to have $\varepsilon_r = 9$ and conductivity $\rho = 0.4$ S/m. As a buried mate-

rial defect, a metal sphere with a radius of 6.34 mm was selected. The sphere was positioned at a vertical distance of 2.5 cm above the center of the block's bottom plane. The frequency sweep contained 100 points in the range from 50 MHz to 5.05 GHz (harmonic grid).

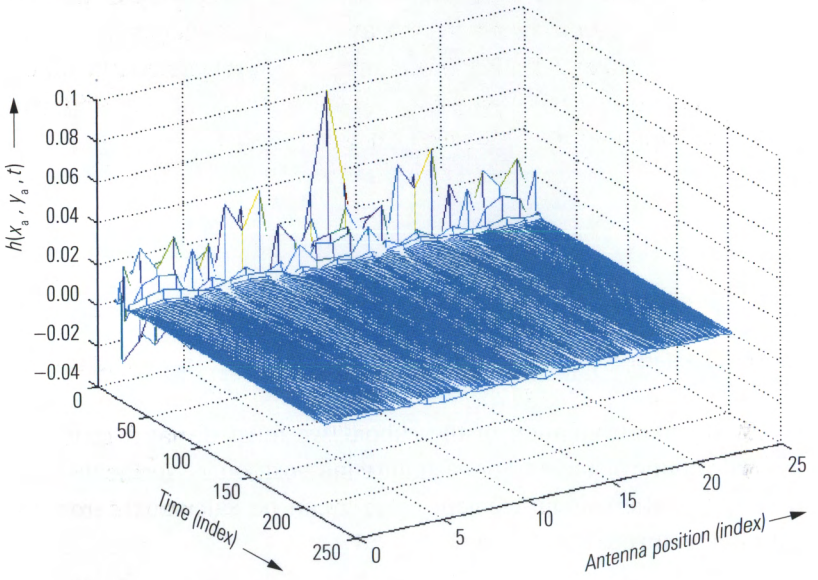


Fig. 6.3.1 Impulse response set $h(x_a, y_a, t)$.

The impulses remaining in the response are caused by scattering in the volume block. For example, they arise when an electrically conductive material or a material with a permittivity that is substantially different from its surroundings is enclosed in the volume. If we assume, within certain limits, constant wave propagation v_p, we can determine the distance r at which the reflection point is located based on the time offset t_p. Since the path is covered twice by the electromagnetic wave (once there and once back), we have the following relationship:

$$r = \frac{t_p}{2v_p} \qquad \textit{(6.3-2)}$$

The quantity r corresponds to the radius of a hemisphere around the antenna. Based on the signal from a single antenna, it is not possible to determine at which point in the surface of the hemisphere the reflection occurs (and thus the location of material defect). However, by over-

laying at least three hemispheres, we can determine the precise location in x, y, z.

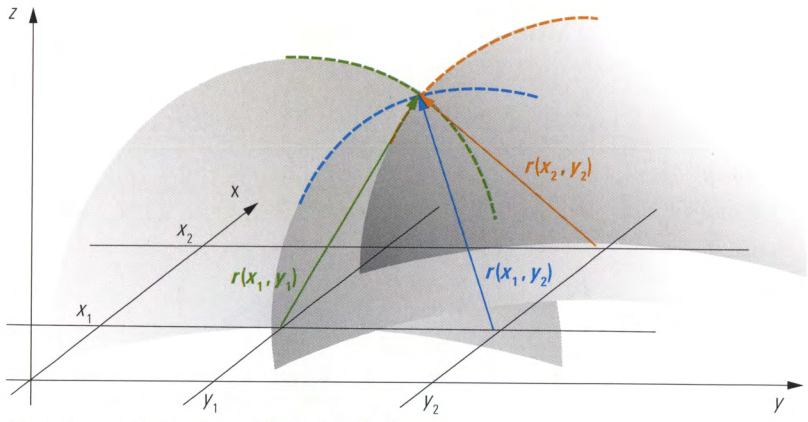


Fig. 6.3.2 Intersection of three hemispheres.

For systematic evaluation of the cuboid, we must divide it into cube-shaped volume elements. For each antenna position x_a, y_a, the distance $r(x_a, y_a)$ to the volume element x, y, z can be calculated using the Pythagoras theorem as follows:

$$r(x_a, y_a) = \sqrt{(x - x_a)^2 + (y - y_a)^2 + z^2} \qquad \textit{(6.3-3)}$$

Based on the distance $r(x_a, y_a)$, the time t_p is determined using formula (6.3-2). From this, the associated value $h(x_a, y_a, t_p)$ can be determined from the impulse response. As the average value of at least three of these values $h(x_a, y_a, t_p)$ that were determined for different antenna positions x_a, y_a, we obtain the measured value associated with a volume element. The evaluation can be programmed, for example, using the Matlab® script language developed by the company The MathWorks. Fig. 6.3.3 and Fig. 6.3.4 show false color renderings that were generated in this manner. The measured values used for these pictures are the same as those in Fig. 6.3.1.

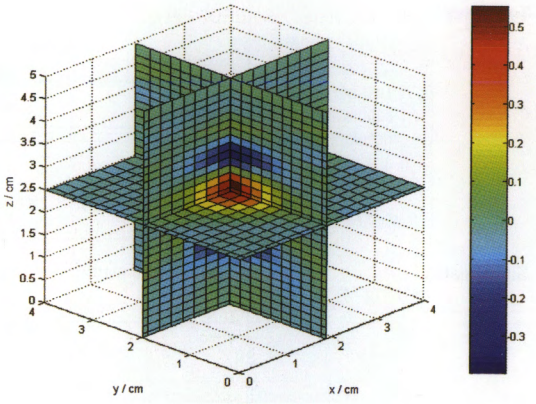


Fig. 6.3.3
Sectional display of the cuboid – View 1.

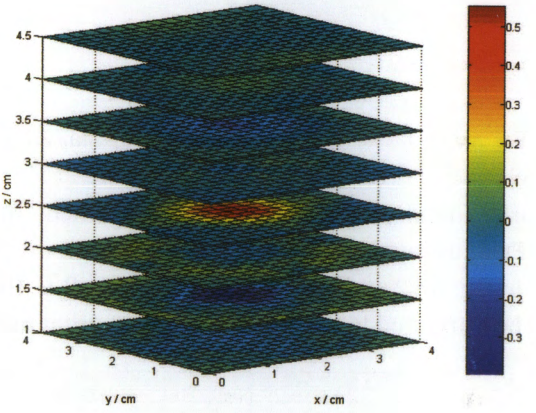


Fig. 6.3.4
Sliced display of the cuboid – View 2.

6.4 Measurement of the complex effective system data and OSML calibration

In section 3.5, we used the magnitudes of the effective system data for verification purposes only. A technique described in [RJ03] can be used to determine the complex effective system data for one-port measurements. These data are used as correction values for the previously calibrated network analyzer in order to further increase its accuracy. We will use this technique here as a measurement example. The test setup for determining the effective system data for the calibrated network analyzer is shown in the following figure.

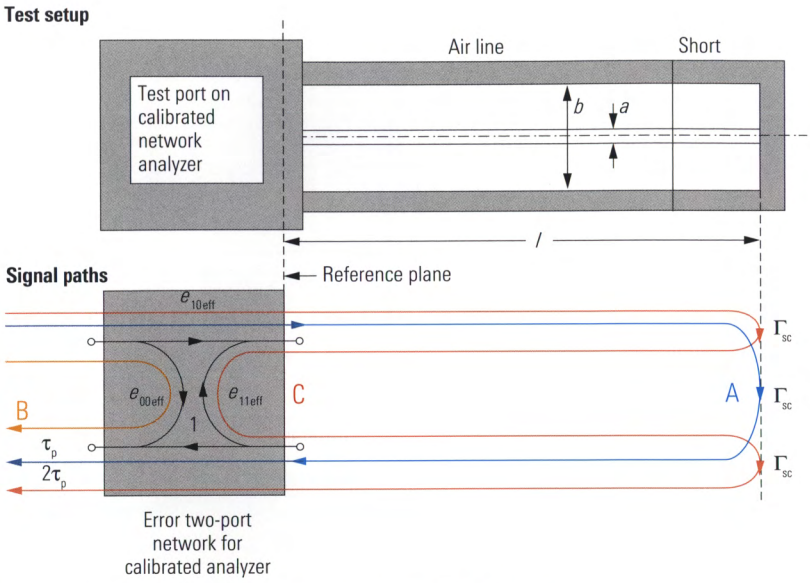


Fig. 6.4.1 Test setup for verification in the time domain.

The impulse response $\Gamma_h(t)$ is calculated from the measured values $\Gamma(f)$ using a time-domain transformation in lowpass mode. It is shown in Fig. 6.4.2 and has three main lobes A, B, C. These three main lobes can be attributed to the signal paths appearing in Fig. 6.4.1.

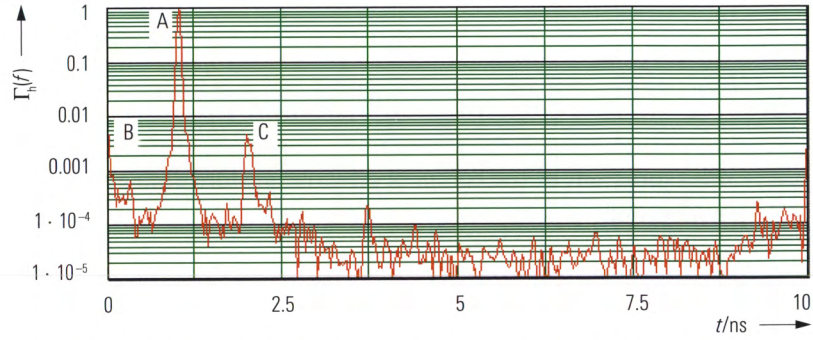


Fig. 6.4.2 Measurement result for verification in the time domain.

By using the time gate the pulses A, B, C can be separated from one another. Based on them, we can determine the complex functions $A(f)$, $B(f)$, $C(f)$ through a reverse transformation. Using the following formulas, we can calculate the complex effective system data as follows: Effective directivity $e_{00\text{eff}}$, effective reflection tracking $e_{10\text{eff}}$ and effective source match $e_{11\text{eff}}$.

$$e_{00\,\mathrm{eff}}(f) = B(f) \tag{6.4-1}$$

$$e_{10\,\mathrm{eff}}(f) = \frac{1 - \dfrac{B(f)}{\Gamma_{\mathrm{sc}}(f)}}{1 + \dfrac{C(f)}{A^2(f)}\left(\Gamma_{\mathrm{sc}}(f) - B(f)\right)} \tag{6.4-2}$$

$$e_{11\,\mathrm{eff}}(f) = \frac{C(f)}{A^2(f)} e_{10\,\mathrm{eff}} \tag{6.4-3}$$

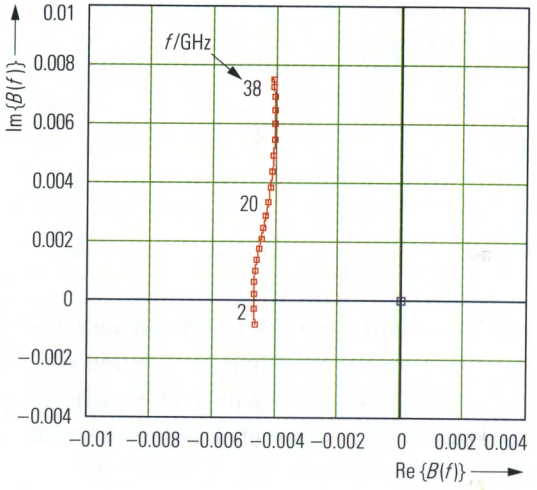


Fig. 6.4.3
Example of a locus curve for the effective directivity, measured using a very narrow time gate.

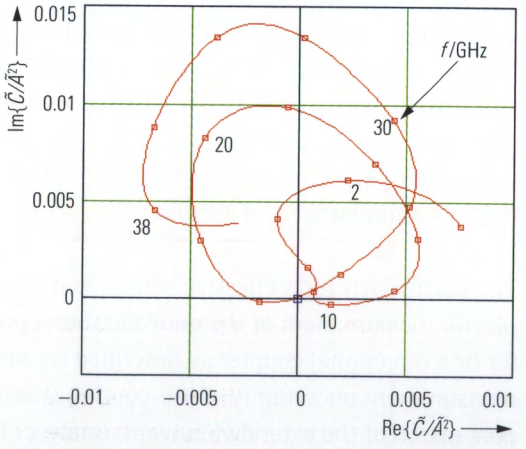


Fig. 6.4.4
Example of a locus curve for the effective source match, measured using a very narrow time gate.

The procedures necessary for this demonstration were programmed using Mathcad from the company Mathsoft™. The effective system data calculated using this technique depend on the accuracy of the characteristic impedance of the air line. It is possible to manufacture and trace

back the relevant dimensions (see Fig. 6.4.1, diameters a and b) with great precision. Since the effective system data $e_{00\text{eff}}$, $e_{10\text{eff}}$, $e_{11\text{eff}}$ were determined as complex quantities, it is possible to use them to further optimize the previously calibrated network analyzer. The following error model is used.

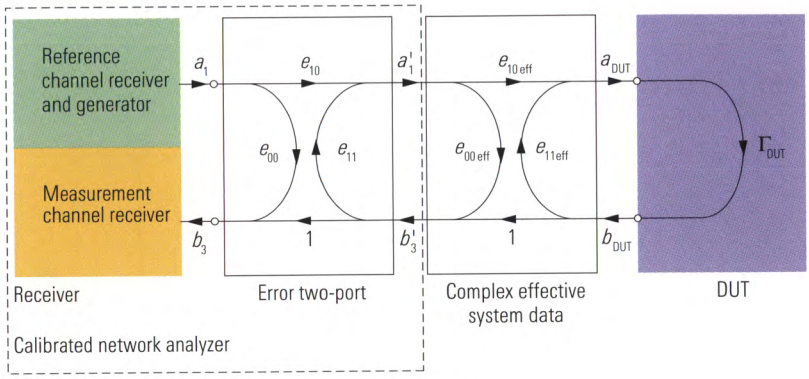


Fig. 6.4.5 Error model with effective system data.

For calculation purposes, the effective system data $e_{00\text{eff}}$, $e_{10\text{eff}}$, $e_{11\text{eff}}$ are combined with the existing error terms e_{00}, e_{10}, e_{11} yielding a new set of error therms that is used to realize the further optimized correction. This technique is named OSML (open, short, match, line) calibration. A network analyzer that is calibrated in this manner exhibits particularly high effective source match which is not attainable with the techniques described in section 3.4. Without using the OSML technique, the effective source matching would be limited by the accuracy of the open, short and match standards. It is not possible to manufacture and trace back these standards with the same accuracy as an air line. At best, we could use a sliding match in place of the match to improve the effective directivity to a similar level of accuracy.

The particularly high effective source match can be exploited, for example, for measurement of the equivalent test port match of a power splitter or a directional coupler as described by John R. Juroshek [Ju97]. The measurement uncertainty of this configuration is determined by the test port match of the extended network analyzer (see Fig. 6.4.6).

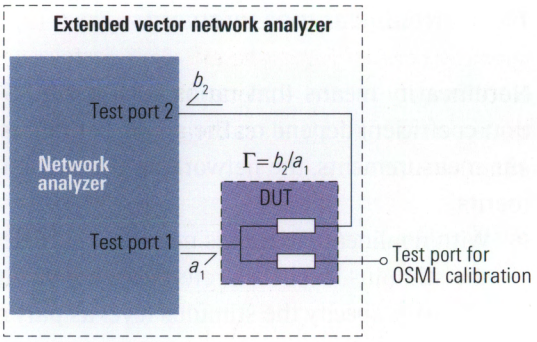


Fig. 6.4.6
Juroshek configuration.

7 Nonlinear measurements

Nonlinearity means that ratios such as the S-parameter or the reflection coefficient depend on the applied stimulus level. To perform nonlinear measurements, the network analyzer must fulfill two main requirements:

- With nonlinear DUTs, it is no longer sufficient to simply measure the stimulus level via a reference receiver. It must be possible to precisely specify the stimulus level as part of the measurement process.
- For a sinusoidal stimulus signal at a single frequency f_1, the response of the nonlinear DUT will have, in addition to the fundamental frequency f_1, additional harmonics at frequencies of $2f_1$, $3f_1$ and so on. The case of a stimulus with two frequencies f_1 and f_2 is similar. Here, however, additional intermodulation products such as $f_1 + f_2$, $f_1 - f_2$, $2f_1 - f_2$, $2f_2 - f_1$ and so on will occur, too. Measurement of their amplitude is part of nonlinear measurements and requires what is known as arbitrary mode[1]. The S-parameters can be supplemented with or replaced by additional measured quantities that are relevant particularly for nonlinear DUTs.

7.1 Features used for nonlinear measurements

This section describes some of the typical features that are provided by network analyzers for use in analyzing nonlinear networks. It also discusses innovative developments such as digital implementation of automatic level control and the true-differential-mode. Both require a suitably designed network analyzer.

7.1.1 Automatic level control

Almost all network analyzers use **automatic level control** (ALC) for a precision control of the stimulus level. The traditional design used for this task comes from the field of signal generators (see Fig. 7.1.1).

1) This mode is some times called "frequency offset mode" or "arbitrary frequency conversion mode". For the sake of simplicity, we use the term "arbitrary mode" in this book.

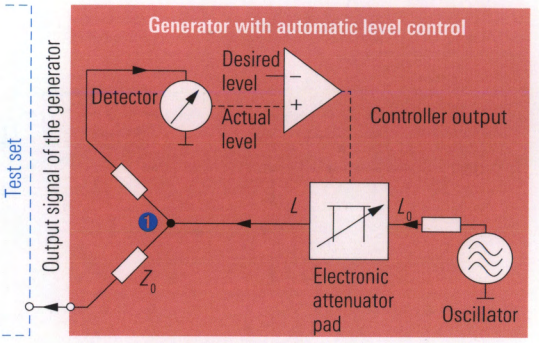


Fig. 7.1.1
Automatic level control for a generator.

In this concept a portion of the generator's output signal is used to feed a detector. The level of this signal is proportional to the level at point ❶. The **actual level** measured by the detector is transferred to a **controller** and compared with the **desired level**. Based on the comparison, an electronic attenuator pad reduces or increases the level at point ❶ until the desired level is attained. To optimize this **closed loop control** for stability and settling time the controller is not just a simple comparator. Normally a **PI controller** that involves a proportional plus an integral component is used. With its proportional component (P-component), it can respond quickly to disruptions. The integral component (I-component) evaluates the behavior of the controlled variable over a time interval and helps in this process to avoid **steady state error**. Due to the level control, the voltage at point ❶ is independent of the load that is connected via the adjacent test set. The source impedance at point ❶ is theoretically equal to 0 Ω. The impedance Z_0 of the power splitter transforms it to the source impedance Z_0. However, this technique has two significant disadvantages. At first, the power splitter has a loss of about 6 dB so that the level L_0 of the generator is no longer available in full. Secondly, the detector measures the power in a broadband fashion. The broadband noise that is inherently present is not separated from the useful signal. This prevents the detector from measuring very low source power levels. Under unfavorable conditions, a noisy signal fed to the controller as an actual level can even result in degradation of the signal-to-noise ratio at the generator's output.

Unlike a conventional signal generator, a vector network analyzer has narrowband receivers already, for example, to monitor the reference channel. The separation of the reference channel from the output signal is already part of the test set functionality.

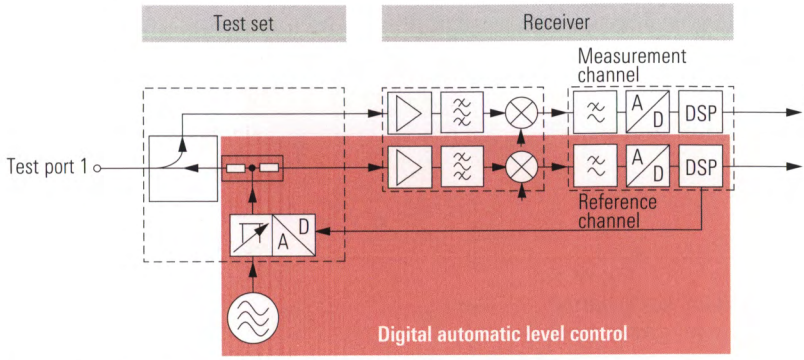


Fig. 7.1.2 Digital automatic level control in a network analyzer.

In other words, we do not need the power splitter and the detector from Fig. 7.1.1 if we use the reference receiver to provide the actual level. However, the numerical measured values provided by the reference receiver must be available continuously, not only during the actual recording of the measured values. Apart from the electronic attenuator, the level control must be implemented digitally. **Digital control** has several advantage: The **controller parameters** can be flexibly adapted to the current test situation. The adaptation process is handled automatically to keep the operation as simple as possible. As a result of the narrowband measurement of the actual level, the signal-to-noise ratio is improved compared to a traditional level control shown in Fig. 7.1.1. Digital control also makes it possible to use the wave quantity b_i from a different test port as the controlled variable (instead of the wave quantity a_k from the active port). This is particularly useful for measurements that must be performed using a constant DUT output level.

In section 3.4, we looked at various calibration techniques for linear measurements. Of course, calibration is just as important for nonlinear measurements. Modern network analyzers use **power calibration** for this purpose. Source power calibration adjusts the absolute level of the generator and includes correction of the reference receiver. Receiver power calibration adjusts the value displayed by the measurement receiver as the wave quantity b_i. All power calibrations are performed using scalar measurement of the errors and resulting in a scalar correction. These calibrations are available independently of the measured quantity and operating mode that are used. In other words, they are also possible for measurements involving frequency conversion and for mixer measurements. In contrast, the vectorial system error correction described

in section 3.4 can only be used for S-parameter measurements. It is not usable, e.g. for a_1 measurements, b_2 measurements or measurements involving frequency conversion. A combination of source power calibration and vector system error correction is possible. The following table summarizes when each of the different correction techniques is possible (or necessary). In particular, we should mention that vector system error correction should always be used for measuring S-parameters on linear or nonlinear DUTs.

Measurement	Vector system error correction (e. g. TOM, TOSM)	Power calibration for G: Generator R: Receiver
S-parameter measurements on linear DUTs	Yes, necessary	G: Not necessary R: Not necessary
S-parameter measurements or hot S-parameter measurement on nonlinear DUTs	Yes, necessary	G: Necessary R: Not necessary
Measurement of wave quantities, wave quantity ratios on linear or nonlinear DUTs	No, not possible	a-wave: G necessary b-wave: R necessary
Measurements involving frequency conversion on nonlinear DUTs	No, not possible	a-wave: G necessary b-wave: R necessary
Power sweep, e. g. for determining the 1 dB compression point based on the S-parameters	Yes, necessary	G: Necessary R: Not necessary
Determination of IP2, IP3 (two-tone measurements)	No, not possible	G: Necessary for both sources R: Necessary

Table 7.1.1 *Relationship between power calibration and vector system error correction.*

7.1.2 Source power calibration

A **source power calibration** corrects the generator level and the associated reference receiver (wave quantity a_i). Network analyzers are factory-equipped with level correction facilities. However, this cannot take into account the influence of the test setup, e.g. the test port cables employed by the user. Source power calibration is required to ensure an exact level at the input of the DUT. In practice, we achieve level accuracies of approx. 0.2 dB. A test setup can also involve a boosted source like shown in Fig. 7.1.3. Here, the reference plane for the calibration is at the output of the boosted source. The reference plane is the point at which

the generator level values are guaranteed after the calibration. A power sensor must be installed at this point to perform the calibration.

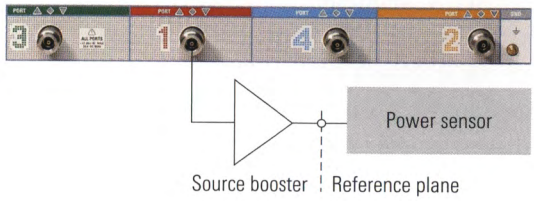


Fig. 7.1.3
Reference plane during source power calibration.

Modern power sensors[1] involve a power meter functionality and can therefore be directly connected to the network analyzer's USB interface. In older network analyzers, there is generally no USB interface available. In this case, a power meter[2] with an associated sensor is required. Communication takes place via the second GPIB interface (system bus) of the network analyzer. This is usually offered as an option. Based on the measured values provided by the power meter, the network analyzer computes the level deviation. The generator level is then corrected by exactly this value. If the computed correction value is relatively large, e.g. due to components with high loss or high gain located between the analyzer and reference plane, nonlinearities in the setup can result in the level not being exactly correct after the first correction. Accordingly, an iterative measurement makes sense. The previously corrected generator level is remeasured again so that the residual deviation can be determined and further correction is applied. This process is repeated until the accuracy specified by the user is attained (or the maximum number of iterations is reached). Prior to power calibration, the channel settings such as the sweep type, sweep range, number of points, source power and step attenuator setting should be chosen according to the subsequent measurement to be performed. Otherwise, interpolation of the calibration data is necessary, which can result in a loss of accuracy. Depending on the sweep type (frequency or power sweep), the correction data are recorded as a function of the frequency or as a function of the level. Power sensors can be classified as two types: Thermal sensors and diode sensors. The following table can be used to select the best power sensor for each case depending on the relevant requirements.

1) E.g. the R&S®NRP-Zx power sensors
2) E.g. the R&S®NRVD power meter

Thermal sensors	Diode sensors
The sensor does not have a lower frequency limit (measurement possible down to DC).	The sensor has a lower frequency limit (typ. 10 MHz).
Due to the thermal conversion, the measurement procedure is relatively slow.	The measurement procedure is very fast.
Due to the larger noise component, measurement of levels <–20 dBm is usually problematic. However, the basic accuracy of thermal sensors is higher than that of diode sensors.	With the exception of peak power sensors, diode sensors can normally measure levels down to approx. –50 dBm.

Table 7.1.2 Comparison of thermal sensors and diode sensors.

7.1.3 Receiver power calibration

A **receiver power calibration**, corrects the numerical values obtained from the measurement receiver in terms of their level. Receiver power calibration requires prior source power calibration. Therefore the relevant channel settings are usually already set on the instrument. But there is one peculiarity that needs to be taken into account if you are planning a measurement involving frequency conversion. The required source power calibration must be configured so that it also covers the relevant frequency range of the measurement receiver. Accordingly, the user interface shown in Fig. 8.3.2 has a suitable dialog box for mixer power calibration. This ensures that power calibration involving frequency conversion is performed easily and correctly (over the proper frequency range). A similar dialog box is provided for power calibration for the harmonic measurement shown in Fig. 7.4.6. For more details about these two special cases of power calibration, see sections 8.3 and 7.4.2, respectively.

To perform the receiver power calibration, the test port to be calibrated (e.g. test port 2) must be connected to another port for which a source power calibration has been performed previously (e.g. test port 1). The following explanation uses the last mentioned test ports 1 and 2 as an example. The reference receiver of port 1 is used to monitor the power delivered by the active test port 1. This power is compared with the one recorded by the measurement receiver of test port 2 to calculate its receiver correction values. Accordingly to these correction values the raw measured values of the b_2-wave are corrected in terms of their mag-

nitude. In the example used in the context of source power calibration, Fig. 7.1.3, the additional components that are used (boosted source and measuring cables) must remain in the test setup since the reference plane was defined at the output of the boosted source and the network analyzer references its a_1-wave to this point. As a consequence, the following test setup should be used in this case.

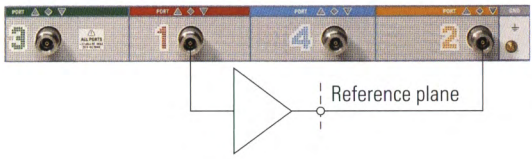


Fig. 7.1.4
Receiver power calibration for test port 2, suitable for the test setup in Fig. 7.1.3.

After the power calibrations have been done (see Fig. 7.1.3 and Fig. 7.1.4), the actual test setup with the DUT is connected (see following Fig. 7.1.5). If necessary, source power calibration should also be performed for test port 2. Due to the boosted source at test port 1, wave quantity b_1 of the DUT cannot be measured, and receiver power calibration at test port 1 is thus unnecessary. Test setups that allow measurement of the b_1-wave despite the boosted source are discussed in section 7.6.

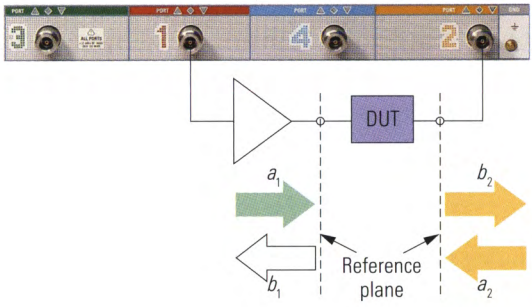


Fig. 7.1.5
Test setup after the power calibrations of Fig. 7.1.3 and Fig. 7.1.4.

7.1.4 Power sweep

The **power sweep** is particularly useful for investigating compression and saturation effects. Using a fixed operating frequency, the measured quantity is recorded as a function of the generator level and plotted with respect to a level axis in dBm. Some typical measurements that require a power sweep are as follows:

- Determination of the x dB compression point
- Recording of a detector characteristic
- Recording of the level-dependent IP3 of field effect transistors

They are discussed within the context of the measurement examples described in sections 7.2, 7.3 and 7.5. A power sweep range that is as large as possible (e.g. 40 dB) is an important quality criteria for a network analyzer. Also very important, however, is the measurement accuracy during the sweep process. When using a power sweep, we recommend that you always perform a power calibration. In advanced network analyzers, the level correction is incorporated into the power sweep so that the level can be individually corrected for each measurement point. Linearity fluctuations in the level control system are compensated in this manner. When measuring S-parameters, an additional system error correction is recommended.

7.1.5 Multiple source concept

A network analyzer typically has a central generator that is selectably switched to one of the test ports via a source switch (see Fig. 2.1.1). This means that the analyzer can only use one test port at a time as the active test port. With the **multiple source concept**, each test port has its own generator stage so that as many test ports as required can be used simultaneously as active test ports. Pairs of test ports share a common generator oscillator. In a three-port or four-port network analyzer, for example, two separate generator oscillators are integrated into the test instrument (see Fig. 7.1.6). This makes it possible for the instrument to simultaneously generate two stimulus signals internally with different frequencies. This is necessary, e.g. for investigating mixers or for two-tone analysis. An external generator is no longer necessary to provide the second RF signal. In addition, it is possible to control a generator integrated into the network analyzer much more efficiently. For the user, this means increased speed combined with many additional features such as the true-differential-mode.

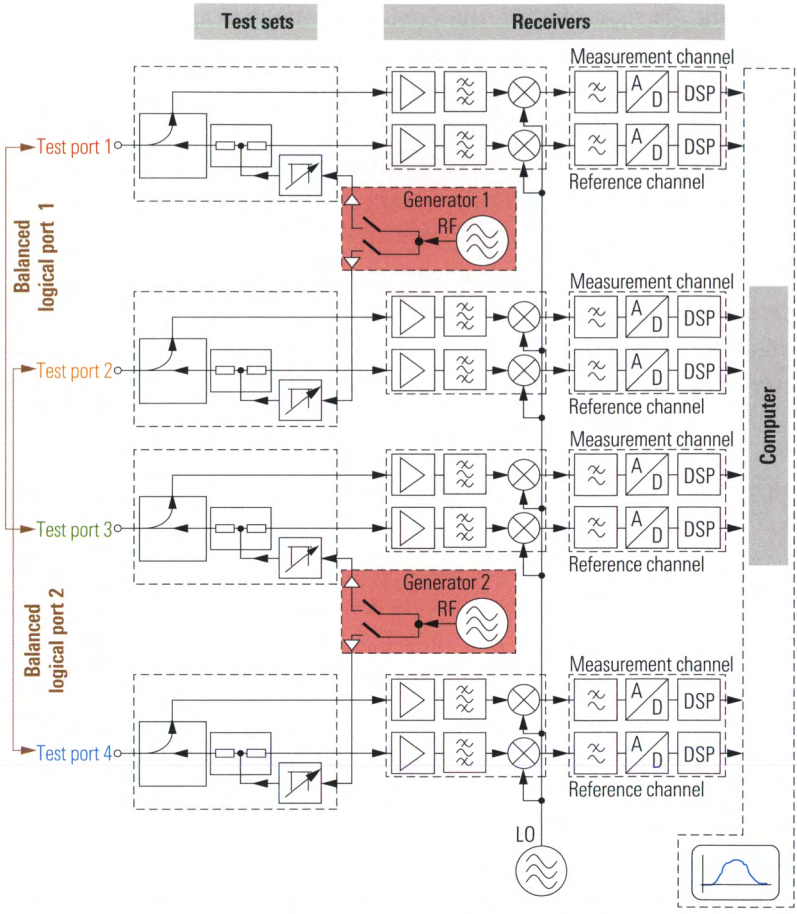


Fig. 7.1.6 The multiple source concept based on the example of the R&S®ZVA four-port analyzer.

7.1.6 Arbitrary mode

In principle, each test port can be assigned an individual stimulus frequency. The receiver frequency is the same for all of the test ports. This means that we can define five individual frequency vectors in a four-port network analyzer. The vector containing the receiver frequencies for all the sweep points can be written as $\mathbf{f}_{\mathrm{Rec}} = [f_{\mathrm{R0}}, f_{\mathrm{R1}}, \ldots f_{\mathrm{R(N-1)}}]$ for example. Normally, all of the frequency vectors are identical. However, **arbitrary mode** allows several linear dependencies between the frequency vectors. To define these dependencies we use a common base frequency vector $\mathbf{f}_{\mathrm{b}} = [f_{\mathrm{b0}}, f_{\mathrm{b1}}, \ldots f_{\mathrm{b(N-1)}}]$ as a reference. This vector normally corresponds

to the stimulus axis displayed with a frequency sweep or to the CW frequency used during a time sweep. Using the base frequency f_b, each of the other frequency vectors is computed using the following formula:

$$f_i = \frac{d_{numerator}}{d_{denominator}} f_{bi} + \frac{c_{numerator}}{c_{denominator}} f_{const} \quad \text{where } i = 0, \ldots (N-1) \qquad (7.1\text{-}1)$$

For each frequency vector, the constants $d_{numerator}$, $d_{denominator}$, $c_{numerator}$, $c_{denominator}$ and f_{const} must be individually specified. They must be chosen such that the frequencies f_i remain within the frequency range of the network analyzer and are not in contradiction with the respective network analyzer's block diagram.[1] The ratio of the constants $d_{denominator}$ and $d_{numerator}$ specifies the stepwidth difference between **f** and $\mathbf{f}_b$. Using the constants $c_{denominator}$, $c_{numerator}$ and f_{const}, a fixed frequency offset can be set.

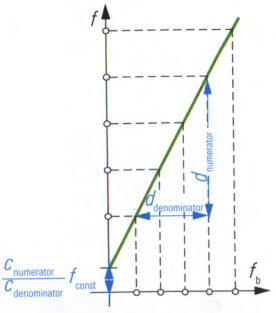


Fig. 7.1.7 Illustration of the terms from formula (7.1-1).

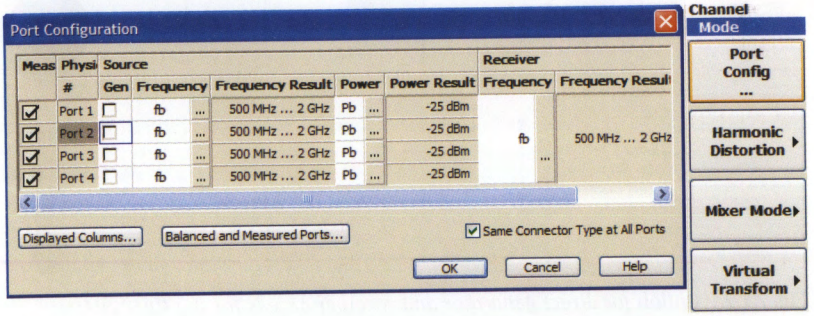


Fig. 7.1.8 Input dialog "Port Configuration" that can involve the arbitrary mode settings

1) Example: Operating test ports 1 and 2 as active test ports at the same time (generators) with different source frequencies would be in contradiction with the block diagram in Fig. 7.1.6. Due to the common generator oscillator, the two stimulus frequencies must be equal. If we use instead test ports 1 and 3 the frequencies are independent even when both ports are used as active ports at the same time.

7.1.7 Direct generator and receiver access

In sections 2.4.1 and 2.4.2, we already saw a way to use receiver and generator step attenuators to handle the extreme level conditions that occur at an amplifier module. In the case of high power amplifiers, however, internal step attenuators are sometimes insufficient so that we have to build an external test set. The prerequisite here is **direct generator and receiver access**. This is available as an option in some instruments. The following figure shows test port 1 together with the associated interfaces for direct generator and receiver access.

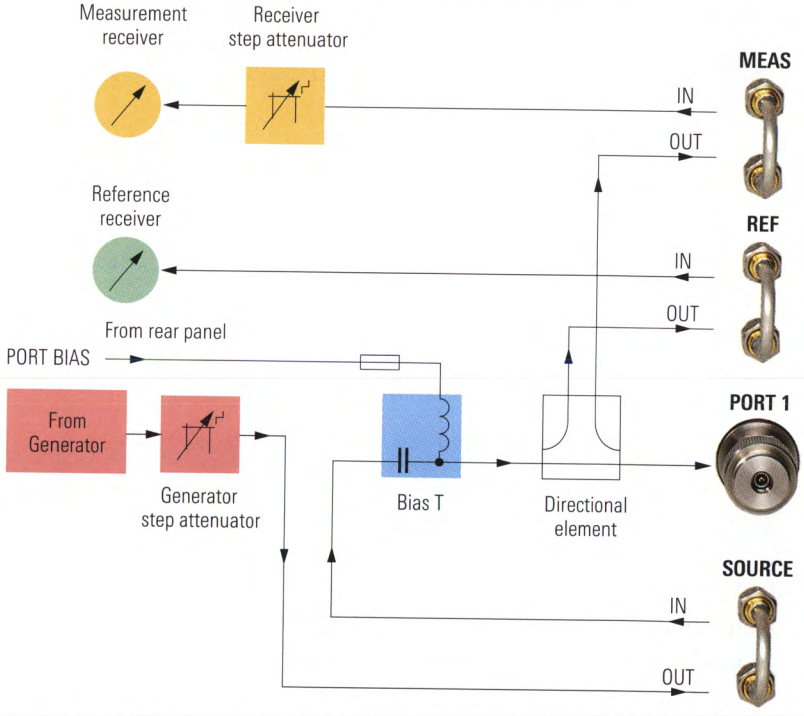


Fig. 7.1.9 Option for direct generator and receiver access for the R&S®ZVA24 network analyzer.

7.1.8 Power sensors as receivers

In some measurements, we do not need any phase information and broadband detection is permissible or even necessary, e.g. if we would like to evaluate the overall spectral output power of an amplifier. Here, the obvious thing to do is to use power sensors and power meters as

additional receivers in the measurement process. The level they measure is recorded synchronously with the sweep and displayed as a trace. Using active USB hubs, the number of USB ports at the network analyzer can be extended to allow additional power sensors with USB interfaces. By cascading multiple USB hubs, we can involve a large number of power sensors. Of course, it is important to keep in mind the significantly greater measurement time for a power meter compared to the internal receivers used in the analyzer.



Fig. 7.1.10 R&S®ZVB4 network analyzer with four R&S®NRP-Z51 power sensors.

7.1.9 External generator control

Many nonlinear measurements involve more than one stimulus generator. The majority of the commercially available network analyzers have only a single internal generator, which makes it necessary to attach an external generator. Even for analyzers providing multiple sources an additional generator can be of interest, e.g. when using the concept shown in Fig. 7.1.6 to do intermodulation measurement on a mixer which requires three generators. The external signal generator is controlled e.g. via GPIB. For this purpose the network analyzer has to be equipped with a second GPIB interface (often named as GPIB system bus) or a USB/GPIB adapter has to be used. If the network analyzer uses the VISA I/O library for control purposes, then additional interfaces such as a LAN will be available for communication. Additional control lines (handshake signals: EXT GEN TRG, EXT GEN BLANK) and the list mode, considerably speed up the tracking of the generator during the sweep process.

In the following figure, the network analyzer forms a GPIB subsystem along with the external generators. Accordingly, the control sequences that are sent to the generators during the sweep process do not block the analyzer's first GPIB interface. During mixer measurements, the

external generator can sometimes be operated at a fixed frequency so that no communication is required between the generator and the network analyzer. In this case the frequency and level of the generator must be set manually.

Besides remote control of the generators, their frequency synchronization is also important. The common 10 MHz reference frequency helps to ensure this. In Fig. 7.1.11, the lower generator acts as the reference master and supplies the reference frequency for all of the other instruments. In some cases, these instruments must be set for an external reference to enable synchronization to the reference master instead of their internal reference.

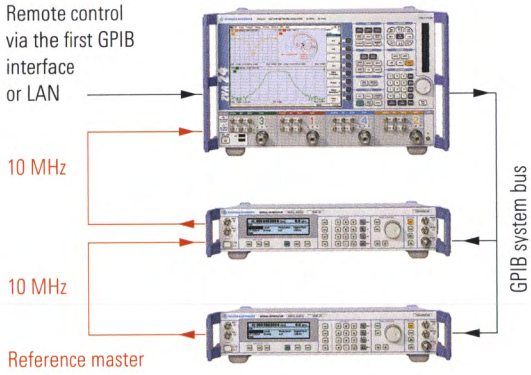


Fig. 7.1.11
R&S®ZVA24 network analyzer with two R&S®SMR40 generators.

7.1.10 Additional equipment

The mechanical receiver and generator step attenuators are built of resistor networks and thus do not exhibit nonlinear behavior. They can be used to distinguish the intermodulation products of the DUT and the network analyzer during certain measurements. For more details, see section 7.5.2.

In two-tone measurements, we must combine signals from two generators. For this purpose, an internal **combiner** is available in some network analyzers. The following two circuits are suitable for this purpose.

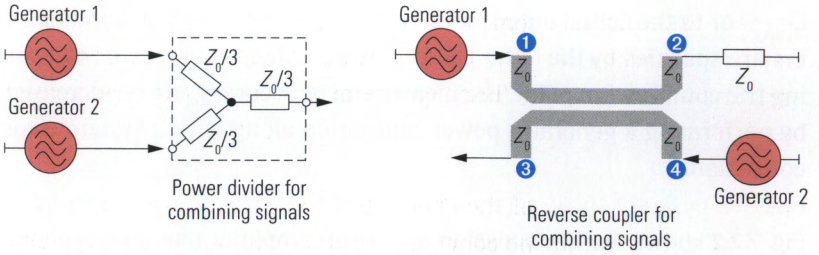


Fig. 7.1.12 Combination of two generator signals.

The broadband bandwidth of a **power divider** is one factor in favor of its usage. Note, however, that it has an insertion loss of 6 dB. On the other hand, a **reverse coupler** can be configured to have an insertion loss of about 3 dB. In addition, such a device offers significantly better decoupling of generators 1 and 2 when properly designed. This is useful for avoiding mutual intermodulation that occurs at the output stage of the generators.

7.2 Measurement of the compression point

When operating a component (e.g. an amplifier) within its linear region the output level L_{OUT} increases at the same rate as the input level L_{IN} for a given frequency. When increasing the input level L_{IN} beyond the linear region the output level L_{OUT} does not keep pace with the linear increase (see Fig. 7.2.1).

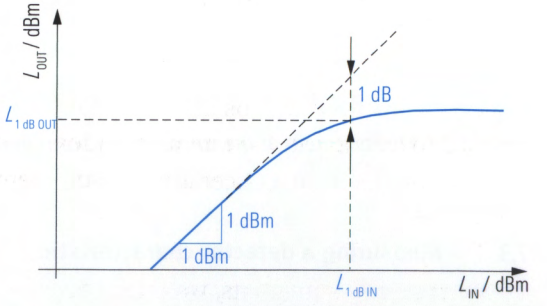


Fig. 7.2.1 Definition of the 1 dB compression point.

The point at which the output level L_{OUT} is 1 dB less than the expected value assuming a linear increase is known as the **1 dB compression point** (sometimes also called **1 dB gain compression point**). It is usually measured with a single-frequency stimulus signal. The level associated with the 1 dB compression point can be referred to the input level as

$L_{1\,\mathrm{dB\,IN}}$ or to the actual output level as $L_{1\,\mathrm{dB\,OUT}}$. Most high power amplifiers are specified by the latter value. A power sweep is useful for measuring the compression point. The measurement accuracy can be increased by performing a generator power calibration along with a system error correction.

Fig. 7.2.2 shows the transmission a_{s21} for an amplifier based on a power sweep. The 1 dB compression point corresponds to the drop in the transmission by 1 dB. A statistical function is used to search for a drop by precisely this amount. It is also possible to specify values other than 1 dB. The levels $L_{1\,\mathrm{dB\,OUT}}$ and $L_{1\,\mathrm{dB\,IN}}$ are displayed as the pair of values Cmp Out and Cmp In.

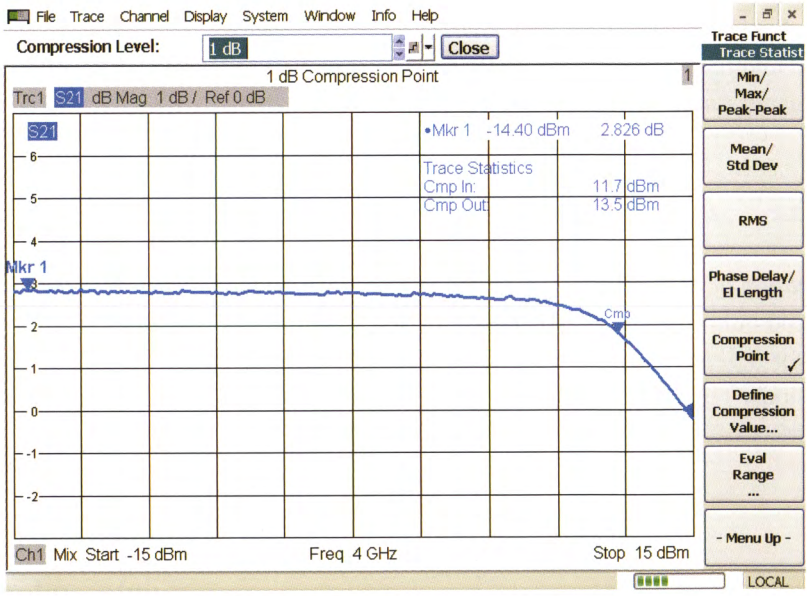


Fig. 7.2.2 Measurement of the compression point of an amplifier.

7.3 Measuring a detector characteristic

Detector circuits offer one approach to level measurements. Due to their simple design, they are used, for example, as overdrive indicators in receivers. Fig. 7.3.1 shows an example of a detector circuit. A diode rectifies the applied RF voltage and the resulting DC voltage is evaluated as a measure of the RF level. The relationship between the RF power P_{IN} and the output voltage U_{DC} of the detector is nonlinear.

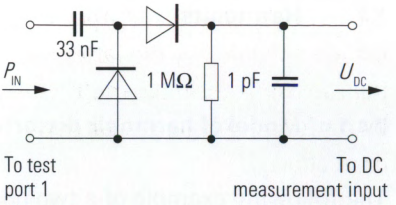


Fig. 7.3.1
Simple detector circuit with a diode.

The network analyzer's **DC measurement inputs** measure the output voltage U_{DC} of the circuit shown in Fig. 7.3.1. The input power P_{IN} applied to the detector circuit is varied using a power sweep. Our measurement result is the detector characteristic shown in Fig. 7.3.2. In the lower part of this figure, the detector output voltage is plotted as a function of the input level in dBm at an operating frequency of 1 GHz. In addition, the level-dependent reflection coefficient is measured, too. It is approximately constant.

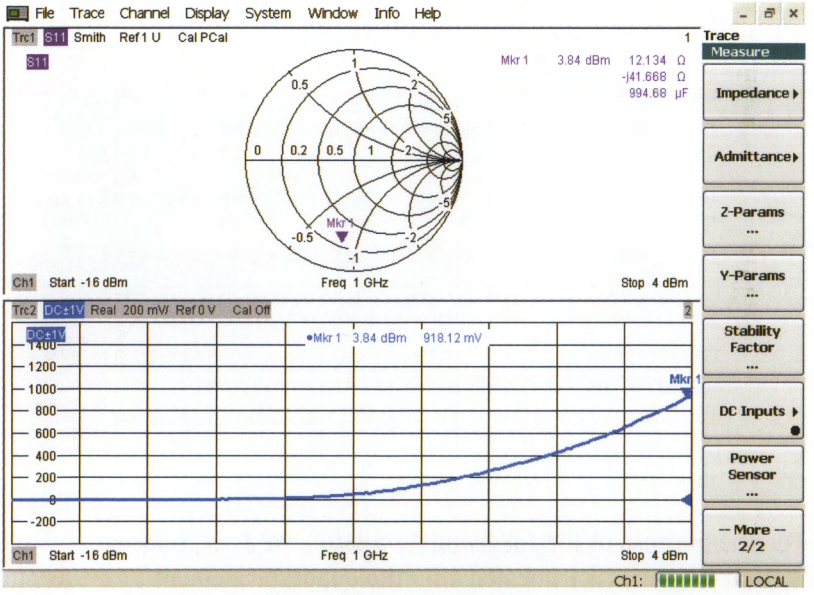


Fig. 7.3.2 Measurement of the detector characteristic.

In some applications, the frequency dependence of the detector's DC voltage is relevant. In such cases, a frequency sweep is used instead of a power sweep.

7.4 Harmonics

7.4.1 Model of harmonic distortions

The following example of a two-port device describes how harmonic distortions arise. We assume a nonlinear relationship $b_2(a_1)$ between the stimulus a_1 at port 1 and the response b_2 at port 2. In principle, the nonlinear function $b_2(a_1)$ can be approximated by the **Taylor series**.

$$b_2(a_1) \approx s_1 \cdot a_1 + s_2 \cdot a_1^2 + s_3 \cdot a_1^3 + \ldots + s_{k_{max}} \cdot a_1^{k_{max}} \qquad \textit{(7.4-1)}$$

The computation of the coefficients s_1, s_2, s_3 and so on will not be discussed in further detail here. The relevant details can be found if needed in mathematics text books[1]. As seen in the following table, the quantity k_{max} determines the quality of approximation.

k_{max}	Description
1	The Taylor series consists only of the term $b_2 = s_1 a_1$. This corresponds to the linear behavior described in the previous chapters.
2	The remaining deviations are mostly small. The curvature of the nonlinear function is exactly only at $a_1 = 0$.
3	The remaining deviations are relatively small. Besides the curvature, the change in the curvature of the nonlinear function is also measured. The Taylor series thus gets very close to the function $b_2(a_1)$ not only in the vicinity of $a_1 = 0$.
4	Compared to $k_{max} = 3$, a small improvement is achieved. However, in most cases it does not justify the significant higher computing expense.
∞	The Taylor series characterizes the function $b_2(a_1)$ exactly without any deviation

Table 7.4.1 *Properties of the Taylor series as a function of* k_{max}.

For the purpose of the following discussion, we have selected $k_{max} = 2$ so that the series is truncated after the quadratic element $s_2 \cdot a_1^2$. As our stimulus $a_1(t)$, we use the oscillation $a_1(t) = \hat{A}_1 \cos(2\pi f_1 t + \varphi_1)$.

$$b_2(a_1) \approx s_1 \cdot \hat{A}_1 \cos\left(2\pi f_1 t + \varphi_1\right) + s_2 \cdot \left(\hat{A}_1 \cos\left(2\pi f_1 t + \varphi_1\right)\right)^2 \qquad \textit{(7.4-3)}$$

Using the trigonometric identity $\cos^2\alpha = \frac{1}{2}(1 + \cos 2\alpha)$, we can simplify the formula given above. The constant term $\frac{1}{2} \cdot s_2 \hat{A}_1^2$ is irrelevant in our discussion and will be neglected.

1) See, for example, [Di87] page 414, [BS91] page 269 or a similar text book.

$$b_2(a_1) \approx s_1 \cdot \hat{A}_1 \cos(2\pi f_1 t + \varphi_1) + \frac{s_2 \hat{A}_1^2}{2} \cos(2 \cdot 2\pi f_1 t + 2\varphi_1) \qquad \textbf{(7.4-4)}$$

For a nonlinear DUT, a stimulus with a **single-tone signal** (frequency f_1) thus leads to a response having multiple **harmonics**. Besides the **first harmonic** at the **fundamental frequency** f_1, it also contains the **second harmonic**[1] at the frequency $2f_1$. If we choose $k_{max} > 2$ for our approach (7.4-3), we can calculate additional harmonics at frequencies of $3f_1$, $4f_1$, ... $k_{max} f_1$. This also corresponds to our practical experience with nonlinear DUTs. While the amplitude of the first harmonic increases linearly for an increase in the input amplitude $\hat{A}_1$, the amplitude of the second harmonic increases quadratically. Instead of the amplitude $\hat{A}_1$, in most cases we will measure the following level[2]:

$$L_{IN} = 20 \lg \left| \hat{A}_1 \right| \text{ dBm} \qquad \textbf{(7.4-5)}$$

Likewise, we obtain the output level for the fundamental frequency f_1 as follows:

$$L_{OUT} = 20 \lg \left| s_1 \hat{A}_1 \right| \text{ dBm} = a_{s1} + L_{IN} \qquad \textbf{(7.4-6)}$$

Here, $a_{s1} = 20\lg|s_1|$ dB represents the transmission of the fundamental frequency. For the second harmonic, i.e. the frequency $2f_1$, we can calculate the output level as follows:

$$\begin{aligned} L_{k2} &= 20 \lg \left| \frac{s_2 \hat{A}_1^2}{2} \right| \text{dBm} = (20 \lg |s_2| + 20 \lg \left| \frac{1}{2} \right| + 20 \lg \left| \hat{A}_1^2 \right|) \text{dBm} \qquad \textbf{(7.4-7)} \\ &= a_{s2} - 6\,\text{dB} + 2 \cdot 20 \lg \left| \hat{A}_1 \right| \text{dBm} = a_{s2} - 6\,\text{dB} + 2L_{IN} \end{aligned}$$

The quantity L_{k2} is also known as the **harmonic level**. The terms $a_{s2} = 20\lg|s_2|$ dB and -6 dB as well as $a_{s1} = 20\lg|s_1|$ are each independent of the input level L_{IN}. An increase in the level at the input by ΔL_{IN} results in an identical change of the level at the output at the first harmonic (fundamental frequency), while the level of the second harmonic increases by $2\Delta L_{IN}$. Likewise, the level L_{k3} of the third harmonic

1) Instead of the term k-th harmonic, it is also common for $k > 1$ to use the term $(k-1)$-th **overtone** especially in the field of acoustics (e.g. second harmonic = first overtone). Note, k is always an integer number.

2) For the sake of simplicity, the Amplitude $\hat{A}_1$ used here is scaled to meet the root mean square value (RMS value).

increases by $3\Delta L_{IN}$. In the following figure, the gray arrows represent the increases of the discussed harmonic output levels.

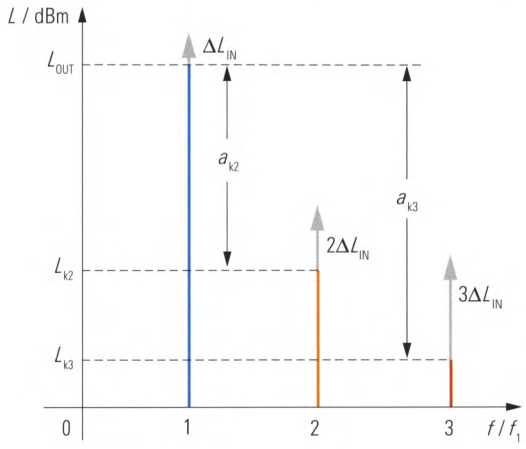


Fig. 7.4.1
Harmonic levels in the output spectrum for a single-tone stimulus.

The **harmonic suppressions**[1] a_{k2} and a_{k3} characterize, the level difference between L_{OUT} and a **higher-order harmonic level** L_{k2}, L_{k3}, ... e. g. $a_{k2} = L_{OUT} - L_{k2}$.

In Fig. 7.4.2, the levels L_{OUT} and L_{k2} of the first and second harmonics are plotted vs. the input level L_{IN}. These curves are typical for nonlinear DUTs.[2] Based on Fig. 7.4.2, for a specified input level L_{IN1} the associated harmonic suppression $a_{k2}(L_{IN1})$ can be read off as the difference between the two functions $L_{OUT}(L_{IN})$ and $L_{k2}(L_{IN})$.

1) In some literature you will find the harmonic suppression to be called the harmonic distortion value. But we are using the phrase suppression because it is more intuitive when talking about a level difference that is measured with respect to the level at fundamental frequency. For the same reason, this book and e.g. [MG92] page F28 use the unit dBc (decibel below carrier) with the harmonic suppression and the later mentioned intermodulation suppression. But do not feel confused to find other literatures like [ZB93] volume 2 page 306 or [CE05] in which the harmonic suppression just takes the unit dB. Alternatively to a dB scale the harmonic suppression can be converted in to a linear scale. This so-called harmonic distortion factor can be understood as the ratio of the n-th harmonic's power to the fundamental power. Summing up all the harmonic distortion factors in the range of n = 2 to ∞ leads to the so-called total harmonic distortion factor (THD). It is usually scaled in % and compares the total harmonic power of all higher harmonics with the power related to the fundamental frequency. The THD is typically used in the field of audio engineering and power supply industry several other definitions are possible. The THD is mentioned here only for the sake of completeness.

2) **Field effect transistors** (FETs) are an exception, for example. With them, the series approach (7.4.1) works only under certain conditions. For more details on this limitation, see, e.g. the study [CE05].

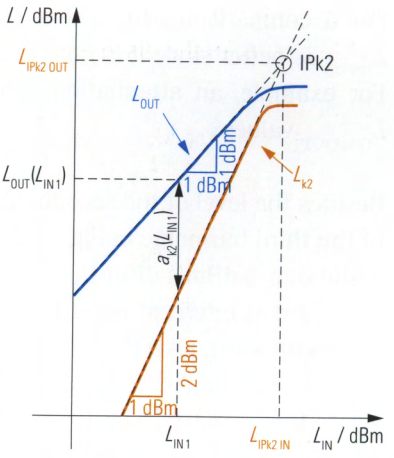


Fig. 7.4.2
Second harmonic intercept point.

All of the quantities we have discussed so far have in common that they are dependent on the input level L_{IN}. Accordingly, it is hard to find some global characterization for them. This is why the **second harmonic intercept point** IPk2 was introduced. It is shown in Fig. 7.4.2 as the intersection IPk2. The level L_{IPk2IN} corresponds to the fictitious input level at which the output level L_{OUT} of the fundamental frequency and the output level L_{k2} of the second harmonic have the same value. The point IPk2 is not attained in most cases due to saturation effects, but it is still useful for a global characterization. Using it, we can immediately construct the two straight lines for L_{OUT} and L_{k2} without any additional information.

The higher the levels L_{IPk2IN} and $L_{IPk2OUT}$ associated with the intercept point are, the less pronounced the harmonic distortions. To determine the intercept point IPk2, all we have to do is measure a pair of values L_{OUT} and L_{k2} at the same input level L_{IN}. The latter must be selected like the level L_{IN1} in Fig. 7.4.2 so that we will not experience any saturation effects. With $a_{k2} = L_{OUT} - L_{k2}$ we calculate:

$$L_{IPk2IN} = L_{IN} + a_{k2} \qquad \textit{(7.4-8)}$$

From Fig. 7.4.2, we can also read off the level of the intercept point that is referred to the output. It is denoted by $L_{IPk2OUT}$ and can be calculated from the transmission of the fundamental a_{s1} in dB and L_{IPk2IN}.

$$L_{IPk2\,OUT} = a_{s1} + L_{IPk2\,IN} \qquad \textit{(7.4-9)}$$

For a comparison of different components, the input intercept level L_{IPk2IN} is better since it is not dependent on the gain of the component. For example, an attenuation value $a_{s1} < 0$ would result in improved $L_{IPk2OUT}$ values.

Besides the level of the second harmonic, we can also draw in the level of the third harmonic in Fig. 7.4.2. It would result in a straight line with a slope of 3 dBm/1 dBm that would intersect with the line of the fundamental at intercept point IPk3. This is known as the **third harmonic intercept point**. Like the intercept point IPk2, it represents a fictitious intersection. Similar to the formulas given above, we can calculate the levels L_{IPk3IN} and $L_{IPk3OUT}$ for it.

$$L_{IPk3IN} = L_{IN} + \tfrac{1}{2} a_{k3} \qquad \textit{(7.4-10)}$$

$$L_{IPk3OUT} = a_{s1} + L_{IPk3IN} \qquad \textit{(7.4-11)}$$

7.4.2 Measurement of the harmonics and their intercept point

Optimizing the harmonic distortion figure using a lowpass filter

To measure harmonics, we need a generator signal with very good harmonic suppression. This harmonic suppression should be at least approx. 10 dB greater than the expected harmonic suppression of the DUT. The harmonic suppressions a_{k2} and a_{k3} for the test port output signal generally have a guaranteed minimum value ranging from 20 dBc to 30 dBc. The actual observed values will often be much greater. In many cases, the harmonic suppression of the generator is sufficient. If this is not the case, however, the suppression can be improved using a lowpass filter between the active test port and the DUT. Fig. 7.4.3 shows a measurement of the harmonic level L_{k3} in which a lowpass filter has been arranged at the test port. The left part of the figure shows the harmonic frequency in the passband of the filter, while the right part shows that the harmonic level has been reduced by the stopband attenuation of the filter to a value of < -94 dB. To obtain the harmonic suppression from this, we must also take into account the insertion loss of the filter in the passband and the source power (here, 4 dB and 0 dBm). In the current

case, we calculate a harmonic suppression $a_{k3} > 90$ dBc at the filters output for $f > 430$ MHz.

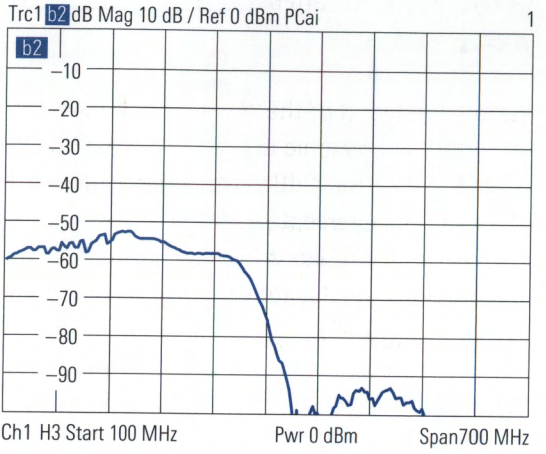


Fig. 7.4.3
Measurement of the harmonic level L_{k3} at the output of the lowpass at test port 1.

Between the filter and the DUT, an attenuator pad should be introduced in order to ensure good source match over a broad bandwidth (see Fig. 7.4.4). In the stopband range, the filter exhibits s_{11} and s_{22} values close to total reflection. Using a 10 dB attenuator pad, we can ensure source match of at least 20 dB. When measuring harmonics, good source match is an absolute must since vector system error correction cannot be used.

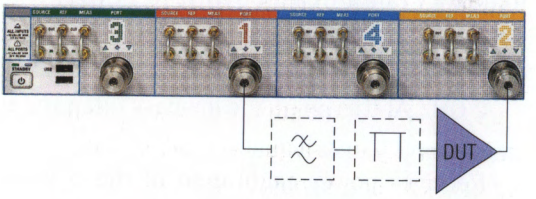


Fig. 7.4.4
Harmonic measurement using a lowpass filter.

To make use of the filter effect, we must limit the available sweep range to less than one octave. If f_c is the corner frequency of the filter and f_s is the stop frequency, the generator frequency must be equal to at least $\frac{1}{2}f_s$ for the filter to suppress the second harmonic and all of the higher harmonics with its stopband attenuation (see Fig. 7.4.5). The stimulus frequency must be below the corner frequency f_c since the generator signal would be attenuated otherwise contrary to our objective.

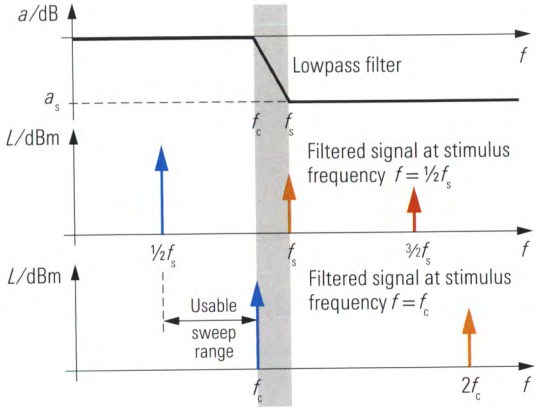


Fig. 7.4.5 Illustration of the usable sweep range with a lowpass filter.

Power calibration for measuring harmonics

As part of preparations for measuring harmonics, we recommend power calibration for all of the measured quantities. Receiver power calibration is necessary for the stimulus frequency range as well as for the frequency range of the harmonics to be measured. The calibration work is simplified by the use of a special dialog box (see Fig. 7.4.6). If your network analyzer does not offer this feature or you are using a lowpass filter, you will need to individually plan and execute the different calibration steps. These steps are described for the example of a harmonic suppression measurement a_{k2} for the second harmonic and a fundamental frequency range from 300 MHz to 1 GHz:

1. Source power calibration for the frequency range from 600 MHz to 2 GHz. At this point, the lowpass filter and the attenuator pad should not yet be installed at the test port.
2. Receiver power calibration of the b_2-wave based on the previous source power calibration. For this, connect test port 1 and test port 2 using a through connection. The measurement is made in the frequency range from 600 MHz to 2 GHz.
3. Source power calibration for the fundamental frequency from 300 MHz to 1 GHz. Install the lowpass filter and attenuator pad. The level is now measured at the output of the attenuator pad. This means the power sensor replaces the DUT of Fig. 7.4.4.
4. Receiver power calibration of the b_2-wave based on the previous source power calibration. In the present case, connect the output of the attenuator pad and test port 2 using a through connection.

The source power calibration described under step 1 is only intended to prepare the receiver power calibration for the frequency range of the second harmonic (step 2). Step 4 is required only because we need the exact fundamental frequency output level L_{OUT} in addition to the harmonic level L_{k2} to calculate the harmonic suppression a_{k2}.

Example 1: Measurement of intercept points IPk2 and IPk3 using a frequency sweep

The following measurement example uses a low-noise amplifier with 18 dB gain. The initial measurement quantities to be determined are the b_2 wave $b_2(f_b)$ as well as the second and third harmonics $b_2(2f_b)$, $b_2(3f_b)$. Using the softkey bar, we can directly select the harmonics (see Fig. 7.4.6, right). Since all three measurements have different channel settings, three independent channels are used to handle them (see section 2.8.2).

In the first step of the measurement, we must perform a power calibration for all of the measured quantities. To do this, we use the dialog box shown in Fig. 7.4.6 for the corresponding traces of the second and third harmonics, i.e. the measured quantities $b_2(2f_b)$ and $b_2(3f_b)$. The measured quantities $a_1(f_b)$ and $b_2(f_b)$ are corrected using the standard power calibration.

In the second step, the harmonic suppressions for the generator signal are checked using a through connection. In the current case, we read off (except for the range above 3 GHz) a suppression of typically >40 dB (see Fig. 7.4.6). This is sufficient for the expected results and a lowpass filter is not required.

In the third step of this measurement, connect the DUT in place of the through connection. You should obtain traces like the ones shown in the upper part of Fig. 7.4.8. They reveal harmonic suppressions that are largely independent of frequency. They have values of $a_{k2} = 8.8\ \text{dBm} - (-23.7\ \text{dBm}) = 32.5\ \text{dBc}$ and $a_{k3} = 8.8\ \text{dBm} - (-27.0\ \text{dBm}) = 35.8\ \text{dBc}$.

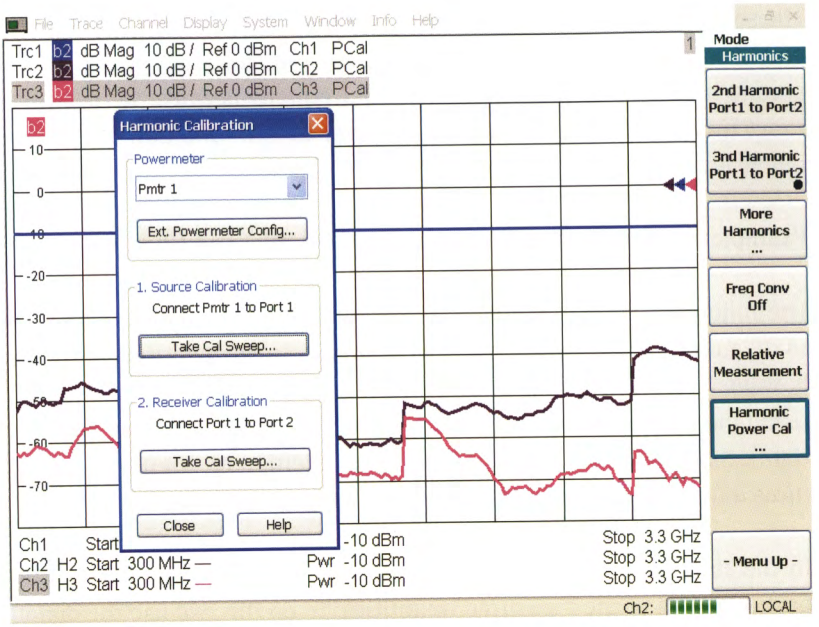


Fig. 7.4.6 Checking the harmonic distortion figures on the generator (test port 1).

Using trace mathematics, we will now calculate the input levels of the intercept points IPk2 and IPk3. Since trace mathematics deals with the linear complex measured values, formulas (7.4-8) and (7.4.10) must be expressed in terms of linear quantities. Instead of the levels L_{IPk2}, L_{IN}, L_{OUT} and L_{k2}, we can use the power levels P_{IPk2}, P_{IPk3}, $P_{\mathrm{IN}} = |a_1|^2$, $P_{\mathrm{OUT}} = |b_2(f_b)|^2$ as well as $P_{\mathrm{k2}} = |b_2(2f_b)|^2$ and $P_{\mathrm{k3}} = |b_2(3f_b)|^2$. This leads to the following power relationships.

$$P_{\mathrm{IPk2}} = P_{\mathrm{IN}} \frac{P_{\mathrm{OUT}}}{P_{\mathrm{k2}}} \qquad \textit{(7.4-13)}$$

$$P_{\mathrm{IPk3}} = P_{\mathrm{IN}} \left(\frac{P_{\mathrm{OUT}}}{P_{\mathrm{k2}}} \right)^{1/2} \qquad \textit{(7.4-14)}$$

However, since the result of the computation is to be interpreted as a wave quantity, only the magnitudes of the wave quantities are used instead of their squares. Exact entry of the trace mathematics for IPk3 is shown in the next figure. The corresponding IPk2 and IPk3 traces are shown in the lower half of Fig. 7.4.8. The violet trace represents the level L_{IPk2IN} and the pink trace represents the level L_{IPk3IN}.

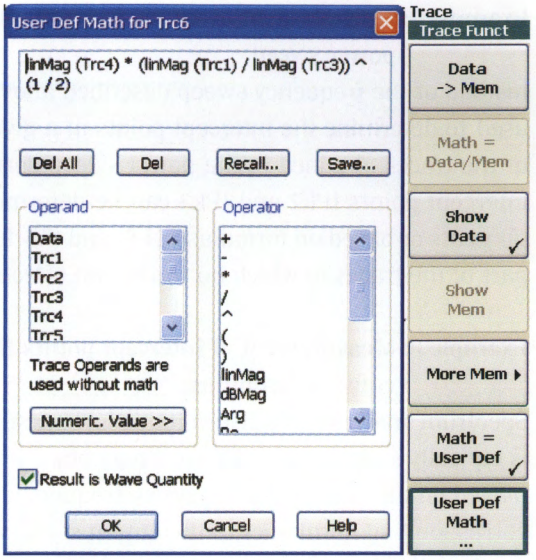


Fig. 7.4.7 Entry of trace mathematics for computation of the intercept point IPk3.

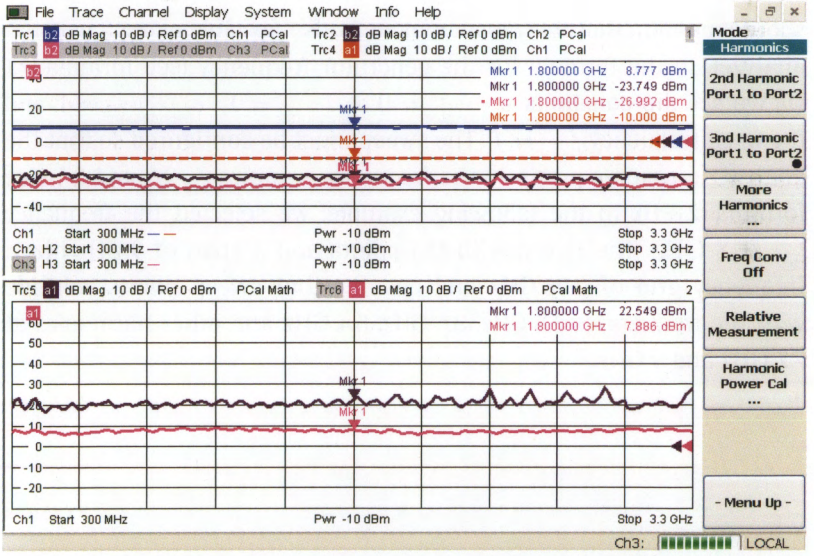


Fig. 7.4.8 Measurement of harmonic levels and the intercept points IPk2, IPk3 of an amplifier.

In view of the rather complicated trace mathematics a manual verification of the result is advisable. Based on the harmonic suppressions $a_{k2} = 32.5$ dBc, $a_{k3} = 35.8$ dBc and the input level -10 dBm, we obtain the values $L_{IPk2IN} = 22.5$ dB and $L_{IPk3IN} = 7.9$ dB using formulas (7.4-8) and (7.4-10).

Example 2: Measurement of intercept points IPk2 and IPk3 using a power sweep

Instead of the frequency sweep described above, a power sweep can be used to determine the intercept points at a given frequency. The slopes of the traces obtained in the power sweep correspond to Fig. 7.4.2. The intercept points IPk2 and IPk3 can be determined by constructing the tangents or based on formulas (7.4-8) and (7.4-10). In both cases, only the part of the traces in which no saturation effects occur is relevant.

Example 3: Measurement of intercept points IPk2 and IPk3 at constant stimulus

Spectrum analyzers are also suitable for measuring harmonics. The DUT is typically supplied with a fixed frequency f_1 and its output signal is displayed on the spectrum analyzer vs. frequency. We can then observe the corresponding harmonics in the output signal at frequencies f_1, $2f_1$ and $3f_1$. The disadvantage of this technique is that the harmonics are measured only at one stimulus frequency. For the sake of completeness, we should mention that this measurement is also possible using a network analyzer. In the formula for the generator frequency (see formula (7.1-1)), we set the constants $d_{\text{numerator}} = 0$, $c_{\text{numerator}} = 1$, $c_{\text{denominator}} = 1$ and the fixed frequency $f_{\text{const}} = f_1$. The sweep must be configured so that the analyzer measures precisely the frequency samples f_1, $2f_1$ and $3f_1$ (see formula (2.8-1)). In the following example, we selected the frequency $f_1 = 1.8$ GHz. The sweep with 181 points and a span of 9 GHz has a frequency grid of 0.05 GHz and ensures that sweep points coincide with the frequencies 1.8 GHz, 3.6 GHz, 5.4 GHz and other harmonics at 7.2 GHz and 9 GHz.

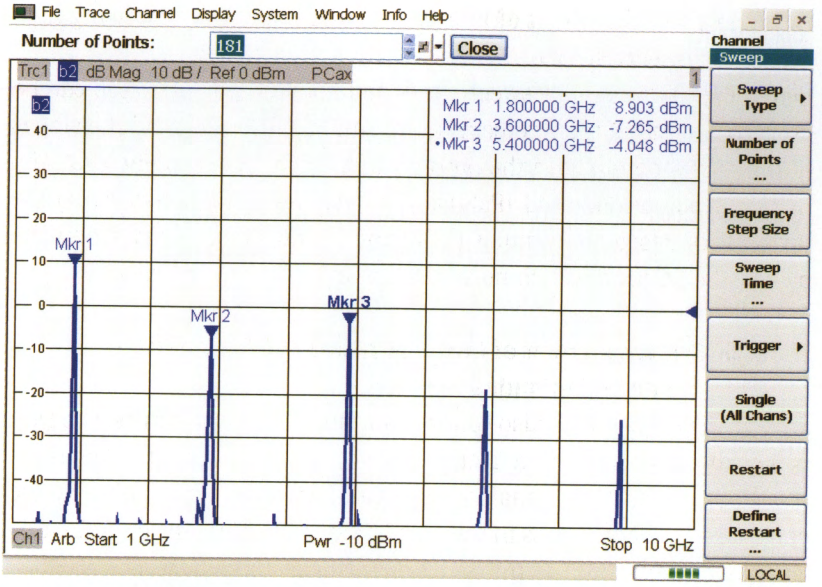


Fig. 7.4.9 Measurement of the harmonics of an amplifier with stimulus at a fixed frequency.

7.5 Intermodulation

7.5.1 Model of intermodulation distortions

Radio channels used to transmit signals occupy a frequency bandwidth B. This means that the circuit components that are involved no longer have just a single frequency applied to them. A simple test signal that at least partially mimics the occupied bandwidth B consists of two sinusoidal oscillations with the same amplitude $\hat{A}_1 = \hat{A}_2$ and a frequency spacing B. This is known as a **two-tone signal**. We will typically encounter frequency spacings $f_1 - f_2 = 10$ kHz to 10 MHz.

$$a(t) = \hat{A}_1 \cos(2\pi f_1 t + \varphi_1) + \hat{A}_2 \cos(2\pi f_2 t + \varphi_2) \tag{7.5-1}$$

Like in harmonic analysis, the signal (7.5-1) is inserted into the Taylor series (7.4-1) with $k_{\max} = 3$. This results in what is known as **intermodulation products**. Once we have simplified and sorted out the different terms, we can summarize the results in the following table.

Component	Formula	Note
DC component	$0.5 \cdot s_2(\hat{A}_1^2 + \hat{A}_2^2)$	1)
Fundamentals (First harmonic)	$s_1\hat{A}_1 \cdot \cos(2\pi f_1 t + \varphi_1)$ $s_1\hat{A}_2 \cdot \cos(2\pi f_2 t + \varphi_2)$	2)
Cross modulation	$(0.75 \cdot s_3\hat{A}_1^3 + 1.5 \cdot s_3\hat{A}_1\,\hat{A}_2^2) \cdot \cos(2\pi f_1 t + \varphi_1)$ $(0.75 \cdot s_3\hat{A}_2^3 + 1.5 \cdot s_3\hat{A}_1^2\,\hat{A}_2) \cdot \cos(2\pi f_2 t + \varphi_2)$	3)
2nd harmonic	$(0.5 \cdot s_2\hat{A}_1^2) \cdot \cos(2 \cdot 2\pi f_1 t + 2\varphi_1)$ $(0.5 \cdot s_2\hat{A}_2^2) \cdot \cos(2 \cdot 2\pi f_2 t + 2\varphi_2)$	4)
Intermodulation products, 2nd order	$(s_2\hat{A}_1\hat{A}_2) \cdot \cos(2\pi(f_1 - f_2)\, t + \varphi_1 - \varphi_2)$ $(s_2\hat{A}_1\hat{A}_2) \cdot \cos(2\pi(f_1 + f_2)\, t + \varphi_1 + \varphi_2)$	5)
3rd harmonic	$0.25 \cdot s_3\hat{A}_1^3 \cdot \cos(3 \cdot 2\pi f_1 t + 3\varphi_1)$ $0.25 \cdot s_3\hat{A}_2^3 \cdot \cos(3 \cdot 2\pi f_2 t + 3\varphi_2)$	4)
Intermodulation products, 3rd order	$0.75 \cdot s_3\hat{A}_1^2\hat{A}_2 \cdot \cos(2\pi(2f_1 + f_2)t + 2\varphi_1 + \varphi_2)$ $0.75 \cdot s_3\hat{A}_1^2\hat{A}_2 \cdot \cos(2\pi(2f_1 - f_2)t + 2\varphi_1 - \varphi_2)$ $0.75 \cdot s_3\hat{A}_1\hat{A}_2^2 \cdot \cos(2\pi(2f_2 + f_1)t + 2\varphi_2 + \varphi_1)$ $0.75 \cdot s_3\hat{A}_1\hat{A}_2^2 \cdot \cos(2\pi(2f_2 - f_1)t + 2\varphi_2 - \varphi_1)$	6)

Table 7.5.1 *Intermodulation products.*

Notes on the table:

1) This component has a frequency of 0 Hz and is usually irrelevant due to the DC decoupling.
2) The first-order harmonics usually represent the useful signal. They correspond to the amplified or attenuated sinusoidal oscillations that were fed in.
3) The component known as the **cross modulation** has the same frequency as the useful signal and thus cannot be distinguished from it in a measurement.
4) This signal component represents the higher-order harmonics of the useful signal which have a relatively large frequency spacing with respect to it of f_1, f_2 or $2f_1$, $2f_2$, respectively. They are easy to suppress with filters.
5) The **2nd order intermodulation products** (also known as **differential tones**) are used, for example, in a mixer as the upper sideband $f_1 + f_2$ or the lower sideband $f_1 - f_2$. They are considered to be undesirable in the following and are characterized within the context of the second order intercept (SOI).
6) The third-order intermodulation products are handled in the context of the third order intercept (TOI). The signal components at $2f_1 - f_2$ and $2f_2 - f_1$ are particularly disruptive since they are relatively close to the useful signal.

The following figure illustrates the terms listed in Table 7.5.1. The color coding corresponds to the table (blue: fundamentals, orange: second-order products, red: third-order products). The **order** of the intermodulation products corresponds to the sum of the orders of the components that are involved. For the product at the frequency $2f_1 - f_2$, we calculate the order as $2 + 1 = 3$, for example. All the products of table 7.5.1 despite

the fundamentals are collected by the expression intermodulation distortion (IMD).

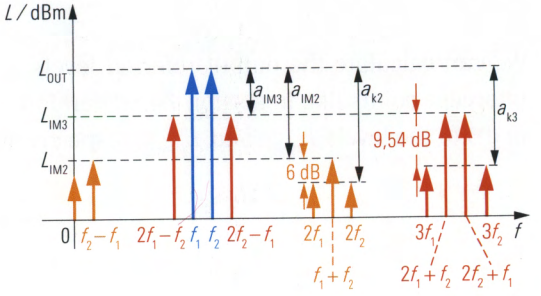


Fig. 7.5.1
Output spectrum of a nonlinear circuit with two-tone input signal at $f_1 < f_2$.

The definition of the **intermodulation suppression** a_{IM2}, a_{IM3} is visible in the figure above.

$$a_{IM2} = L_{OUT} - L_{IM2} \qquad \textit{(7.5-2)}$$

$$a_{IM3} = L_{OUT} - L_{IM3} \qquad \textit{(7.5-3)}$$

As with the harmonic analysis, an increase in the level of the two sinusoidal oscillations at the input by ΔL_{IN} also results in a level change in the respective intermodulation product by $n\Delta L_{IN}$, where n is the order of the product and the level changes are expressed in dB.

Consequently, we can calculate a **n-th order intercept point**. This point is assigned a dBm input level L_{IPnIN}.

$$L_{IPn\,IN} = \frac{a_{IMn}}{n-1} + L_{IN} \qquad \textit{(7.5-4)}$$

At the fictitious input level L_{IPnIN}, the n-th intermodulation product has the same level as the first harmonic at f_1 or f_2. The other quantities found in formula (7.5-4) are the intermodulation suppression a_{IMn} for the n-th intermodulation product in dBc and the associated input level L_{IN} in dBm. This level refers to one of the two input signals with the same amplitude. In most cases, the **second order intercept** (SOI) and the **third order intercept** (TOI) are specified. For their input level L_{IP2IN} or L_{IP3IN}, we have the following:

$$L_{IP2\,IN} = a_{IM\,2} + L_{IN} \qquad (7.5\text{-}5)$$

$$L_{IP3\,IN} = \frac{a_{IM\,3}}{2} + L_{IN} \qquad (7.5\text{-}6)$$

We can calculate the output referred levels L_{IP2OUT} and L_{IP3OUT} of the intercept points like in section 7.4.1. Therefore, we just add the gain a_{s21} in dB to the levels L_{IP2IN} and L_{IP3IN}, respectively.

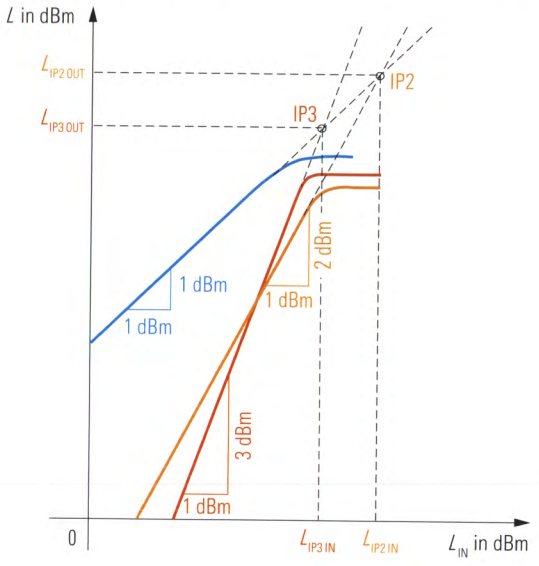


Fig. 7.5.2
Construction of intercept points IP2 and IP3.

7.5.2 Measurement of the intermodulation products and their intercept point

To measure intermodulation, we need a two-tone signal with frequencies of f_1 and f_2. A three- or four-port network analyzer based on the multiple source concept, is capable of simultaneously generating the frequencies f_1 and f_2 with its two internal generators. These stimulus signals are available, for example, at test ports 1 and 3. Network analyzers that only have one internal generator will require an external signal generator for this application. Control of the external signal generator was previously discussed in section 7.1.9. Arbitrary mode can be used to configure a constant frequency offset between all involved generators (internal and external generators).

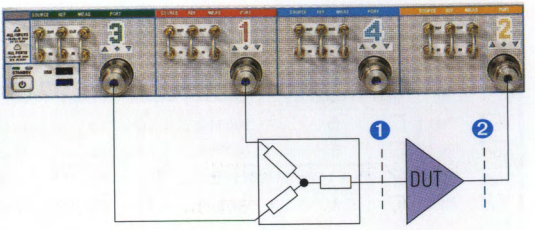


Fig. 7.5.3
Intermodulation measurement using the R&S®ZVA24 and a power divider for combination of the signals from test ports 1 and 3.

Possible implementations of the combiner are discussed in section 7.1.10. The sinusoidal oscillations at frequencies f_1 and f_2 that are used as stimulus signals must have the same level at the DUT input (reference plane ❶).

Example 1: Intermodulation measurement using a frequency sweep

As our first measurement example, we will consider an intermodulation measurement using a frequency sweep. The test setup shown in Fig. 7.5.3 is assumed. The stimulus power is set to -13 dBm. The center frequency of 1 GHz and the sweep range of 200 MHz determine the limits of the base frequency f_b. In arbitrary mode, we will use default settings except for test port 3 (see Fig. 7.5.4 next page). Test port 3 is operated as a permanently active test port ("Gen" column is checked) and it is assigned a generator frequency of $f_b + 1$ MHz. First, the source power L_{IN} at test port 1 is displayed as Trc1. It is not necessary to display the source power of test port 3 since it is equal to the level on test port 1. The output level of the DUT at the frequency f_b is measured at test port 2 as Trc2. For the intermodulation measurement, a new trace Trc3 is generated from trace Trc2 and assigned to a separate channel Ch2. This is a prerequisite for using Trc3 as shown below. Because the new channel Ch2 was derived from Ch1 it has been assigned the same arbitrary mode configuration which is shown in Fig. 7.5.4. But in difference to this setting Trc3 should be used to measure the power of the third-order intermodulation product. To do so it is necessary to use a receiver frequency of $2f_1 - f_2 = 2(f_b + 1\ \text{MHz}) - f_b = f_b + 2\ \text{MHz}$. The dialog "Port Configuration" associated to channel Ch2 is modified at column "Receiver Frequency" to $f_b + 2$ MHz to meet these needs. As the measured quantity the wave b_2 is selected because it corresponds to the DUT's output signal.

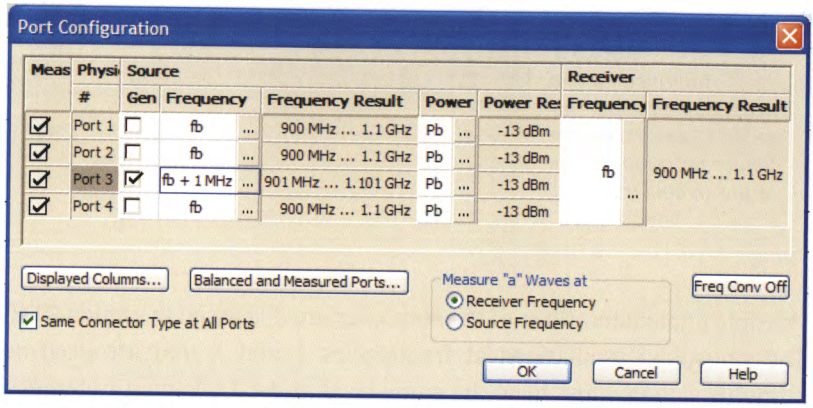


Fig. 7.5.4 Configuration of arbitrary mode.

In the intermodulation measurement, we are primarily interested in the third order intercept point IP3. It can be determined from the intermodulation suppression a_{IM3} and the input level L_{IN}. In Fig. 7.5.5, trace mathematics is used to compute a trace for the intercept point IP3. The corresponding trace is shown in the lower part of Fig. 7.5.7 and Fig. 7.5.8. Trace mathematics is based on linear complex measured values. Therefore, we must express formula (7.5-6) in terms of the linear power. We use the input power $P_{in} = |a_1|^2$, the output power of the first harmonic $P_{OUT} = |b_2(f_b)|^2$ and the output power of the intermodulation product. Here, this is $P_{IM3} = |b_2(f_b + 2\ \text{MHz})|^2$.

$$P_{IP3\,IN} = \left(\frac{P_{OUT}}{P_{IM3}}\right)^{1/2} \cdot P_{IN} \qquad (7.5\text{-}7)$$

Since the result of the calculation will be a wave quantity, only the magnitudes of the complex wave quantities are used in the mathematical expression Fig. 7.5.5 instead of the squares of the magnitudes.

Power calibration is necessary in order to perform the intermodulation measurement accurately. The reference planes for this calibration are located immediately before and after the DUT (see ❶ and ❷ in Fig. 7.5.3).

Since the power divider is already present in the test setup during the power calibration, its insertion loss of 6 dB is compensated by the calibration process.

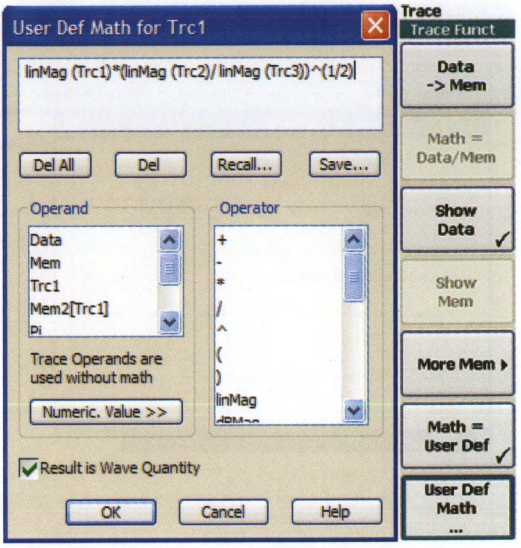


Fig. 7.5.5
Trace mathematics for computation of intercept point IP3.

1. First, we must perform a source power calibration of the wave quantities a_1 and a_3 for channel Ch1. During this calibration procedure, only the test port to be calibrated must be activated.[1] In addition, receiver power calibration for the wave quantity $b_2(f_b)$ as described in section 7.1.3 is required.
2. For channel Ch2, i.e. the intermodulation product, we first perform a source power calibration covering the frequency range of the intermodulation product (here, 902 MHz to 1.002 GHz). This source power calibration is used for a receiver power calibration for the intermodulation product $b_2(f_b + 2\ \text{MHz})$.
3. For the wave quantities a_1 and a_3 of channel Ch2, the source power calibration from channel Ch1 has to be adopted.

It is also possible to skip step 2 and use the receiver power calibration from channel Ch1 for CH2. If the frequency points of the calibrated channel Ch2 do not exactly coincide with the sweep points of Trc2, the correction data are normally calculated by interpolation / extrapolation. Of course, this will result in a loss of accuracy.

1) In the user interface shown in Fig. 7.5.6, click the "Modify Settings" button in the "Source Power Cal" dialog. This should open the "Modify Source Power Cal Settings" dialog box. Check "Switch Off All Other Sources During Cal Sweep" so that only the test port to be calibrated, e.g. "Source: Port 1" will be active during source power calibration.

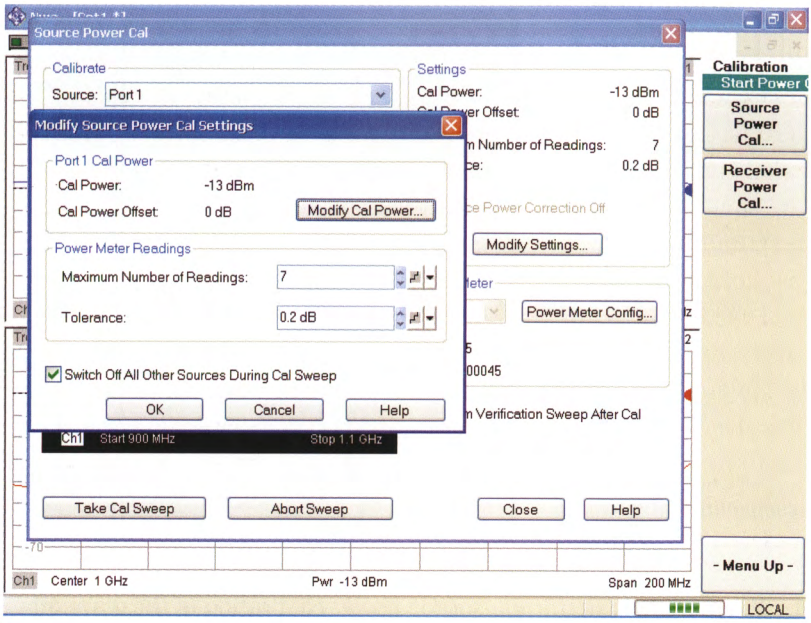


Fig. 7.5.6 Source power calibration on channel Ch1.

Once the calibration process is completed, the DUT is connected. The corresponding measurement result is shown in Fig. 7.5.7.

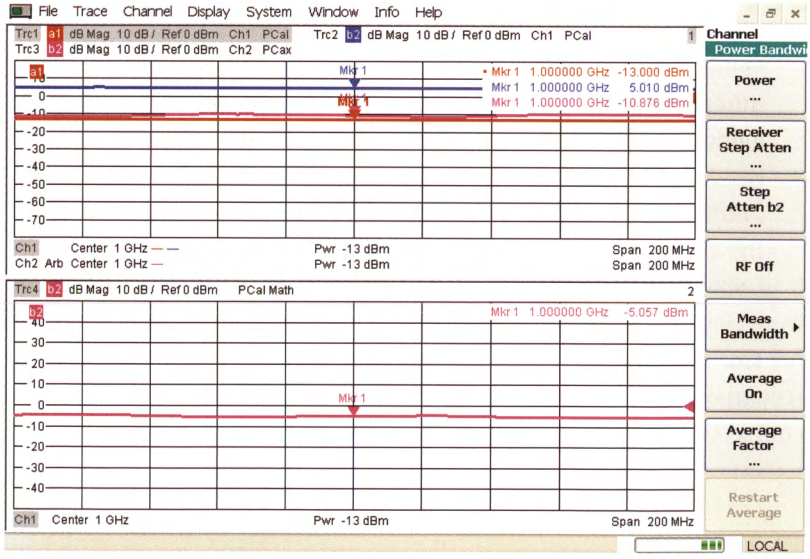


Fig. 7.5.7 Intermodulation measurement in the frequency range 900 MHz to 1.1 GHz on an amplifier.

Intermodulation effects occur in any analog signal processing chain that contains nonlinear components and is fed with at least two sinusoidal tones. Accordingly, it is impossible to prevent the network analyzer from also having a limited intermodulation suppression. Since signal generation is handled separately for the frequencies f_1 and f_2 and the signals are combined using a linear circuit component, we can assume that the resulting signal will be free of intermodulation as long as the generators are sufficiently decoupled by the combiner. If necessary, an attenuator pad can be inserted between the combiner and each generator to insure better decoupling and to prevent the generators from mutually intermodulate one another. Intermodulation in the network analyzer should then be encountered only in the measurement receiver for the test port that measures the combined signal. In the current example, this is test port 2 in Fig. 7.5.3. To asses the intermodulation suppression of the network analyzer, we can use two different approaches:

1. The receiver step attenuators can be used to decrease the level at the measurement receivers. Since these mechanical attenuators consist of passive resistor networks, they do not influence the intermodulation content of the measured signal. The input level at the measurement receiver can be reduced by increasing the attenuation of the step attenuator. This decreases the receivers contribution to the intermodulation. If this contribution dominates the measurement result, the change in attenuation will change the measured intermodulation suppression. If no change is observed in the trace, the measurement can be considered to be valid since the measurement result can be clearly attributed to the DUT.
2. The DUT is replaced by a through connection. The output level from the DUT, known from Trc2 in Fig. 7.5.7, is set as the source power for both channels. This means that the measurement receiver at test port 2 has the same level applied to it as in the measurement in Fig. 7.5.7. Because the input signal at test port 2 is free of intermodulation for the reasons stated above, the network analyzer measures its own intermodulation suppression (Fig. 7.5.8). To confirm that intermodulation in Fig. 7.5.7 is due to the DUT, we must verify that the intermodulation suppression of test port 2 is at least 10 dB better than the (expected) intermodulation suppression of the DUT. That is the case here.

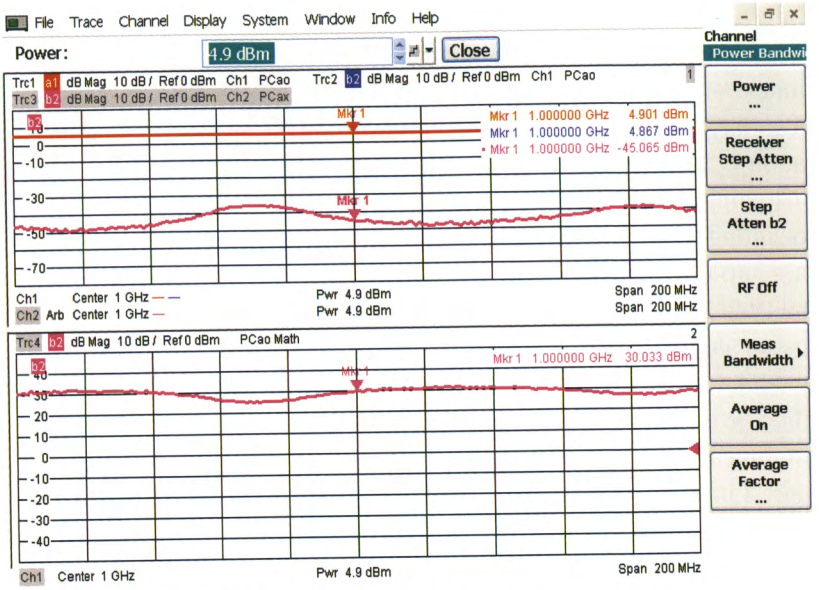


Fig. 7.5.8 Measurement of the network analyzer's intermodulation.

Of course, it would be nice to have some kind of correction mechanism to compensate intermodulation effects of the network analyzer. But the problem is that the intermodulation products of the DUT and the analyzer are superimposed in a rather complicated way. So far computational correction for this effect is not offered by vector network analyzers.

Example 2: Intermodulation measurement using a power sweep

Our second example will involve an IP3 measurement using a power sweep. The test setup and the analyzer settings are the same as in example 1. As the sweep type, we selected a power sweep with a center frequency of 1 GHz and a level range of -25 dBm to $+5$ dBm. We can repeat the power calibration to increase the measurement accuracy. In the obtained measurement result shown in Fig. 7.5.9, the saturation effect of the output levels (Trc2 and Trc3) is clearly visible. In this figure trace Trc4 was computed also for levels at which the DUT was in compression. An IP3 calculation based on the measured level ratios using formula (7.5-7) does not make any sense in the DUT's saturation range. In the current case, these are input levels $L_{\mathrm{IN}} > -16$ dBm. However, Fig. 7.5.9 already reveals a fluctuation of the IP3 below the saturation range. Behavior of this sort is typical of amplifiers based on field effect transistors.

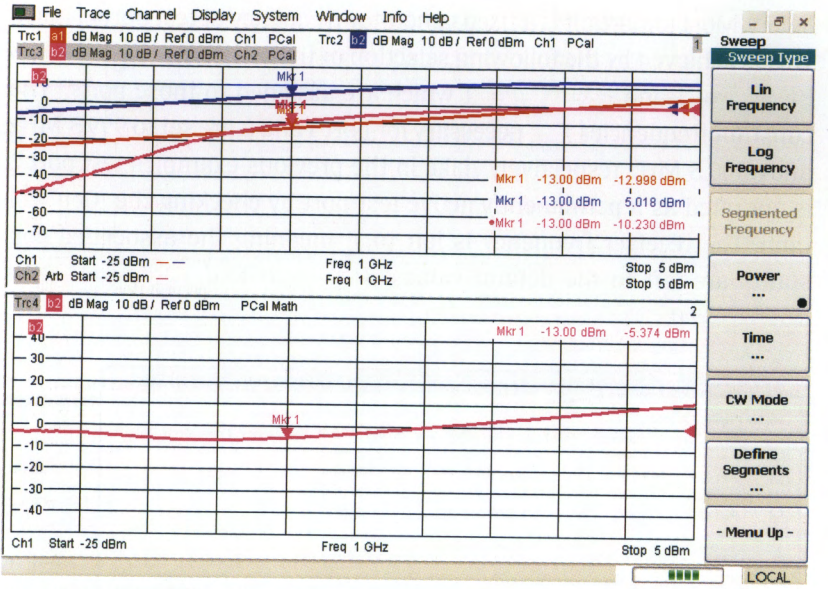


Fig. 7.5.9 Intermodulation measurement using a power sweep in the range -25 dBm to 5 dBm on an amplifier.

Example 3: Intermodulation measurement with constant stimuli

We can also use spectrum analyzers to make intermodulation measurements. Two additional signal generators are required which are normally operated at fixed frequencies, e.g. at $f_1 = 1\ \text{GHz} - 0.5\ \text{MHz}$ and $f_2 = 1\ \text{GHz} + 0.5\ \text{MHz}$. The power density spectrum is measured at the DUT's output. We can identify the spectral components shown in Fig. 7.5.1 at the frequencies $2f_1 - f_2$, f_1, f_2, $2f_2 - f_1$. Based on these values, we can then use formulas (7.5-2) to (7.5-6) to calculate the intermodulation suppression a_{IM3} and the intercept point IP3. Naturally, we must repeat this measurement for multiple frequency samples in order to assess the frequency-dependent behavior.

For the sake of completeness, we will demonstrate in the following section that this measurement can also be performed using a network analyzer. First, we must configure the frequency sweep so that the measurement receiver measures exactly at the frequencies $2f_1 - f_2$, f_1, f_2, $2f_2 - f_1$ (see formula (2.8-1)). In the current case, we set the instrument for 201 frequency samples and a sweep range from 997.5 MHz to 1002.5 MHz. We calculate a step size of 25 kHz, so that 4 sweep points exactly coincide with the frequencies at 998.5 MHz, 999.5 MHz, 1000.5 MHz and 1001.1 MHz. The required parameters for arbitrary mode are shown

below. Most important is a fixed generator frequency at test ports 1 and 3. This is achieved by the following selection of the constants: $d_{\text{numerator}} = 0$, $c_{\text{numerator}} = 1$ and $c_{\text{denominator}} = 1$ which are assigned to those ports. The constant frequencies f_{const} necessary for port 1 and 3 are set to 999.5 MHz and 1000.5 MHz respectively. Like in the previous examples test port 3 is operated as a permanently active test port by checking the "Gen" column. The receiver frequency is left to f_{b} meaning the associated constants are left to the default values $d_{\text{numerator}} = 1$, $d_{\text{denominator}} = 1$ and $c_{\text{numerator}} = 0$.

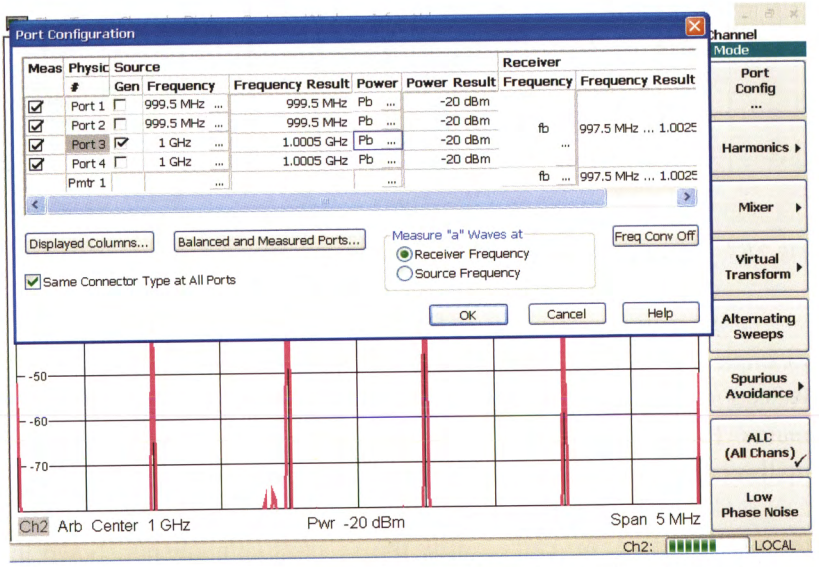


Fig. 7.5.10 *Parameters for arbitrary mode during an intermodulation measurement using fixed-frequency generators.*

Delta markers have been placed on the maxima of the trace shown in Fig. 7.5.11. They are useful for directly reading off the intermodulation suppression a_{IM3} and the frequency offset with respect to the frequency f_1.

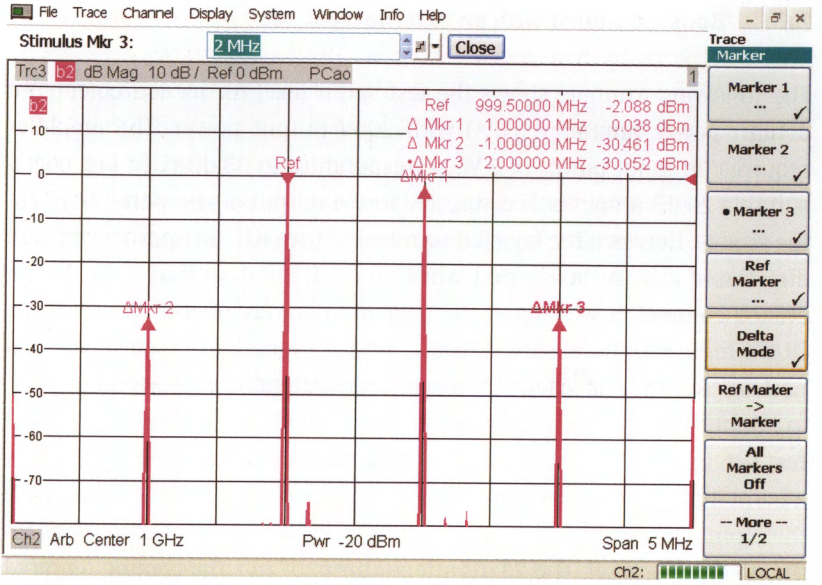


Fig. 7.5.11 Intermodulation measurement using fixed-frequency generators.

The configuration shown in Fig. 7.5.3 can be set up without direct generator and receiver access. However, if we also want to measure the input reflection factor of the DUT the combiner should not be located between test port 1 and the DUT. The "Source" connectors of test port 1 and 3 can be used as shown in Fig. 7.5.12 to locate the combiner in such a manner, that the combiner will not degrade the directivity of test port 1.

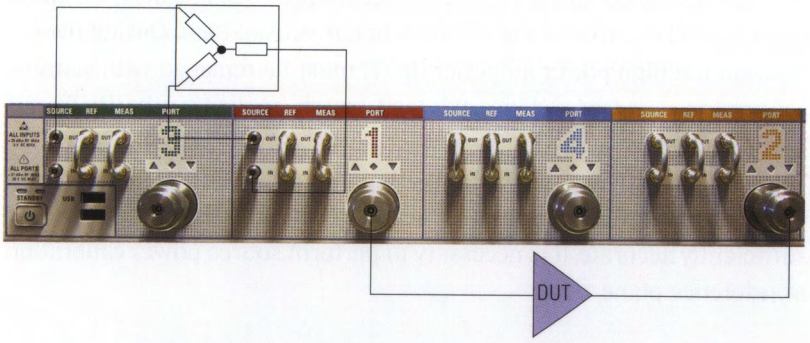


Fig. 7.5.12 Intermodulation measurement using the R&S®ZVA24 with direct generator and receiver access.

7.6 Boosted source with an external test set

The following example shows the test setup used for measurements on a **high power amplifier** (HPA) with 20 W output power. The amplifier requires an input power of 2 W, corresponding to 33 dBm. In Fig. 7.6.1, a suitable 25 dB amplifier is connected to the output of test port 1, to boost the source. Between the **boosted source** and the DUT, an appropriate dual directional coupler is inserted which is configured so that both the reference channel (a_1-wave) and the measurement channel (b_1-wave) of the DUT's input can be observed. The a_1 and b_1 signals are feed to the connectors "Ref In" and "Meas In" respectively. Attenuator pads can be used to adapt their levels. This part of the test setup is also called an **external test set**. In principle, the dual directional coupler can also be replaced by a combination of a power splitter and three-port directional coupler. With this combination, however, we should expect a higher insertion loss for the external test set. The 23 dB attenuator pad[1)] has the cooling required to absorb a large share of the DUT's output power before it reaches test port 2.

The attenuator pad at test port 2 results in a significant degradation in the raw directivity at this test port. Section 2.4.1 explains the reason for this degradation. The test setup Fig. 7.6.1 is not suited for s_{22} measurements due to the very limited raw directivity. For this reason the one-path two-port calibration technique described in section 3.4.6 is recommended. This incomplete calibration technique does not need the quantities s_{22} and s_{12}. That is the reason while it is used here. During the calibration the high power amplifier (DUT) must be replaced with suitable calibration standards[2)]. Since the load match applied to the DUT's output is not corrected by this technique, it should be as good as possible. This can be achieved by selecting a high-quality 23 dB attenuator pad with good matching. To ensure that the level of the stimulus signal is sufficiently accurate, it is necessary to perform source power calibration at reference plane 1.

1) Attenuator pads capable of handling a power of some watts (e.g. here at least 20W) typically have a designated input port. If the user interchanges the attenuators input and output port the power capability of the attenuator pad is reduced dramatically which can lead to burn out of the pad and to some secondary damage (e.g. damage to the analyzer, risk of fire).

2) Make sure the calibration standards meet the requirements in terms of their power handling capability.

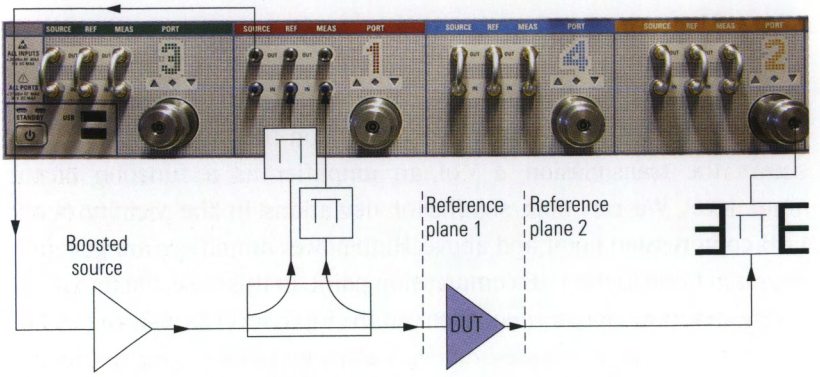


Fig. 7.6.1 Test setup with boosted source and external test set.

In Fig.7.6.2, the test setup has been extended to include a second external test set. This eliminates the directivity problem at test port 2 so that the techniques described in sections 3.4.2 to 3.4.4 can be used for calibration. This enables s_{22} measurements directly at reference plane 2. The load match provided by the non-ideal 23 dB attenuator pad is corrected in a linear manner as part of the system error correction. Read the next section for more details on interpreting of the measured quantity s_{22} for a high power amplifier.

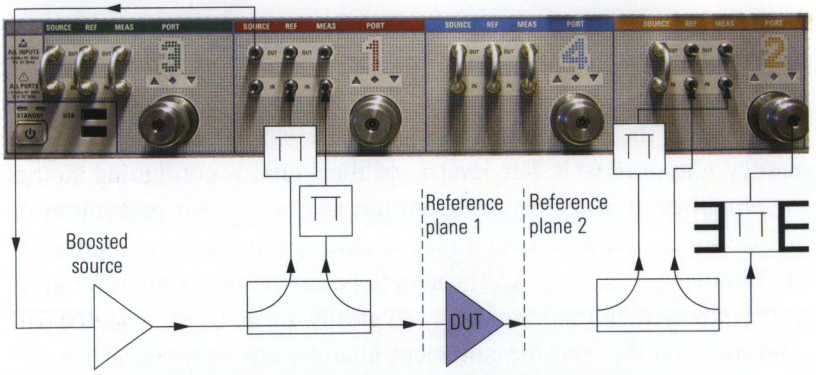


Fig. 7.6.2 Test setup with boosted source and two external test sets.

7.7 Measuring hot S-parameters

The characteristics of nonlinear components are level-dependent. Typically most important is their dependence on the input level. Fig. 7.2.2 shows the transmission a_{s21} of an amplifier as a function of the input level. We can note significant deviations in the vicinity of the 1 dB compression point and above. High-power amplifiers are generally operated close to the 1 dB compression point. In this case, the remaining S-parameters are mostly dependent on the input level as well. In Fig. 7.6.1, the parameter s_{11} was measured with 2 W input power. One alternative would be, for example, an s_{11} measurement without the boosted source, in which case the external test set could be omitted. Due to the lower level, however, this s_{11} measurement would not correspond to real operating conditions and the obtained result would not be useful in practical terms.

The parameter s_{22} is particularly important for optimizing the output match. In the measurement circuits we have discussed so far, this parameter is recorded during reverse operation. This means that the supply from port 1 is switched off and instead, we feed in a signal from port 2. As a consequence the amplifier does not receive any input signal during the s_{22} measurement so that the measurement does not correspond to the actual operating situation.

In the hot s_{22} measurement, the amplifier has an input signal at the frequency f_1 applied to it. The level L_1 of the signal is configured so that the amplifier supplies the proper output level L_{Output} for its application. At the same time, a reflection measurement is made at the output at the frequency $f_2 = f_1 + \Delta f$. The source power L_2 at test port 2 is significantly lower than the level L_{Output} (typically, $L_2 = L_{\text{Output}} - 30$ dB). The reference channel and measurement channel are recorded at the test port connected to the amplifier output at the frequency f_2. To ensure sufficient separation to the signal used at test port 1, the frequency spacing Δf should be equal to at least three to five times the value of the IF bandwidth. If the network analyzer allows a selection, we recommend using an IF filter type with high selectivity. Otherwise, the spacing Δf will need to be further increased. In the test setup shown in Fig. 7.7.1, the network analyzer from Fig. 7.1.6 was used. It has two internal generators. In the example, test ports 1 and 4 were selected since they can be simultaneously operated as active test ports with different frequencies.

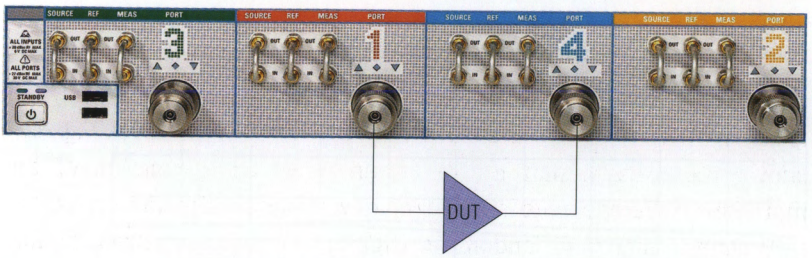


Fig. 7.7.1 Hot s_{22} measurement using the R&S®ZVA24 network analyzer.

In principle, we can also use other test ports. What is critical here is that both ports are assigned to different generators (see Fig. 7.1.6). Insofar as a boosted source and an attenuator pad are necessary, the test setup from Fig. 7.7.1 must be extended like in Fig. 7.6.2.

Fig. 7.7.2 shows the required settings in arbitrary mode for the hot s_{22} parameter measurement. The generator at test port 1 is permanently switched on for the measurement. The source power at test port 1 should be set to a fixed value of +10 dBm[1] in accordance with the normal operating state of our DUT. It is thus independent of the -25 dBm level used for the other test ports (specified using the POWER softkey). In our hot s_{22} parameter measurement, a level of -25 dBm is used at test port 2. The frequency offset of the source at test port 1 is set in accordance with formula (7.1-1) as $d_{\text{numerator}} = d_{\text{denominator}} = 1$ and $c_{\text{numerator}} = c_{\text{denominator}} = 1$ as well as $f_{\text{const}} = 10$ MHz. The 10 MHz frequency offset is clearly sufficient in conjunction with the 10 kHz IF bandwidth.

1) In the user interface shown in Fig. 7.7.2, there is a dialog box for the stimulus level like the one used to configure the stimulus frequency. It also has the constants found in formula (7.1-1). At test port 1, they were set as follows: $d_{\text{numerator}} = 0$; $c_{\text{numerator}} = 1$; $c_{\text{denominator}} = 1$ as well as $P_{\text{const}} = 10$ dBm.

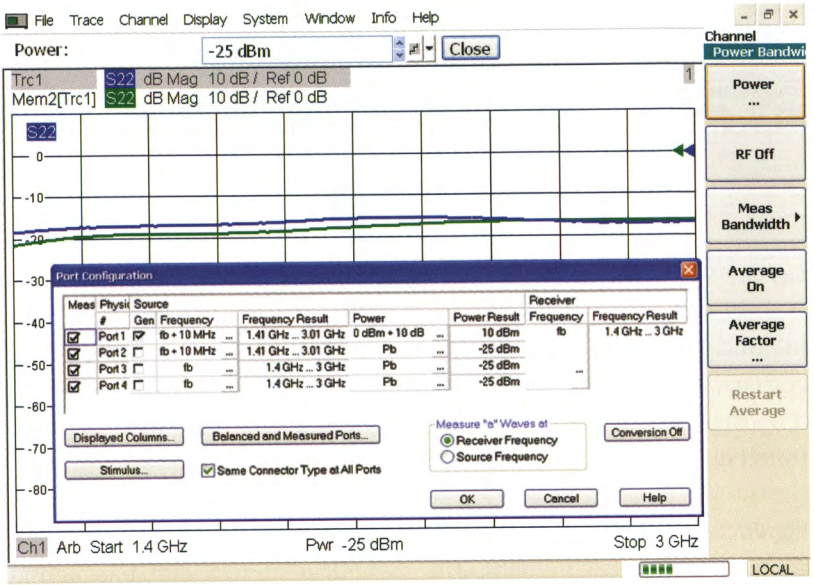


Fig. 7.7.2 *Example settings for arbitrary mode for the hot s_{22} measurement using the R&S®ZVA 24.*

7.8 Load-pull measurements

If the load impedance Z is suitably configured, it is possible to draw the maximum possible power P_A from a source. This is known as the **available power** P_A and the corresponding operating case is known as **power matching**. Provided a source impedance Z_S this case corresponds to a load impedance designed to be $Z = Z_S^*$ (maximum power theorem). In case of a linear source, the source impedance Z_S is constant and independent of the terminal quantities u and i. That is the reason, way it can be calculated from the source voltage u_s (open-circuit voltage) and a valid (u, i) pair of the terminal quantities within the fields low frequency design. In RF engineering a vector network analyzer must be used to determine the source reflection coefficient Γ_S based on this measurement the source impedance Z_s can be calculated. Nevertheless, the vector network analyzer takes advantage of the fact, that Z_s and in conjunction Γ_S is independent from the actual wave quantities (a, b).

The following figure shows the model of a nonlinear source, e.g. the output of a high power amplifier. It is an ideal voltage source with a nonlinear source impedance. The source impedance $Z_S(u, i)$ is dependent

on the terminal voltage u and the terminal current i. Both parameters are influenced by the load impedance Z that is used. A measurement of the source impedance $Z_S(u, i)$ in which a specific load impedance Z_0 is used can thus not be applied to operation with a different load impedance $Z \neq Z_0$.

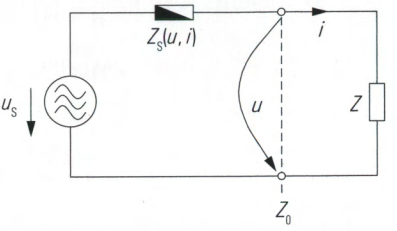


Fig. 7.8.1
Source with a nonlinear internal impedance.

By varying the load impedance Z using the so-called load-pull technique, we can set different operating states for u and i and then determine the actual delivered power P_d that is absorbed by the load. This level is assigned to the point in the Smith chart that corresponds to the used load impedance. It is customary to combine points with the same delivered power to form curves. The point in the Smith chart at which the greatest power P_d occurs corresponds to the case of power matching (which is the point ✶ in the next figure).

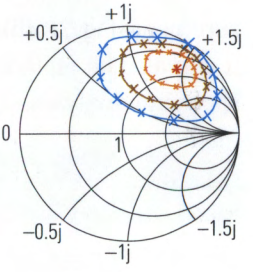


Fig. 7.8.2
Example of traces with equivalent delivered power.

To do the measurement described above, we need a mechanism that can synthesize arbitrary load impedances as necessary. It is important that the impedance must be physically applied to the DUT. Consequently, a change in the reference impedance using embedding, for example, is no alternative. For a load-pull measurement, we use a single-frequency stimulus at the frequency f_0. At the first sight, it seems like it will be sufficient to implement the required load in a narrow bandwidth at f_0. The synthetic load at f_0 must be configured so that if possible it can be used to track all of the points in the Smith chart. The best-known solution is a **tuner**. Its basic principle of operation is illustrated in Fig. 7.8.3a. It consists of an air line with a sliding contact arranged on the inner conductor. On this contact, a **stub** is connected that has a movable short. By

varying the two adjustment quantities x_{pos} (position of the sliding contact) and y_{pos} (position of the short relative to the sliding contact), the impedance Z_0 of the test instrument can be transformed to any arbitrary point in the Smith chart. In Fig. 7.8.3b, for example, it was transformed to the point ❷. However, losses in the setup prevents us from setting reflection coefficients $|\Gamma(f)| \approx 1$ (typically, the restriction is $|\Gamma(f)| < 0.95$).

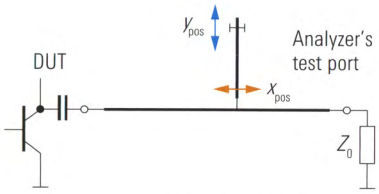


Fig. 7.8.3a *Basic principle of a tuner.*

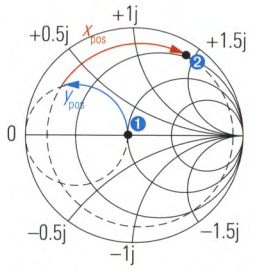


Fig. 7.8.3b *Transformation paths.*

Prior to the actual measurement, the tuner must be calibrated. To do this, connect its two ports to a network analyzer (Fig. 7.8.4a). The analyzer records the S-parameter s_{11} of the tuner for about 400 to 1600 different tuner settings (i.e. x_{pos} - / y_{pos} pairs of values). For the actual measurement (Fig. 7.8.4b), the network analyzer is not absolutely necessary. It can also be replaced by the combination of a signal generator and a spectrum analyzer.

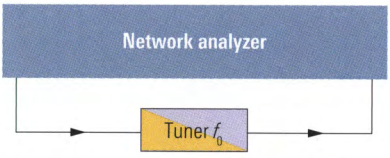


Fig. 7.8.4a *Calibration of the tuner.*

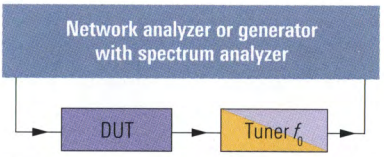


Fig. 7.8.4b *Test setup for load pull measurements.*

During the measurement (Fig. 7.8.4b), the reflection coefficient provided by the tuner cannot be monitored. The repeatability of the tuner is thus an important quality criterion. For the mechanical tuners described here, deviations of <-60 dB are typical. Using special interpolation algorithms in the control software, the second important parameter, i.e. the resolution of the setting grid, is significantly improved. Besides mechanical tuners, electronic implementations based on PIN diodes are also commercially available.

With the nonlinear DUT discussed here, additional harmonics at the output have to be considered. This is why it is also customary to synthesize additional impedances for the harmonics at $2f_0$ and $3f_0$. This is also done with a narrow bandwidth and called **harmonic tuning**. We have to extend the setup in Fig. 7.8.4b to include two additional special tuners. They are situated directly at the DUT. The following figure provides an example for the load impedance $Z(f)$ during harmonic tuning.

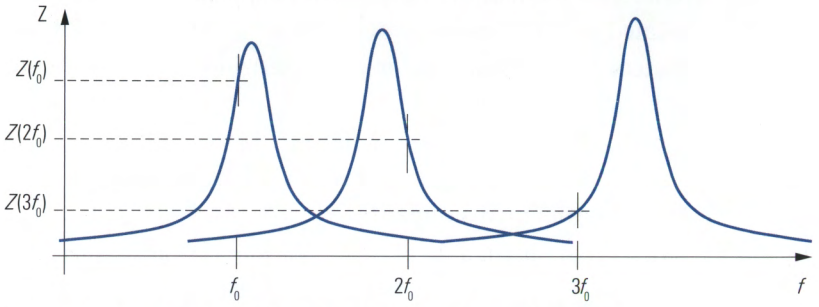


Fig. 7.8.5 Synthesis of an impedance during harmonic tuning.

By using an additional tuner at the amplifier's input, we are able to simulate and optimize the surrounding circuitry with greater precision. Fig. 7.8.6 shows a setup with four tuners. Since these devices are relatively expensive, this solution is used only in exceptional cases.

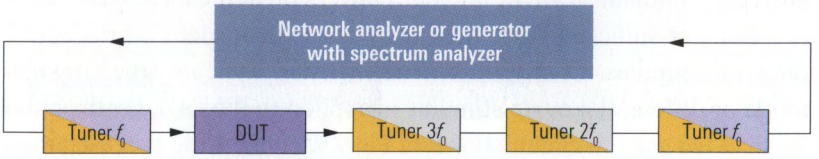


Fig. 7.8.6 Test setup for a load-pull measurement with harmonic tuning.

A combination of a hot S-parameter measurement and a load pull measurement is also possible. In technical literature [GB04], a setup is proposed that records all four hot S-parameters of an amplifier with a simultaneous load-pull measurement on the input and output. The hot S-parameters are referred to the fixed reference impedance Z_0. This enables direct comparison of the hot S-parameters for different load and source impedances. As a general rule, the hot S-parameters of a nonlinear DUT are dependent on the following:

- Measurement frequency f_0
- Source level P
- Source reflection coefficient $\Gamma_S(f_0)$

- Load reflection coefficients $\Gamma(f_0), \Gamma_S(2f_0), \Gamma_S(3f_0)$
- Biasing u_{Bias}, i_{Bias}
- Possible additional non-electrical quantities (e.g. temperature υ).

However, it is important to keep in mind that the hot S-parameters remain a small-signal model even if additional influence quantities such as the power P are included in the characterization. As is typical with small-signal models, the Taylor series expansion from formula (7.4-1) is truncated after the linear element (i.e. $k_{max} = 1$). Accordingly, the hot S-parameters are not suitable for characterizing the formation of higher-order harmonics, for example.

7.9 True-differential-measurements

Section 4.10 includes examples of balanced measurements on linear networks. These measurements are performed as single-ended measurements (i.e. unbalanced). The balanced measurement result is obtained after suitable conversion of the single-ended S-parameters. This is generally performed by the network analyzer and can be interpreted as a type of embedding.

With highly nonlinear DUTs, numeric conversion of the measured values is no longer sufficient since their behavior is dependent on the actual physical stimulus. In an unbalanced measurement, no true common mode or differential mode stimulus is applied to the DUT. On the other hand, if we use the circuit shown in Fig. 7.9.1, one of the two excitation modes (differential mode or common mode) can be generated depending on which of the unbalanced transmission lines we supply.

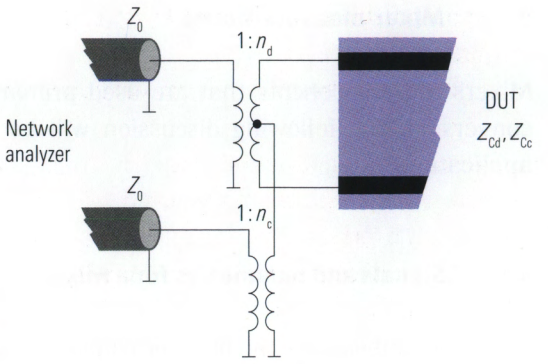


Fig. 7.9.1
Adaptation using transformers.

For the transformers, we calculate the turn ratios n_d and n_c as follows:

$$n_d = \sqrt{\frac{Z_{Cd}}{Z_0}} \qquad n_c = \sqrt{\frac{Z_{Cc}}{Z_0}} \qquad \textbf{(7.9-1/2)}$$

Note, however, that the nonideal properties of the transformers will introduce measurement errors. In addition, they are usually only available with a narrow bandwidth. On the network analyzer shown in Fig. 7.1.6, a true-differential-measurement mode is provided. The two physical test ports of a balanced port are assigned to different generators, which are operated at the same frequency but with a phase difference of 180°. In this manner, it is possible to generate a **true-differential-stimulus** at the balanced input port of the DUT. The problems with the transformers mentioned above are avoided. Since we can specify values besides 180° as the phase difference, a combination of common mode and differential mode stimulus is also possible. On the front panel of the network analyzer, the test ports are arranged so that the adjacent test ports 3 and 1 (or 2 and 4) are assigned to different generators so that they can each form a true-differential test port.

8 Mixer measurements

Mixers are components that are used primarily to handle frequency conversion. The following discussion will be limited to this primary application.

8.1 Signals and parameters for a mixer

This section discusses the basic principle of a mixer along with the relevant signals and their parameters. This should make it easier to understand section 8.2 and later sections. We will not discuss the possible implementations of mixers and their functioning.

8.1.1 Input and output signals of a mixer

A mixer is fed two signals at its inputs: A **radio frequency signal** (RF signal) and a **local oscillator** signal (LO signal). These two signals are characterized by the incident wave quantities a_{RF} and a_{LO}. The output signal from the mixer is the **intermediate frequency signal** (IF signal) and it is characterized by the wave quantity b_{IF}.

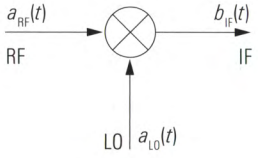


Fig. 8.1.1
Graphical symbol for a mixer with its intput and output signals.

Mixers are nonlinear components that are characterized to first approximation in the time domain as **multipliers**.

$$b_{\mathrm{IF}}(t) = Ma_{\mathrm{RF}}(t) \cdot a_{\mathrm{LO}}(t) \qquad \textit{(8.1-1)}$$

Let us assume that the **mixer constant** has a value $M = 1/\sqrt{\mathrm{W}}$. For cosine-shaped input signals in time domain $a_{\mathrm{RF}}(t) = \hat{A}_{\mathrm{RF}}\cos(2\pi f_{\mathrm{RF}}t + \varphi_{\mathrm{RF}})$ and $a_{\mathrm{LO}}(t) = \hat{A}_{\mathrm{LO}}\cos(2\pi f_{\mathrm{LO}}t + \varphi_{\mathrm{LO}})$, formula (8.1-1) yields the following IF signal:

$$b_{\mathrm{IF}}(t) = M\hat{A}_{\mathrm{RF}}\hat{A}_{\mathrm{LO}}\cos\left(2\pi f_{\mathrm{LO}}t + \varphi_{\mathrm{LO}}\right)\cos\left(2\pi f_{\mathrm{RF}}t + \varphi_{\mathrm{RF}}\right) \qquad \textit{(8.1-2)}$$

The figure below shows an example of the signal $b_{IF}(t)$. It was generated using two different input frequencies $f_{RF} \neq f_{LO}$.

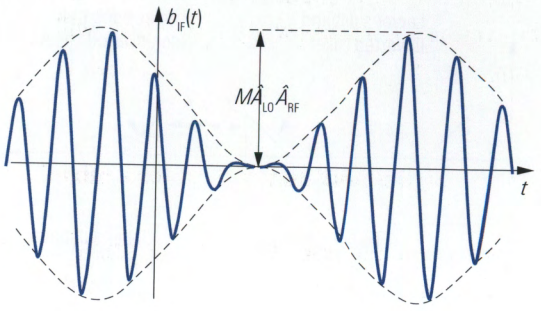


Fig. 8.1.2
Example of the IF signal vs. time $b_{IF}(t)$.

By rewriting $\cos(\alpha) \cdot \cos(\beta) = \frac{1}{2}[\cos(\alpha - \beta) + \cos(\alpha + \beta)]$, we obtain the following equations from formula (8.1-2):

$$b_{IF}(t) = \frac{M}{2} \hat{A}_{RF} \hat{A}_{LO} \left[\cos\left(2\pi(f_{LO} - f_{RF})t + \varphi_{LO} - \varphi_{RF}\right) + \right. \qquad \textbf{(8.1-3)}$$
$$\left. + \cos\left(2\pi(f_{RF} + f_{LO})t + \varphi_{LO} + \varphi_{RF}\right) \right]$$

or

$$b_{IF}(t) = \frac{M}{2} \hat{A}_{RF} \hat{A}_{LO} \left[\cos\left(2\pi(f_{RF} - f_{LO})t + \varphi_{RF} - \varphi_{LO}\right) + \right. \qquad \textbf{(8.1-4)}$$
$$\left. + \cos\left(2\pi(f_{RF} + f_{LO})t + \varphi_{LO} + \varphi_{RF}\right) \right]$$

Due to the symmetry of the cosine function, the two formulas (8.1-3) and (8.1-4) are equivalent. For the frequency of the IF signal, we have the following equations:

$$f_{IF} = |f_{RF} - f_{LO}| \qquad \textbf{(8.1-5)}$$

$$f_{IF} = f_{RF} + f_{LO} \qquad \textbf{(8.1-6)}$$

The following three figures illustrate the frequency relationships. A dashed line represents the LO frequency. Instead of the frequency f_{RF}, a frequency band is assumed that is shown as a dashed triangle. Solid triangles symbolize the resulting IF bands. If the orientation of the triangles for the RF and IF bands are the same, we use the term **noninverted case**. If the orientation is different, we use the term **inverted case**. The IF signal has two bands in each case. The higher frequency band is known

as the **upper sideband** (USB) and the lower frequency band is known as the **lower sideband** (LSB).

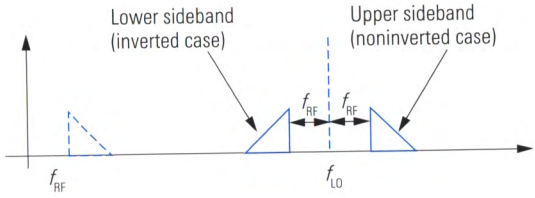


Fig. 8.1.3
Up-conversion in the inverted and noninverted case.

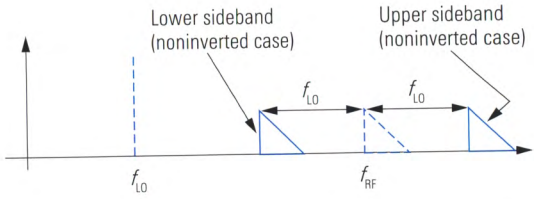


Fig. 8.1.4
Up and down-conversion in the noninverted case.

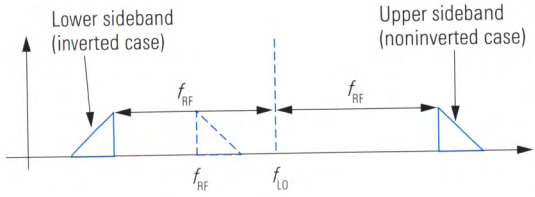


Fig. 8.1.5
Inverted case down-conversion and noninverted case up-conversion.

Using suitable filters, it is possible to select one of the sidebands. If the IF signal has a higher frequency than the RF signal, i.e. $f_{IF} > f_{RF}$, then we use the term **up-conversion**. Vice versa, the case where $f_{IF} < f_{RF}$ is known as a **down-conversion**. If a sideband in the inverted case is selected, then we speak of an **inverting conversion**. There exists a sideband in the inverted case for $f_{LO} > 2f_{RF}$ (up-conversion) and $f_{LO}/2 < f_{RF} < f_{LO}$ (down-conversion). In the remaining cases, we speak of a **noninverting conversion**.

If we use the mixer in a receiver, we have to keep in mind that the mixing process is not unambiguous. Assuming we would like to receive the frequency $f_{RF} = 200$ MHz, and that our IF frequency is 5 MHz. The LO frequency can be obtained using formula (8.1-5) as $f_{LO} = 200$ MHz $-$ 5 MHz, i.e. 195 MHz. In addition to the frequency $f_{RF} = 200$ MHz, however, we will also receive the frequency $f_{RF} = 200$ MHz -10 MHz. This situation becomes clear if we insert the calculated frequency f_{LO} into formula (8.1-6). In general, two different frequencies f_{RF} will be mixed to the same frequency f_{IF} for a given LO frequency. This phenomenon is known as **image reception**.

8.1.2 Higher order mixing products

When analyzing mixers, we can observe signal components that cannot be explained within the framework of the multiplier model. To understand this, we simply need to recall that mixers also include nonlinear components such as field effect transistors (FETs) and semiconductor diodes. These circuit components are exposed to a two-tone signal which results from the RF and LO frequency. The result is comparable to the components listed in table 7.5.1. In difference to section 7.5 the useful signal is now given by the second order intermodulation product. This signal corresponds to the IF signal derived by the simple mixer model of formula (8.1-1). The other spectral components are unwanted but they remain in the signal although a good mixer design involves frequency selective components to reduce their level. For a more general approach, we can use the following equations in place of formulas (8.1-5) and (8.1-6):

$$f_{IF} = |m \cdot f_{RF} - n \cdot f_{LO}| \qquad \textit{(8.1-7)}$$

$$f_{IF} = m \cdot f_{RF} + n \cdot f_{LO} \qquad \textit{(8.1-8)}$$

where $n, m = \{0, 1, 2, 3, 4, 5, ...\}$ *(8.1-9)*

The following measurement confirms formulas (8.1-7) and (8.1-8). The mixer output spectrum was measured using a fixed RF and a fixed LO frequency.

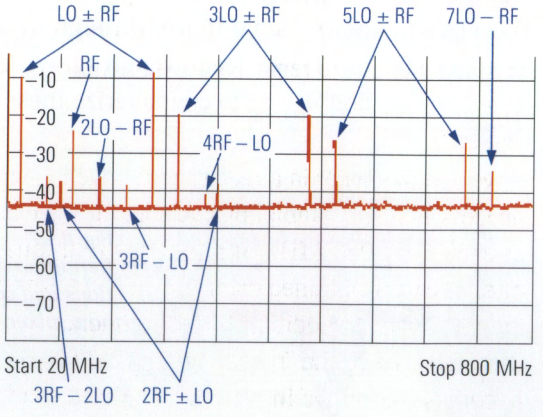


Fig. 8.1.6
IF spectrum, measured using fixed frequencies f_{RF} and f_{LO}.

The **higher order mixing products** from Fig. 8.1.6 can cause problems e.g. in a receiver. If we solve formulas (8.1-7) and (8.1-8) for f_{RF}, we will

find numerous reception frequencies f'_{RF} for the specified frequencies f_{IF} and f_{LO}.

$$f'_{RF} = \left| \frac{1}{m} \cdot f_{IF} \pm \frac{n}{m} \cdot f_{LO} \right| \qquad n,\ m \text{ see (8.1-9)} \qquad \textit{(8.1-10)}$$

These RF frequencies f'_{RF} where reception is not desired are known as **spurious responses.** Typical for them is that they can be received only if an RF signal is present at f'_{RF}. In contrast to this **inherent spurious responses** arise even if no RF signal is applied.

8.1.3 Important mixer parameters

Frequency range

It is difficult to implement mixers for a broad frequency range. This is why we assign specific ranges to the frequencies f_{RF}, f_{LO} and f_{IF}. If there is sufficient spacing between these different frequency ranges, we will be able to decouple the RF, LO and IF ports using frequency-selective components. In some cases, however, it is necessary to have an **overlap** between the different frequency ranges (particularly the RF and LO signals). This entails additional complexity if we want to ensure sufficient decoupling between the signals. To achieve, the desired isolation (e.g. between the RF and LO signals) a balanced setup is normally used inside the mixer.

Isolation

The Fig. 8.1.7 shows the **LO-IF feedthrough** (red), the **RF-IF feedthrough** (orange) and the **LO-RF feedthrough** (gray). We define the isolations d_{LO-IF}, d_{RF-IF} and d_{LO-RF} to characterize these undesired crosstalk processes.

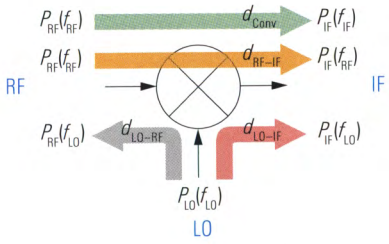


Fig. 8.1.7
Definition of the isolations and conversion loss on a mixer.

Take the **LO-IF isolation** d_{LO-IF} as an example. It describes the crosstalk between the LO port and the IF port. The definition is based on the LO

power $P_{LO}(f_{LO})$ that is measured at the LO frequency f_{LO} and the portion of power $P_{IF}(f_{LO})$ that is transmitted to the IF port due to the crosstalk process. The power $P_{IF}(f_{LO})$ is measured at the IF port at the frequency f_{LO} meaning no frequency conversion is assumed within this crosstalk. Based on the powers $P_{LO}(f_{LO})$ and $P_{IF}(f_{LO})$ we can form a power ratio that is typically specified using a dB scale. If we assume that the mixer is properly terminated with the required reference impedance, we can also calculate the LO-IF isolation based on the wave quantities $a_{LO}(f_{LO})$ and $b_{IF}(f_{LO})$. The **RF-IF isolation** d_{RF-IF} and the **LO-RF isolation** d_{LO-RF} are defined in an similar manner. They are also based on a power ratio with both powers forming the ratio that are measured at the same frequency f_{RF} or f_{LO} respectively.

$$d_{LO-IF} = -10\lg\frac{P_{IF}(f_{LO})}{P_{LO}(f_{LO})}\text{dB} = -20\lg\frac{|b_{IF}(f_{LO})|}{|a_{LO}(f_{LO})|}\text{dB} \qquad \textbf{(8.1-11)}$$

$$d_{RF-IF} = -10\lg\frac{P_{IF}(f_{RF})}{P_{RF}(f_{RF})}\text{dB} = -20\lg\frac{|b_{IF}(f_{RF})|}{|a_{RF}(f_{RF})|}\text{dB} \qquad \textbf{(8.1-12)}$$

$$d_{LO-RF} = -10\lg\frac{P_{RF}(f_{LO})}{P_{LO}(f_{LO})}\text{dB} = -20\lg\frac{|b_{RF}(f_{LO})|}{|a_{LO}(f_{LO})|}\text{dB} \qquad \textbf{(8.1-13)}$$

Conversion loss

The conversion loss is a measure of the losses that arise for the desired frequency conversion. This quantity is the ratio of the power at the RF port at the frequency f_{RF} and the power at the IF port at the frequency f_{IF}. The conversion loss is usually specified in dB and typical values are 6 dB to 7 dB.

$$d_{Conv} = -10\lg\frac{P_{IF}(f_{IF})}{P_{RF}(f_{RF})}\text{dB} = -20\lg\frac{|b_{IF}(f_{IF})|}{|a_{RF}(f_{RF})|}\text{dB} \qquad \textbf{(8.1-14)}$$

The conversion loss is frequency dependent typically. With insufficient LO power an abnormally high conversion loss can be seen, meaning the mixer does not work as it was designed for.

Return loss

As is the case with other three-port networks, we can also define three return losses for a mixer. We have the return losses d_{RF} at the RF port, d_{IF} at the IF port and d_{LO} at the LO port. However, it is not customary to specify the return loss d_{LO}.

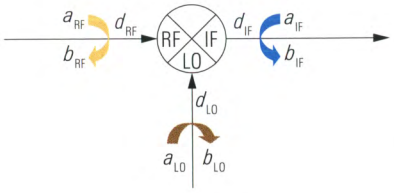


Fig. 8.1.8
Three return losses on a mixer.

The return losses that occur are frequency-dependent for the most part. In addition, the LO level has a major influence on the return losses. This will be illustrated within the later shown example in section 8.3.

$$d_{IF} = -20 \lg \frac{b_{IF}(f_{IF})}{a_{IF}(f_{IF})} \qquad \textit{(8.1-15)}$$

$$d_{RF} = -20 \lg \frac{b_{RF}(f_{RF})}{a_{RF}(f_{RF})} \qquad \textit{(8.1-16)}$$

$$d_{LO} = -20 \lg \frac{b_{LO}(f_{LO})}{a_{LO}(f_{LO})} \qquad \textit{(8.1-17)}$$

Compared with other components, the ports of a mixer tend to be poorly matched. This means that the circuitry around the mixer needs to have the best possible matching to prevent multiple reflections. This is particularly critical for the IF and RF ports! Manufacturers often recommend a minimum matching that can sometimes be referenced to an impedance Z_0 different from 50 Ω. For the IF port, the matching should be ensured not only for the entire IF frequency range but also for the LO and RF feedthroughs.

1 dB compression point

The input power required for this quantity is measured at the RF port at the frequency f_{RF}. The associated output power is determined at the IF port at the corresponding IF frequency f_{IF}. There are mixers for different LO levels. Mixers that are designed to handle a higher LO level generally have a higher 1 dB compression point.

Intermodulation distortion, intercept points

To measure the intercept point, we need to apply a two-tone signal with the frequencies f_{RF1} and f_{RF2} to the RF input of the mixer. The frequency spacing $f_{RF2} - f_{RF1}$ and the associated RF power levels $P(f_{RF1})$ and $P(f_{RF2})$ should be chosen in accordance with on the guidelines described in section 7.5.1. Corresponding to the frequency conversion

desired on the mixer, the output frequencies f_{IF1} and f_{IF2} at the IF port are associated with the frequencies f_{RF1} and f_{RF2}. Consequently the second and third order intermodulation products are expected at $f_{IF2} \pm f_{IF1}$ and at $2f_{IF2} - f_{IF1}$ or $2f_{IF1} - f_{IF2}$. The intercept points are calculated in accordance with the description in section 7.5.

LO level

Mixer manufacturers usually recommend an optimum LO level $P_{LO}(f_{LO})$ and refer their specifications for the conversion loss, isolation and so on to this level. Typical LO levels are in the range from 6 dBm to 20 dBm.

RF level

The RF level is generally 3 dB to 10 dB less than the LO level. At very low RF levels (typically <-50 dBm), we will generally observe a poor signal-to-noise ratio at the IF output. Using a suitable IF filter (bandpass filter), we can suppress the broadband component of the noise outside the passband around f_{IF}, thereby improving the signal-to-noise ratio (see Fig. 2.8.5). If we want to specify a minimum RF level, we need to take into account the entire signal processing chain in which the mixer is embedded. An important quantity in this context is the noise factor of the mixer. It has roughly the same numerical value as the conversion loss. The maximum level of the RF signal is determined by the 1 dB compression point, the intercept point and the spectral purity requirements of the application. In addition, there exists a maximum RF level for every mixer known as the **mixer burnout limit**. Typical values are in the range from 20 dBm to 30 dBm. For pulsed applications in which the power of the RF signal is activated only for a defined time interval, it is also customary to specify the **maximum impulse level**.

Linear distortion, group delay

Besides nonlinear distortion, we are also interested in the linear distortion of the mixer. This includes the frequency response of the conversion loss and the group delay. For the definition and interpretation of the group delay τ_G, see section 4.5. The frequencies f_{RF} and f_{LO} determine the output frequency f_{IF} of the mixer. As can be seen from formulas (8.1-3) and (8.1-4), this dependence applies also to the phases φ_{IF}, φ_{RF} and φ_{LO}. For example, we can calculate $\varphi_{IF} = \varphi_{RF} - \varphi_{LO}$, yielding a constant group delay τ_G. Notice that, formulas (8.1-3) and (8.1-4) are based on an ideal multiplier model, while the actual group delays measured may show a different behavior.

When measuring mixers you will sometimes face some special type of mixer: There are some special mixers which use a **biasing current** that is typically less than some mA. Using this current, the mixer can cope with a reduced LO level. Some of these mixers are even **active mixers**, meaning they do show a conversion gain, meaning $d_{\mathrm{Conv}} < 0$. A **harmonic mixer** (also called **subharmonic mixer**) uses a higher order mixing product e.g. $f_{\mathrm{IF}} = 8 \cdot f_{\mathrm{LO}} - f_{\mathrm{RF}}$ offering the advantage that the LO frequency can be reduced to a fraction of its normal value. But this type of mixer exhibits a higher conversion loss. A **sideband separation** mixer or an **image rejection mixer** offers dedicated access to the lower or upper side band. An **I/Q mixer** is used with an IF frequency of 0 Hz typically. It has two IF ports for the inphase component and the quadrature component.

8.2 Features for mixer measurements

8.2.1 Mixer measurement mode

In principle, we can make mixer measurements using the arbitrary mode. However, a special **mixer mode** does simplify the process considerably. It should include the following:

- Specification of frequency relationships at the mixer
- Determination of power levels at the RF and LO ports
- Handling of a mixer power calibration

The frequency sweep and power sweep modes are most commonly used. We will now have a closer look at them.

Frequency sweep

If the frequency at the RF port is specified as an interval and a fixed LO frequency is used, we can determine the frequency range at the IF port using formulas (8.1-5) and (8.1-6). However, it is also customary to choose a fixed IF. The network analyzer then has to adjust the frequency of the LO signal in a suitable manner. In the dialog box for mixer mode in Fig. 8.2.1, the necessary computations are performed automatically. It even supports a fixed RF frequency although this is not a typical case. The frequency range specified by the user as "Swept" determines the stimulus axis. The frequency range specified as "Auto" is computed by the network analyzer based on the "Swept" and "Fixed" frequencies.

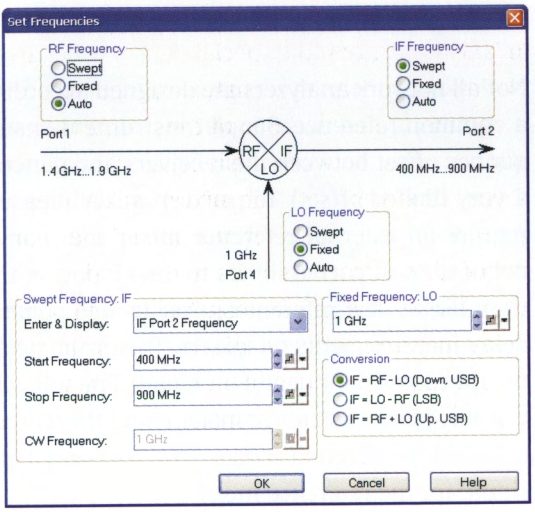


Fig. 8.2.1
Configuration of the frequency ranges for a frequency sweep with a fixed IF.

Power sweep

The level can be swept at one of the mixer's input ports. A fixed level is specified for the other input port. The mixer's IF port is generally operated as a **passive test port** (exception: measurement of the return loss at the IF port).

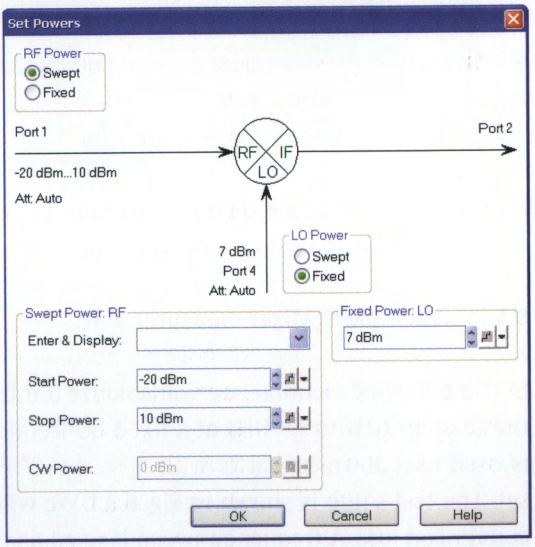


Fig. 8.2.2
Configuration of the level ranges for a power sweep with a fixed LO level.

8.2.2 Connection of a reference mixer

Not all network analyzers are designed according to Fig. 2.6.3a involving a common reference. Simpler instruments usually cannot handle a frequency offset between the receivers and generator (or they handle only a very limited offset). To perform mixer measurements, such analyzers require an external **reference mixer** that converts the reference channel of the test port assigned to the RF port of the DUT (see following figure). Usage of a reference mixer is also common when making group delay measurements on mixers. The group delay of the reference mixer must be known (or negligibly small). The following figure shows how we can integrate a reference mixer using the "Direct generator and receiver access" introduced in section 7.1.7. Test port 4 is used to generate the LO signal for the reference mixer and the DUT.

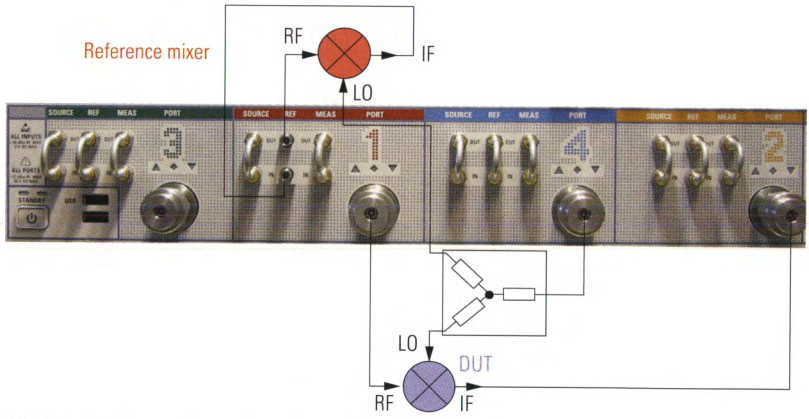


Fig. 8.2.3 Connection of a reference mixer.

8.3 Example 1: Mixer measurement

In the following example, we will analyze a mixer with an RF frequency range of 1.4 GHz to 1.9 GHz at a fixed LO frequency of 1 GHz. This mixer is used as a down-converter with $f_{IF} = f_{RF} - f_{LO}$. An LO level of 7 dBm is set. The test setup is shown in Fig. 8.3.1. We will first look at the mixer's conversion loss. A frequency sweep is used as the sweep mode.

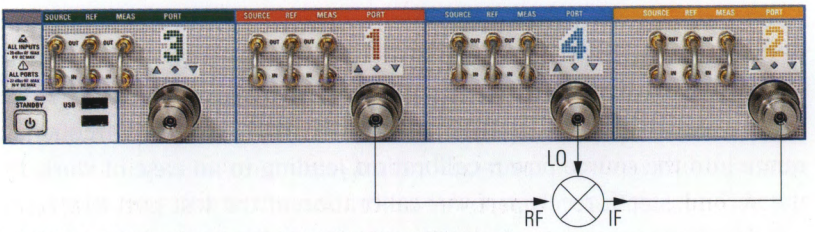


Fig. 8.3.1 Test setup for scalar mixer measurement.

The frequency ranges are defined using the mixer mode shown in Fig. 8.2.1. After setting it to "Swept", the intermediate frequency (IF) provides the frequency axis for the measurement. It is configured for 400 MHz to 900 MHz. The RF frequency is computed automatically when "Auto" is selected. The LO is set to "Fixed" at 1 GHz. Using the main dialog of the mixer measurement (not shown) we select the test port 4 to be our LO port. The dialog box in Fig. 8.2.2 is used to set the LO and RF powers. Both of these quantities are configured as "Fixed" powers.

To measure the compression behavior, we must create a new trace in a separate diagram area. The trace is assigned to channel Ch2 and the power sweep mode is activated for channel Ch2. In the dialog box in Fig. 8.2.2, a power sweep range from –20 dBm to 10 dBm is specified. The RF level is swept. The LO level has a fixed value of 7 dBm. To obtain the highest possible level accuracy, mixer power calibration is performed on both channels Ch1 and Ch2 using the dialog box in Fig. 8.3.2.

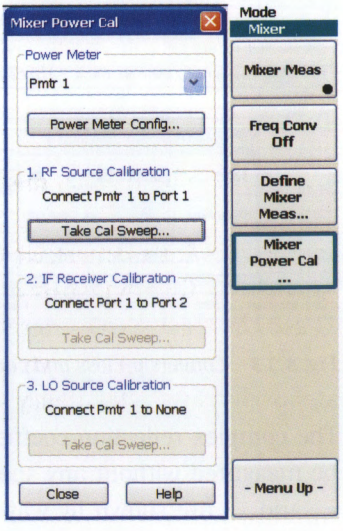


Fig. 8.3.2
Dialog box for mixer power calibration.

To perform a mixer power calibration, a power meter is first connected to the network analyzer instead of the mixer's RF port. At this port a source power calibration that includes also the IF range is done. The Dialog box shown in Fig. 8.3.2 automatically incorporates the IF frequency range into the source power calibration, leading to an ease of work. In the second, step a receiver power calibration at the test port that later will be connected to the mixers IF port is done. This receiver power calibration is based on the previous source power calibration, consequently the test ports which are normally connected to the mixers RF and IF port are now connected by a low loss through connection. The third and last step is a source power calibration of the test port that will provide the LO signal for the mixer measurement. To carry out this calibration a power meter is connected to the test port instead of the mixer's LO port.

Fig. 8.3.3 shows the measurement results. If necessary, the diagrams can be given a **title**. The input field is shown additionally in Fig. 8.3.3. The compression point can be calculated automatically using the statistical functions. For details, see section 7.2.

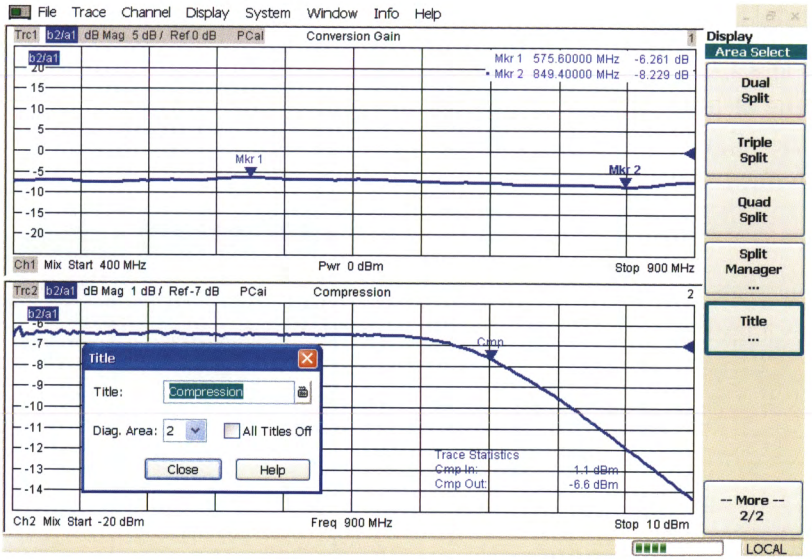


Fig. 8.3.3 Conversion loss and compression point of a mixer.

The complex reflection coefficient at the three ports of the mixer can be measured without any special effort. In Fig. 8.3.4, the S-parameters s_{11} and s_{22} are shown by way of example. As seen from Fig. 8.3.1, they correspond to the reflection coefficients at the RF and IF ports of the

mixer. After a one-port calibration (OSM calibration technique), these measured values can be system error corrected. When measuring these S-parameters, it is important for the LO port to be supplied with the appropriate LO level. This must be ensured either by using an external generator or a suitable configuration of the network analyzer. Each of the diagrams in Fig. 8.3.4 contains a measurement without an LO signal as a red trace to show the significant differences.

To measure the RF-IF isolation, we can make an s_{21} measurement. Using a frequency sweep, the transmission in the RF frequency range is measured without frequency conversion. For this measurement a system error correction is possible. It is performed as a two-port calibration, e.g. using the TOM calibration procedure, between test ports 1 and 2. During the measurement, the necessary LO level must be applied to the LO port on the mixer. Since the LO frequency at 1 GHz lies within the RF frequency range, the LO feedthrough appears as a spike in the trace (see the lower part of Fig. 8.3.4). For the sake of illustration, an isolation measurement without a LO signal was also performed for Fig. 8.3.4 (red trace).

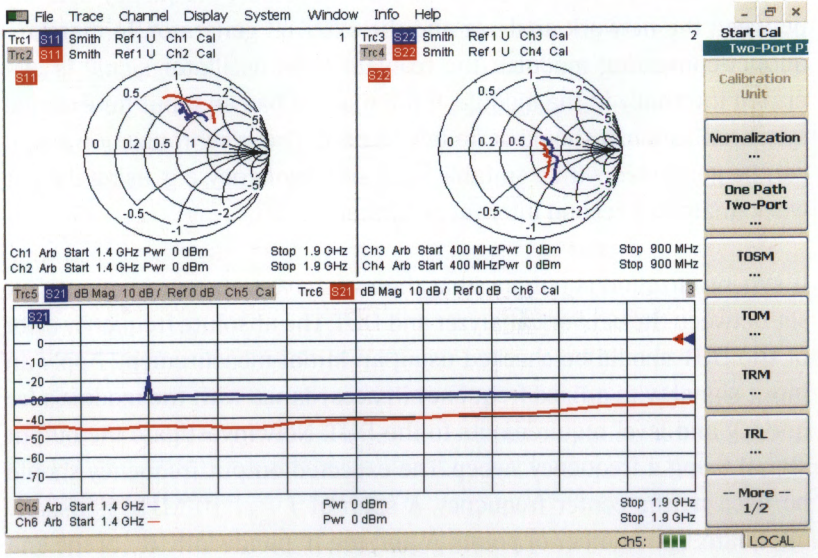


Fig. 8.3.4 Complex reflection coefficient and isolation on a mixer.

8.4 Example 2: Measurements on a down-converting module

In chapter 1, the RF EXTENSION module was introduced in Fig. 1.4.1. We would like to investigate its frequency-converting transmission characteristics for an RF frequency range from 3.9 GHz to 4.1 GHz. The measurement result is obtained at the IF port in the range from 304.4 MHz to 504.4 MHz. Fig. 8.4.1 shows the bandpass behavior of this module. The trace is similar to the one in Fig. 1.4.3 in which only the bandpass filter was analyzed. Using the displayed limit lines, we can check whether the behavior of the module fulfills the relevant requirements. The result of the tolerance check for trace Trc1 will be either a green "PASS Trc1" or a red "FAIL Trc1" in case of a tolerance violation. The entire measurement is automatically configured with the instrument running under remote control. The result of the tolerance check is available as a digital signal or via remote control commands. Using the "Go to Local" softkey, you can terminate remote control operation and make manual entries.

In the current measurement, an external generator provides the LO signal for the module. A common 10 MHz reference frequency (synchronization similar to Fig. 7.1.11) provides the necessary frequency stability between the network analyzer and the external generator. In many frequency-converting modules, the relevant local oscillator signal is generated internally in the module. If the module has an input for external synchronization, it should definitely be used. The synchronization signal can be generated with a suitable level and frequency, e.g. using the network analyzer's second internal generator.

If synchronization is not possible, then we have to face a frequency offset between the network analyzer and DUT. The absolute frequency error of the DUT should be checked using an initial measurement. Apply an input signal with a fixed frequency in accordance with the module's frequency and level requirements to the DUT. Now investigate the output signal using a frequency sweep. The expected output frequency should be used as the center frequency. A span of $f_{\text{span}} = 10\ \text{MHz}$ is usually adequate. The number of points N and the IF bandwidth B_{IF} of the analyzer should be chosen so that the condition $f_{\text{span}} < N \cdot B_{\text{IF}}$ is fulfilled, e.g. $B_{\text{IF}} = 100\ \text{kHz}$, $N = 201$. The frequency error can be read off by comparing the actual and expected output frequency. It must be taken into account during subsequent measurements. When configuring the frequency axis with incorporating the frequency error, we can neglect the

condition $f_{span} < N \cdot B_{IF}$ for the remainder of the measurement. The short-term stability of the DUT is generally not a problem since the oscillators that are commonly used nowadays are linked to a crystal reference via a phase-locked loop. For IF bandwidths up to 1 kHz, the spurious FM that is associated with the short-term stability is usually negligible.

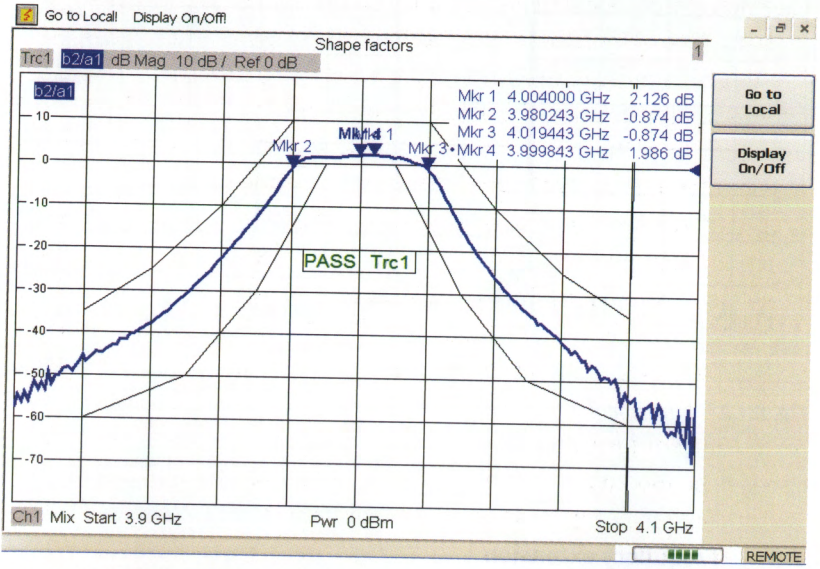


Fig. 8.4.1 *Measurement involving frequency conversion on the RF EXTENSION module from Fig. 1.4.1.*

The intercept point IP3 is also measured customarily on a mixer or, as in the current case, on a module that performs a frequency conversion. An additional signal source and a combiner element are both necessary (section 7.1.10). Fig. 8.4.2 shows an intermodulation measurement on the RF EXTENSION module using a frequency spacing $f_{RF2} - f_{RF1}$ of 2 MHz. Besides the useful signals at f_{IF1} and f_{IF2}, the figure also reveals the level of the intermodulation products $2f_{IF2} - f_{IF1}$ and $2f_{IF1} - f_{IF2}$. Like in measurement example 3 in section 7.5.2, we must configure the frequency sweep so that the measurement receiver precisely measures the frequency samples f_{IF1}, f_{IF2}, $2f_{IF2} - f_{IF1}$ and $2f_{IF1} - f_{IF2}$.

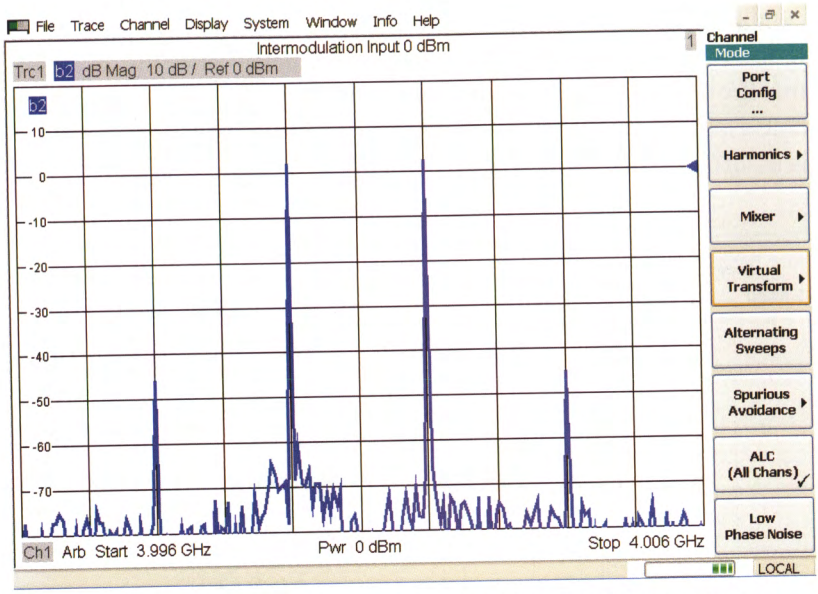


Fig. 8.4.2 *Intermodulation measurement on the RF EXTENSION module from Fig. 1.4.1.*

8.5 Frequency extension

The combination of a measurement involving a frequency conversion and an external test set is known as a **frequency extension**. This is a way of extending a suitable vector network analyzer to a maximum operating frequency of several hundred GHz. The test ports of the **frequency extension modules** and the directive element integrated in them are typically implemented using rectangular waveguides. The operating frequency range of the module is then mostly determined by the frequency range of the waveguides H_{10}-mode. The following table gives an overview of these typical operating frequencies and the related waveguide dimensions.

Operating frequency in GHz[1]	Waveguide band[2]	Cutoff frequency f_c in GHz	Waveguide designation		Inner dimensions	
			EIA (US)[3]	RCSR (UK)	in mm	in inches
50 – 75	V	39.8616	WR-15	WG-25	3.7592 × 1.8796	0.148 × 0.074
60 – 90	E	48.3567	WR-12	WG-26	3.0988 × 1.5494	0.122 × 0.061
75 – 110	W	58.9951	WR-10	WG-27	2.5400 × 1.2700	0.100 × 0.050
90 – 140	F	73.7439	WR-8	WG-28	2.0320 × 1.0160	0.080 × 0.040
110 – 170	D	90.7617	WR-6	WG-29	1.6510 × 0.8255	0.065 × 0.0325
140 – 220	G	115.6767	WR-5	WG-30	1.2954 × 0.6477	0.051 × 0.0255
170 – 260	(Y)	137.1980	WR-4	WG-31	1.0922 × 0.5461	0.043 × 0.0215
220 – 325	J(H)	173.5151	WR-3	WG-32	0.8636 × 0.4318	0.0340 × 0.0170
325 – 500	Y	268.1596	WR-2.2	–/–	0.5588 × 0.2794	0.0220 × 0.0110
500 – 750		393.3008	WR-1.5	–/–	0.3810 × 0.1910	0.0150 × 0.0075
750 – 1100		589.9512	WR-1	–/–	0.2540 × 0.1270	0.0100 × 0.0050

Table 8.5.1 *Important waveguide frequency bands in accordance with UK and US standards.*

The frequency extension modules generate the stimulus signal by performing a frequency multiplication on the generator signal from the network analyzer. Instead of frequency multiplication, the stimulus signal can also be generated using suitable mixing processes. The reference and measurement channels separated by the directional element (e.g. a dual directional coupler) are mixed to a fixed IF frequency that is typically in the range 15 MHz to 300 MHz. The modules include suitable harmonic mixers for this purpose. The necessary LO signal is generated by either an external generator or the network analyzer's second generator. In the following figures, the latter case applies.

1) As we went to press, commercial frequency extension modules were available up to a frequency of 500 GHz.
2) The lettering scheme applies solely to waveguides! In addition, various divergent schemes for frequency bands are commonly used in radar technology and in the standards of different American companies. The letters listed here starting with WR-4 are not used in a uniform manner. The indications in ( ) comply with the nomenclature used by Oleson Microwave Labs, for example.
3) The WR-2.2, WR-1.5 and WR-1 waveguide designations are provided only as a logical continuation of the standard. As we went to press, there was no final specification yet.

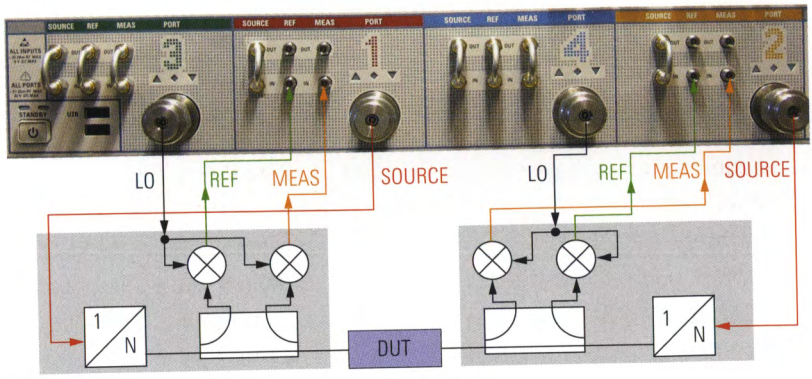


Fig. 8.5.1 Frequency extension on an R&S®ZVA24.

The configuration shown in Fig. 8.5.1 enables both a forward and a reverse measurement. To reduce the technical complexity, unidirectional test setups are also common. In this case the module located at the DUT output includes only a harmonic mixer to record the measurement channel (see Fig. 8.5.2). The waveguide directional coupler, the mixer for the reference channel and the stimulus signal processing for this module are no longer necessary. However, complete system error correction is not possible under these circumstances. If we want to correct the influence of the load match, we need to measure the values for the parameters s_{12} and s_{22}. However, the simplified module does not measure these quantities. To obtain sufficient measurement accuracy, the simplified module and its connection to the reference plane must be configured to form a match to the DUT that ensures as little reflection as possible. A precision attenuator pad can be integrated into the simplified module for this purpose.

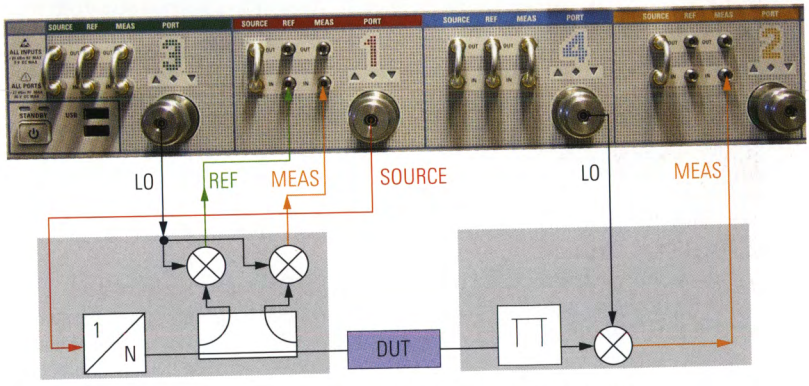


Fig. 8.5.2 Unidirectional frequency extension on an R&S®ZVA24.

Measurement example 1: We will investigate a 1" long waveguide section of type WR03 over its frequency range from 220 GHz to 325 GHz. Our frequency extension modules are type V03VNA2-T/R modules from Oleson Microwave Labs. The vector network analyzer must include the "External generator and receiver access" option to allow the cabling shown in Fig. 8.5.1. The network analyzer can be configured using arbitrary mode. In the present case, we chose a fixed receive frequency of 25.44 MHz at the test ports. The modules that are used have a multiplication factor of 18 for generation of the stimulus signal. The stimulus frequency at the analyzer's test ports 1 and 2 is thus equal to $^1/_{18} \cdot f_b$. The mixers in the modules operate in harmonic mode and use the 18th harmonic of the signal designated as "LO" in Fig. 8.5.1. The stimulus frequency at test ports 3 and 4 should thus be specified as $^1/_{18} \cdot f_b - {}^1/_{18} \cdot 25.44$ MHz. These test ports are operated as active ports regardless of the measurement direction. After configuring the arbitrary mode, the frequency axis is set to f_b = 220 GHz to 325 GHz. On the network analyzer, we chose low-phase-noise mode and an output power to suit the modules. The IF bandwidth is reduced as required. For calibration purposes, we use a waveguide calibration kit and the TRL calibration procedure. The following figure shows the complete test setup.

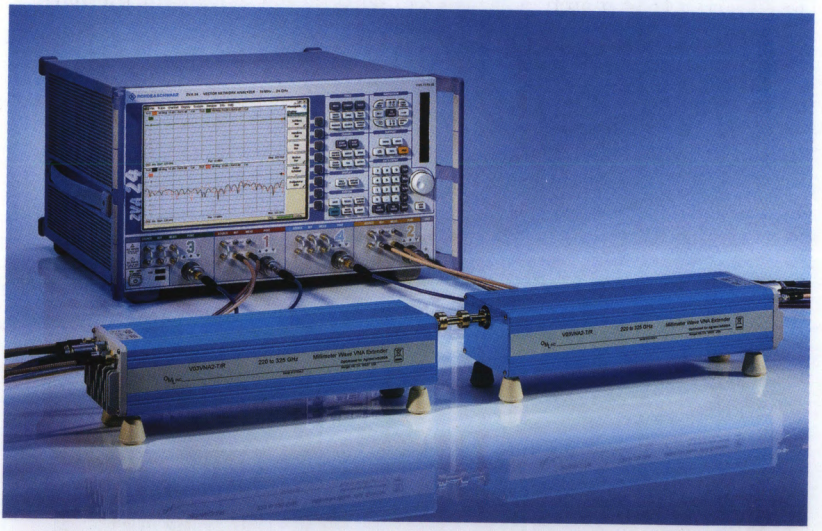


Fig. 8.5.3 Test setup for vector network analysis in the frequency range 220 GHz to 325 GHz.

The WR-03 waveguide cross section used here is highly filigreed with dimensions of less than one millimeter. In Fig. 8.5.4, it can be seen in the center of the waveguide flange. The MIL-F-3922/67B-M03 flange type is extended by two additional alignment pins that can be seen directly above and below the waveguide cross section.



Fig. 8.5.4 MIL-F-3922/67B-M03 waveguide flange with coin for size comparison.

As the measured quantities, we select the transmission and reflection:

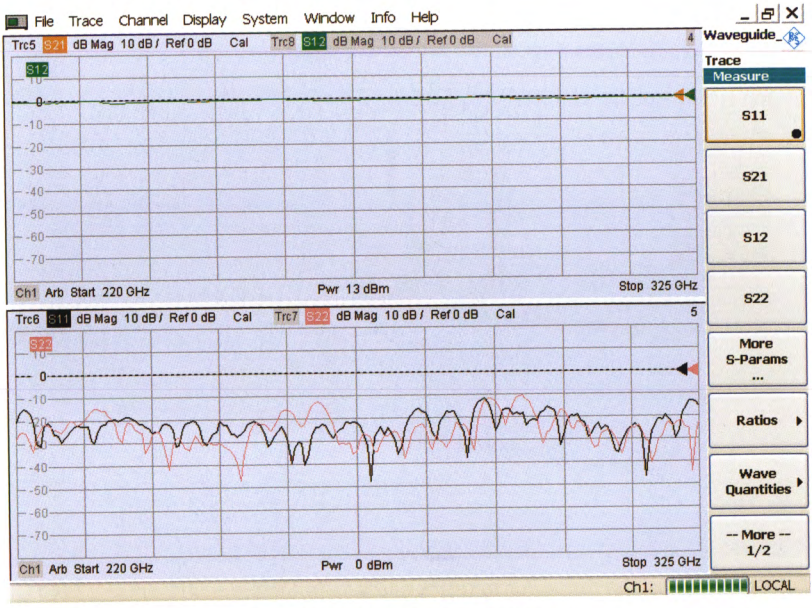


Fig. 8.5.5 Measurement result for the waveguide section.

Measurement example 2: Using the same configuration, we will now investigate a mm-wave amplifier. The amplifier is configured for a cen-

ter frequency of 215 GHz, which is why its transmission decreases significantly above 230 GHz.

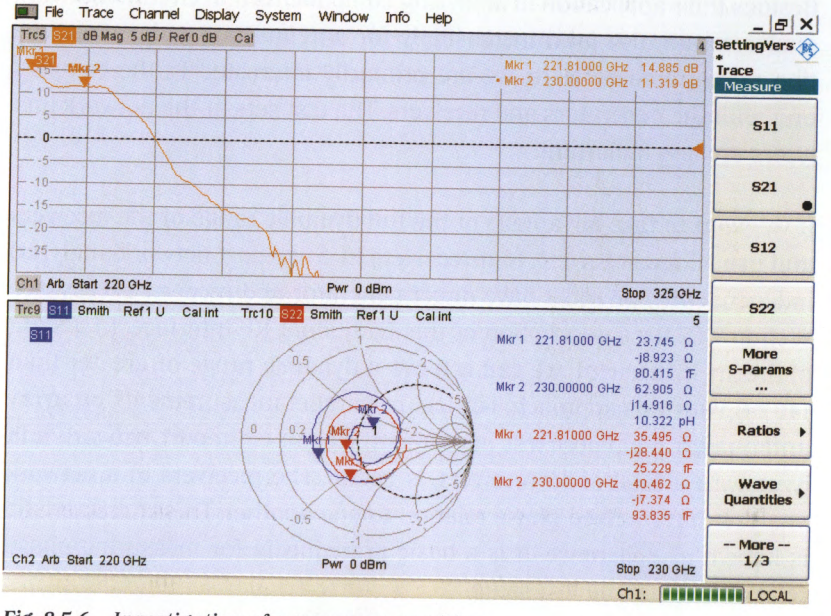


Fig. 8.5.6 Investigation of a mm-wave amplifier.

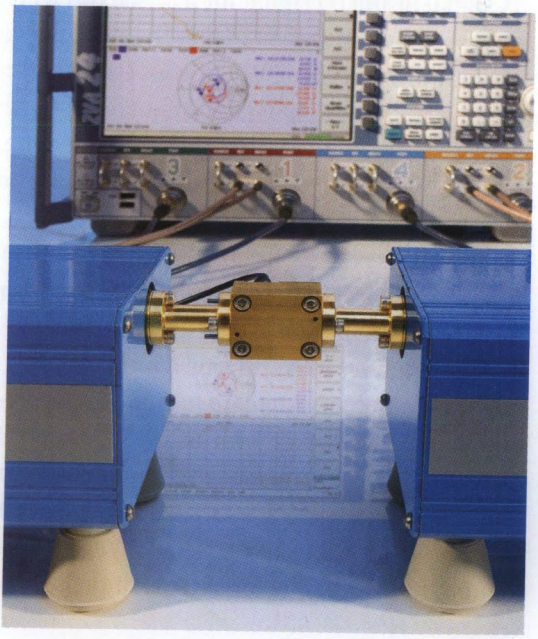


Fig. 8.5.7 mm-wave amplifier (Fraunhofer-Institut IAF) as a DUT on the Oleson V03VNA2-T/R frequency extension modules.

9 Antenna and radar cross section measurements

Besides their application in analyzing components and circuits, network analyzers are also used increasingly for antenna and radar cross section measurements. Here, we are primarily interested in the compact unit with the generators and receivers. The test sets of the network analyzers are less important.

If we want to take advantage of the full dynamic range of the receivers and use all receivers (e.g. four receivers in a two-port network analyzer) independently, we must have direct generator and receiver access (see section 7.1.7 for a description of this additional RF interface. In a state-of-the-art instrument, we can achieve a dynamic range of greater than 140 dB with this approach. For use in parallel measurements on **array antennas**, a multi-port test instrument, e.g. an eight-port network analyzer with a total of 16 receivers, is useful. The receivers in a network analyzer have a fixed phase relation to one another. This characteristic is known as **coherence**. It is a basic prerequisite for measuring phase relations within the test setup.

The network analyzer's generator can be used to generate the stimulus signal or the LO signal for a test setup with external mixing. Most of the test setups described in this chapter involve transmission lines with a length of 20 m and even greater. Using external mixing, it is possible to convert the test signals at frequencies of several GHz down to a frequency of only a few MHz. This simplifies the process of transmitting these signals between the test setup and the analyzer.

The transmission paths discussed here all include the following sections: **Signal feeding - Antenna - Free-space propagation - Antenna - Signal retransmission**. Because of a relatively high overall transmission loss the analyzer must provide a high dynamic range. The DUTs are generally mounted on a **positioner** and moved continuously. Each rotary axis contains a **rotary joint** to pass through the measured signal. The measurement must be made very quickly on the moving DUT. For example, if we are analyzing a DUT in two axes with 0.1° resolution in each axis, we will have to make $3600 \cdot 3600 = 12.96 \cdot 10^6$ individual measurements. The sweep time for each position is therefore a critical parameter.

9.1 Antenna measurements

If we feed an antenna with a sinusoidal signal, it radiates a field that reveals the same sinusoidal behavior when considering the field strength vs. time of any arbitrary point somewhere around the antenna. Compared to the stimulus, the field components at that point have the same frequency but their phase depends on their location. If we chose a fixed phase value we can search for points having the same phase. The result would be a three dimensional surface which we call the **wavefront**. A transmitting antenna radiates **spherical wavefronts** (see Fig. 9.1.1). At the so-called **far-field distance** (which is significantly greater than the free-space wavelength λ_0 or the dimensions of the antenna), however, the radiated waves have approximately **plane wavefronts**. This is known as the antenna's **far-field region**. In **radio transmission** scenarios, the receiver is usually located in the antenna's far-field region.

$$\lambda_0 = c_0 / f \text{ where } c_0 = 2.99792458 \cdot 10^8 \text{ m/s} \qquad \textit{(9.1-1)}$$

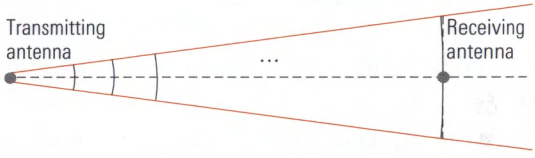


Fig. 9.1.1 Approximately plane wavefronts in the far field of an antenna.

The DUT considered below can be a transmitting antenna or a receiving antenna. Transmitting antennas are usually tested by operating them as receiving antennas. Inversion of the signal direction is permissible since antennas are reciprocal. **Reciprocity** means that their transmission and reception characteristics are identical. When making measurements on antennas, we expose the DUT to waves having plane wavefronts for the following reasons:

- Plane wavefronts generally corresponds to the real-world operating conditions for a receiving antenna.
- With plane wavefronts, the electric field vector **E**, the magnetic field vector **H** and the propagation direction **z** are all perpendicular to one another (see Fig. 9.1.2 next page).

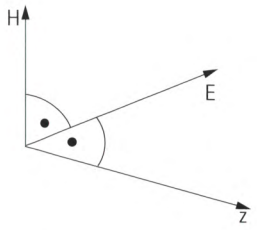


Fig. 9.1.2
Orientation of the field components in the far field.

Due to this geometric relationship and the presence of air as a dielectric[1], the electric field strength $|\mathbf{E}|$ and the magnetic field strength $|\mathbf{H}|$ are related by the **free-space impedance** Z_{F0}.

$$|\mathbf{E}| = Z_{F0} \cdot |\mathbf{H}| \text{ where } Z_{F0} = 120\pi\ \Omega \approx 377\ \Omega. \qquad \textit{(9.1-2)}$$

It is sufficient to just measure (characterize) one of the field components.

- When using a transmitting antenna, we are mostly interested in the far field it generates. Based on the reciprocity mentioned above the characteristics of the transmitting antenna can be measured operating it as a receiving antenna. A stimulus field with plane wavefronts must be applied because it is the reciprocal equivalent for the far field.

Three techniques are commonly used to verify the far-field properties of an antenna. The first approach involves measuring the **near field** of the antenna and computationally determining the far field using a **near-field transformation**. The other two approaches do not involve any conversion of this sort and will be considered in further detail here.

A far-field measurement is guaranteed if we simply choose the distance between the antennas large enough. Depending on the dimensions of the antennas and the wavelength, such measurements might need to be made outdoors. This is known as **open site testing**. Weather conditions (e.g. snow, rain) can influence radio transmission. If influences of this sort are undesired, then the measurements must be performed under specific weather conditions. This can restrict the measurement to some few days of the year where the weather conditions meets. In principle, any signals from other radio services can impair the measurement. Suit-

1) The permittivity of air is approximated by the vacuum permittivity, i.e. $\varepsilon_r = 1$. In addition, lossless wave propagation is assumed.

able control measurements should be performed to detect any potential problem of this sort. The following figure shows a far-field measurement. In this present example the distance is very large. A generator is used to generate the Tx signal. The generator is controlled via a fiber optical link. Suitable converters are employed to connect the fiber optical link to the GPIB interface used in the instruments. The external reference frequency for the network analyzer and generator is obtained from the GPS signal available at both locations.

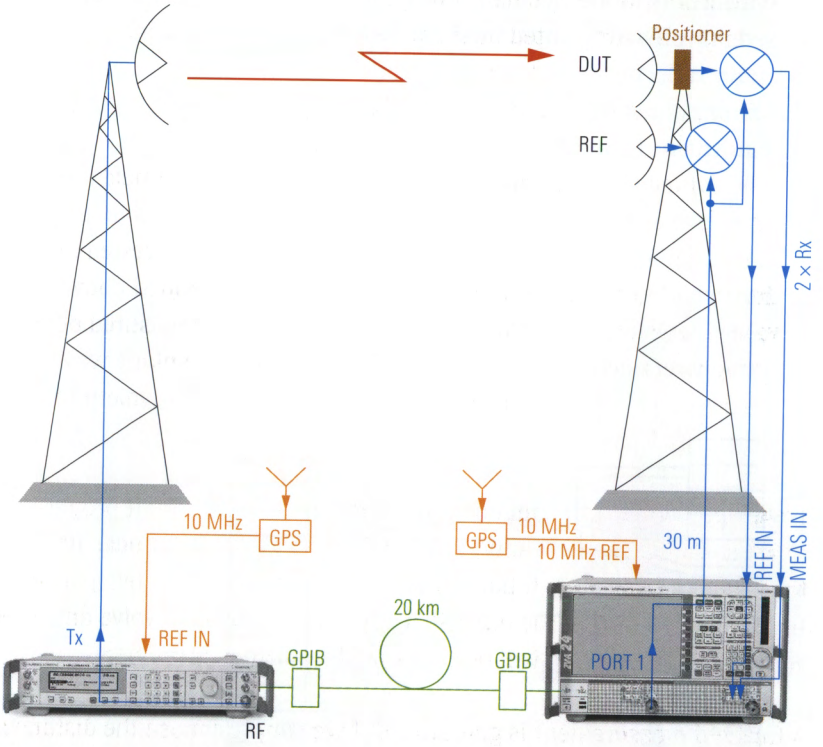


Fig. 9.1.3 Open site test covering a distance of 20 km.

Using external mixers, the Rx signals are converted down to a few MHz so they can be fed to the network analyzer via the 30 m long transmission lines. The antenna designated as "REF" in Fig. 9.1.3 provides the comparison signal for the phase and amplitude at the measurement location.

In **anechoic chambers**, we have constant measurement conditions. Furthermore interference from other transmitters is largely eliminated. The chambers are coated with **pyramidal absorbers** to suppress reflections.

Of course, the spatial dimensions of an anechoic chamber are limited. To generate plane wavefronts in this compact space, we must convert the spherical wavefronts generated by a signal source - **feed** - into plane wavefronts using one or more curved reflectors. The setup used to do this is known as a **compact antenna test range** (CATR). There are three different implementations:

- A **single-reflector compact test range** (SCR) has one reflector for converting the feed into waves that exhibit approximately plane wavefronts in the vicinity of the DUT.
- A **dual-reflector compact test range** has a parabolic cylinder as a **main reflector** and also as a **subreflector** to do the same. Assuming similar dimensions, these test ranges have a significantly greater **equivalent focus length** than single-reflector compact test ranges.
- The **compensated compact test range** (CCR) also has two reflectors. However, it manages to avoid system-inherent **cross-polarization** with its special shaped reflectors (double curved sections of hyperbolic and parabolic functions). This means that, the decoupling between the vertical and horizontal polarization component is largely retained.

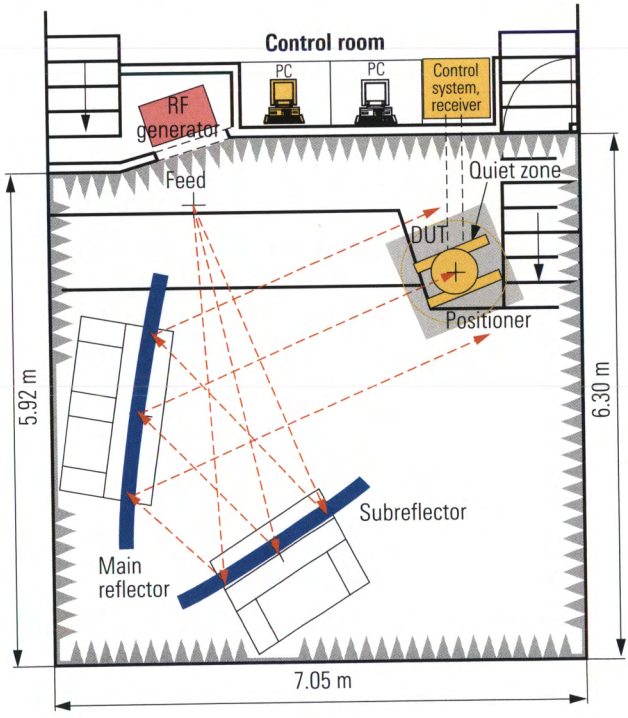


Fig. 9.1.4 Compensated compact test range at the Munich University of Applied Sciences (system type: CCR 20/17 from EADS Astrium).

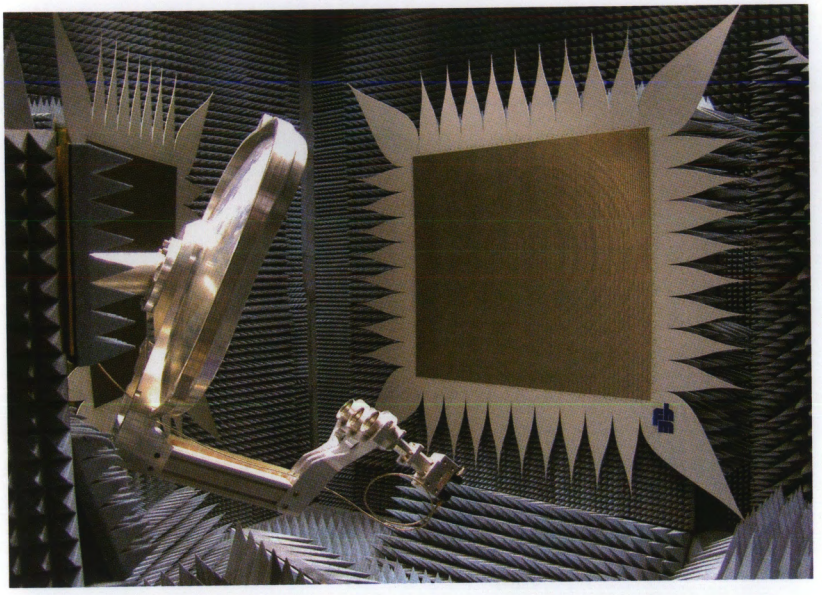


Fig. 9.1.5 *Test of a multi beam / shaped beam reflector antenna in the 12 GHz range (compensated compact test range: Munich University of Applied Sciences; antenna: KATHREIN).*

In Fig. 9.1.4 the **top view** of a compensated compact test range is shown. For this medium-sized test range the dimensions of the test zone (typically called the **quiet zone**) are 1.3 m × 1.3 m. Using a positioner, we can rotate the DUT around multiple axes with high angular resolution. The surface precision of the reflectors is equal to 25 μm. The main reflector and subreflector have dimensions of 2 m × 1.7 m and 1.5 m × 1.5 m, respectively. The system is designed to handle the frequency range from 1.7 GHz to 200 GHz. The waves at the DUT location have a maximum amplitude and phase ripple of 0.5 dB and ±5°, respectively. The equivalent focus length is 37 m [Fa05]. Fig. 9.1.5 shows the test range from the perspective of the positioner.

The measurement can be affected by undesired crosstalk. For example, in the setup shown in Fig. 9.1.4, part of the power delivered by the feed reaches the DUT directly without passing via the reflectors. However, we can distinguish crosstalk mechanisms of this sort based on the delay. This is possible using the pulsed operating mode known as **hardgating**, for example [HF98]. Alternatively, the time gate discussed in section 5.4 can be used. Of course, the measurement must then be made using a frequency sweep, whereas a single measurement frequency is sufficient

with hardgating. Fig. 9.1.6 shows a block diagram for implementation of an compact antenna test range (CATR) using hardgating. For an example of a typical measurement result and the influence of the hardgating, see Fig. 9.2.2.

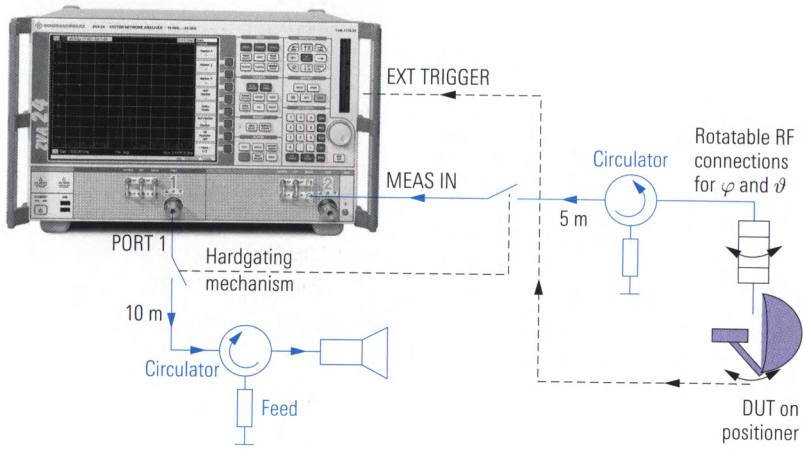


Fig. 9.1.6 Block diagram of an antenna measurement using hardgating.

9.2 Important antenna measurement quantities

Most antenna measurements are intended to record the radiation pattern. They provide us with the radiation diagram, the half power beam width, the side-lobe suppression and (following calibration using a reference antenna) the antenna gain. The definition of these important antenna measurement quantities is discussed below.

The **radiation pattern** of an antenna describes the directional dependence of its radiation. For example, directional antennas have a very pronounced direction of radiation. A radiation diagram is a graphical representation of the radiation pattern in a plane of intersection. The measured quantity is taken in the far-field region of the antenna. It is plotted with respect to the angle ϑ for a fixed angle φ or with respect to the angle φ for a fixed angle ϑ. The following figure illustrates the relationship between the **azimuth angle** φ and the **elevation angle** ϑ.

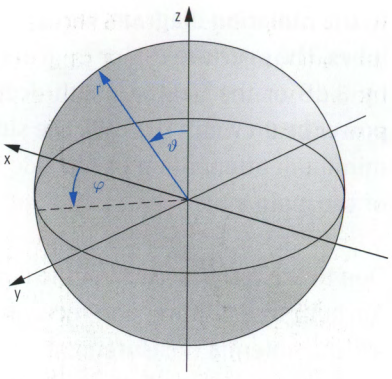


Fig. 9.2.1 Relationship between spherical coordinates and Cartesian coordinates.

In **radiation diagrams**, we can plot either the field strength $E(\vartheta, \varphi)$ or the power per unit solid $P(\vartheta, \varphi)$ on a linear scale or, as is more common, on a dB scale. The graphs are normalized to their maximum linear value. This means that the radiation diagram will always have a maximum value of 1 or 0 dB, respectively. Display of the diagram using polar coordinates is very descriptive. However, we usually prefer to show the angular dependence using a Cartesian coordinate system and a dB scale (see Fig. 9.2.2) since this provides a much more detailed picture of the radiation characteristics.

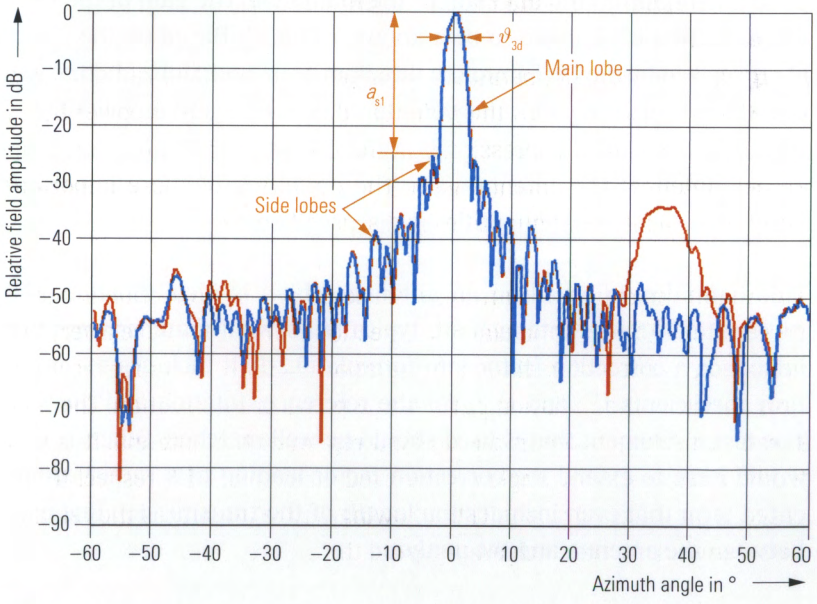


Fig. 9.2.2 Radiation diagram for a 12 GHz offset parabolic antenna with high gain without hardgating (red) and with hardgating (blue).

In the **radiation diagram**, we can distinguish the **main lobe** and the **side lobes**. The aperture ϑ_{3dB} or φ_{3dB} at which the level of the main lobe drops by 3 dB or the field strength decreases to $1/\sqrt{2}$ is known as the **half power beam width** (HPBW). The **side-lobe suppression** a_{sl} designates the minimum attenuation of the side lobes relative to the maximum value of the main lobe.

Due to the reciprocity, all of the definitions discussed above can be used for both transmitting and receiving antennas. This fact forms the basis for the antenna measurements described here. The antenna under test is rotated by the angle ϑ or φ and exposed to a field of plane wavefronts. What we measure is the power $P(\vartheta, \varphi)$ or the electric field strength $E(\vartheta, \varphi)$ received by the antenna as a function of the angle ϑ or φ.

The **antenna gain** G_i represents the ratio of the power P_m received in the direction of the main lobe to the power that an **isotropic radiator** would receive for the same incident field strength. Exposure to plane wavefronts is assumed along with power and polarization matching. Using a measurement with a reference antenna in place of the DUT, we can determine the reference power P_{ref} that is received by the reference antenna from the direction of the main lobe. The gain of the reference antenna is assumed to be a known value G_r. Based on the power P_{DUT} received by the DUT from the direction of its main lobe and the two quantities P_{ref} and G_r, we can calculate the antenna gain as follows:

$$G_i = \frac{P_{DUT}}{P_{ref}} G_r \qquad \textit{(9.2-1)}$$

This definition is based on an antenna, which is ideally matched at its terminals. Since antennas are typically not well matched, we must introduce a correction factor into formula (9.2-1). It includes the reflection coefficients Γ_{ref} and Γ_{DUT} for the reference antenna and the DUT. The test instrument that is used should be well matched. Otherwise we would have to extend the correction factor leading to a rather complicated term that even includes the length of the transmission line used between the antenna and the analyzer.

$$G_i = \frac{P_{DUT}}{P_{ref}} \frac{1 - |\Gamma_{ref}|^2}{1 - |\Gamma_{DUT}|^2} G_r \qquad \textit{(9.2-2)}$$

The antenna gain G_i is a unit-less ratio. It is usually specified by its logarithmic equivalent g_i on a dB scale. To indicate that the values are referenced to an isotropic radiator, we use the dBi unit.

$$g_i = 10 \lg\left(G_i\right) \text{dBi} \qquad \textit{(9.2-3)}$$

9.3 Radar cross section measurements

When assessing the radar visibility of objects (e.g. an aircraft), we also need to expose the DUT (target) to plane wavefronts since it will generally be situated many free-space wavelengths away from the radar station.[1] For a **radar cross section measurement**, we can use a compact antenna test range, for example. The reflectors convert the spherical wavefronts transmitted by the feed into plane wavefronts and concentrate the waves reflected by the DUT (target) back into a focus point. The antenna used for the feed also acts as the receiving antenna. The term we use here is **mono-static radar**. Systems with separate transmitting antenna (Tx antenna) and receiving antenna (Rx antenna) are known as **bistatic radar**. We usually use a scale model as a sample DUT for radar cross section measurements. The measurement frequency is increased by a factor that is inversely proportional to the reduction scale. In mono-static radar, pulsed operation is used to ensure separation of the received signal (which has a level that is lower by, say, 100 dB) from the transmitted signal. Fig. 9.3.1 shows a block diagram for a bistatic radar system. The antennas that are used allow us to distinguish horizontal and vertical polarization.

1) The location technique for material defects described in section 6.3 is an exception, for example.

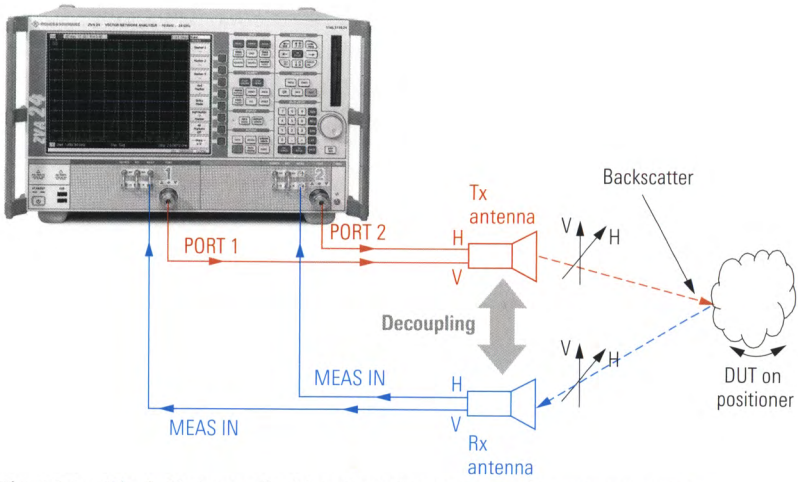


Fig. 9.3.1 Block diagram of a bistatic radar cross section measurement.

The measurement result we obtain is usually the **radar cross section** (RCS) as a function of the rotation of the DUT. The ratio of the received radar power P_{Rx} and transmitted radar power P_{Tx} is given by the **radar range equation**. Assuming bistatic radar where G_{Tx} and G_{Rx} are the gain of the transmitting and receiving antennas, respectively, and d is the distance, λ the wavelength and σ the radar cross section, we can calculate as follows:

$$\frac{P_{\mathrm{Rx}}}{P_{\mathrm{Tx}}} = \sigma \frac{G_{\mathrm{Tx}} G_{\mathrm{Rx}} \lambda^2}{(4\pi)^3 d^4} \qquad \textit{(9.3-1)}$$

The radar cross section σ can be obtained from the measured results P_{Rx} and P_{Tx} using formula (9.3-1). The factor $G_{\mathrm{Tx}} G_{\mathrm{Rx}} \lambda^2 / [(4\pi)^3 d^4]$ involved in the formula is often determined based on a calibration measurement with a reference object having a known radar cross section σ. As the reference object, we can use, for example, a sphere with the radius of two wavelength $r = 2\lambda$. Its radar cross section is approximately $\sigma = \pi r^2$.

The radar cross section σ is often specified using a dB scale. The measured radar cross section in m^2 is referred to an area of $1\ \mathrm{m}^2$. This ratio quantity can then be displayed on a dB scale. The unit dBsm was introduced for this purpose.

$$\Xi = 10 \lg \frac{\sigma}{1\ \mathrm{m}^2}\,\mathrm{dBsm} \qquad \textit{(9.3-2)}$$

The following table shows the radar cross sections for some typical objects.

Object	Radar cross section in m^2	Radar cross section in dBsm
Automobile	100	20
Commercial airline	40	16
Fighter plane	2 to 6	3 to 7.78
Adult person	1	0
Bird	0.01	–20
Insect	0.00001	–50
Advanced stealth fighter	0.000001	–60

Table 9.3.1 *Typical radar cross sections, taken from [1EZ52].*

To increase the resolution when observing moving objects, **inverse synthetic aperture radar** (ISAR) is commonly used. By superimposing the radar data from several images measured during the movement, it is possible to significantly increase the detailed resolution. As part of the example measurement shown in Fig. 9.3.2, the rotation of the DUT is exploited in this manner. The figure shows the radar cross section of an aircraft as an ISAR image. These measurements were made by the company March Microwave Systems on a scaled model of a Boeing 747. The company's Advanced Antenna and RCS Measurement Software (ARCS) was used to control the measurement and evaluate the results.

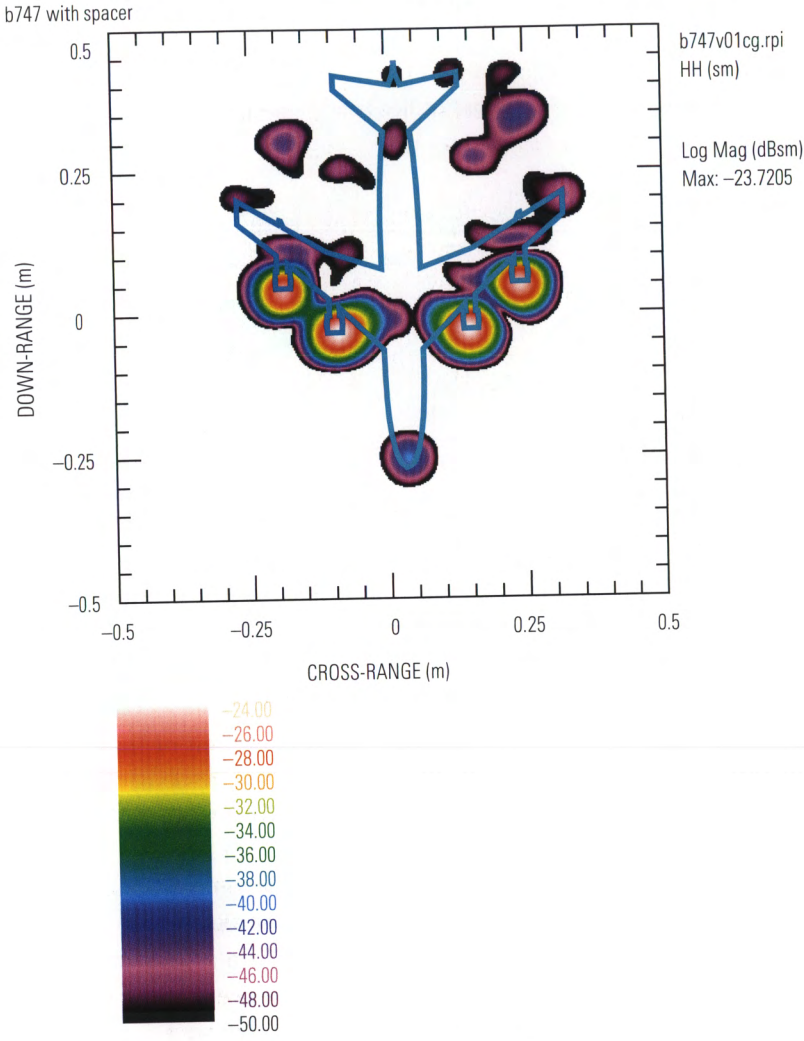


Fig. 9.3.2 ISAR image of the model of a Boeing 747.

10 Conclusion and acknowledgement

Vector network analysis provides an impressive example of how digital data and signal processing techniques have impacted the different features of test instruments over the past 50 years.

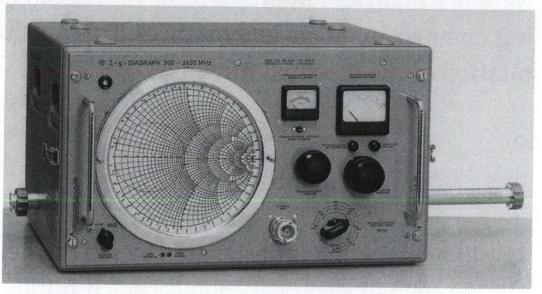


Fig. 10.1.1
The famous Z-g-Diagraph, a vector network analyzer family from Rohde&Schwarz that was manufactured from 1950 to 1977.

In conjunction with technical advances in the field of microwave components, these instruments have evolved to exceed their original limitations by far. Careful trade-off among the analog hardware, digital signal processing and software is very important when designing such an instrument. The proportion of analog hardware has continued to decrease. This has made it possible to develop sophisticated instruments such as our 8-port network analyzer. Besides the number of ports and increasingly compact design of these instruments, progress can also be witnessed in parameters such as the maximum measurement frequency, dynamic range and sweep time.

I personally find this area fascinating due to its location at the crossroads where the descriptive science of physics meets the field of engineering, which is focused on development of new instruments and techniques.

I am very thankful for the information and advice I have received from my circle of colleagues. I prefer not to name anyone individually due to the many people who have helped me in some significant way. Of course, I would like to thank the heads of our lab and department. Since my very first idea they have been continuously interested in this book. I have also had help from outside the company. I would like to thank Prof. Dr-Ing. Fasold (Munich University of Applied Sciences) who contributed figures and advice for chapter 9. I would also like to thank Dr Schlechtweg (Fraunhofer-Institut IAF) for providing the DUTs used in section 8.5. Special thanks for double checking the wording of the English edition to Dr Jetter (Germany) and Mr Lunn (UK).

Michael Hiebel, December 2006

A Appendix

A1 Mathematical principles

Network analysis requires a basic understanding of calculations involving complex numbers and matrices. This appendix is intended as a compilation of the essential formulas.

A1.1 Complex numbers

The imaginary unit is designated by i in the field of mathematics and by j in the field of electrical engineering. It fulfills the equations:

$$j = \sqrt{-1}; \quad j^2 = -1; \quad j^3 = -j; \quad j^4 = 1; \text{ etc.} \qquad \textit{(A1.1-1/4)}$$

The complex number $z = x + jy$ represents an ordered pair of real numbers x and y. The mathematical operations addition, multiplication etc. follow well-defined rules (see following table A1.1.1). Complex numbers are fully determined by their real and imaging part:

Real part: $\mathrm{Re}\{z\} = \mathrm{Re}\{x + jy\} = x$ *(A1.1-5)*

Imaginary part: $\mathrm{Im}\{z\} = \mathrm{Im}\{x + jy\} = y$ *(A1.1-6)*

Any complex number can be represented as a point in a rectangular coordinate system with an abscissa x and an ordinate y. The geometric representation of all complex numbers is known as the complex plane. We can use the following formulas to convert complex numbers between polar and Cartesian representation. The magnitude and angle of a complex number z are notated as $|z|$ and $\arg(z)$.

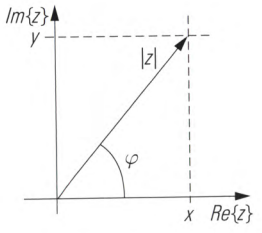


Fig. A1.1.1
Complex plane

Polar → Cartesian conversion:

$$x = |z|\cos\varphi \qquad \textit{(A1.1-7)}$$

$$y = |z|\sin\varphi \qquad \textit{(A1.1-8)}$$

Cartesian → polar conversion:

$$|z| = \sqrt{x^2 + y^2} \qquad \textit{(A1.1-9)}$$

$$\varphi = \arg(z) = \begin{cases} \arccos\dfrac{x}{|z|} & \text{for } y \geq 0 \\ -\arccos\dfrac{x}{|z|} & \text{for } y < 0 \\ \text{undefined} & \text{for } |z| = 0 \end{cases} \qquad \textit{(A1.1-10)}$$

Polar representation of complex numbers leads to **Euler's formula for complex numbers.**

$$z = |z|\cdot[\cos(\varphi) + j\sin(\varphi)] = |z|\cdot e^{j\varphi}.$$

As can be seen in the following table, either Cartesian or polar representation is preferable depending on the particular arithmetic operation to be performed.

Cartesian representation	Polar representation
Addition $z_1 + z_2 = (x_1 + x_2) + j(y_1 + y_2)$	–/–
Subtraction $z_1 - z_2 = (x_1 - x_2) + j(y_1 - y_2)$	–/–
Multiplication $z_1 \cdot z_2 = (x_1x_2 - y_1y_2) + j(x_1y_2 + x_2y_1)$	$z_1z_2 = \lvert z_1\rvert e^{j\varphi_1}\lvert z_2\rvert e^{j\varphi_2} = \lvert z_1\rvert\lvert z_2\rvert e^{j(\varphi_1 + \varphi_2)}$
Division $\dfrac{z_1}{z_2} = \dfrac{x_1 + jy_1}{x_2 + jy_2} = \dfrac{(x_1 + jy_1)(x_2 - jy_2)}{(x_2 + jy_2)(x_2 - jy_2)} = \dfrac{z_1z_2^*}{\lvert z_2\rvert^2}$	$\dfrac{z_1}{z_2} = \dfrac{\lvert z_1\rvert e^{j\varphi_1}}{\lvert z_2\rvert e^{j\varphi_2}} = \dfrac{\lvert z_1\rvert}{\lvert z_2\rvert} e^{j(\varphi_1 - \varphi_2)}$
Power –/–	$z^n = (\lvert z\rvert e^{j\varphi})^n = \lvert z\rvert^n e^{jn\varphi}$
Root –/–	$\sqrt[n]{z} = \sqrt[n]{\lvert z\rvert e^{j\varphi}} = \sqrt[n]{\lvert z\rvert}\cdot e^{j\frac{\varphi + k2\pi}{n}}$ where $k = \{0, \ldots, n-1\}$

Table A1.1.1 Basic operations on complex numbers.

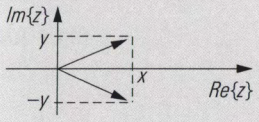
Definition	Calculations
Magnitude $\lvert z\rvert = \sqrt{x^2+y^2} = \sqrt{z^* z}$	$\lvert\lvert z_1\rvert - \lvert z_2\rvert\rvert \le \lvert z_1 - z_2\rvert$; $\lvert z_1 + z_2\rvert \le \lvert z_1\rvert + \lvert z_2\rvert$; $\lvert z_1 z_2\rvert = \lvert z_1\rvert\lvert z_2\rvert$; $\lvert z_1/z_2\rvert = \lvert z_1\rvert/\lvert z_2\rvert$;
Argument See formula (A1.1-10)	$\arg(z_1 z_2) = \arg(z_1) + \arg(z_2)$; $\arg(z_1/z_2) = \arg(z_1) - \arg(z_2)$; $\arg(z^*) = -\arg(z)$; $\arg(z^n) = n \cdot \arg(z)$; $\arg(\sqrt[n]{z}) = \dfrac{2k\pi + \arg(z)}{n}$
Complex conjugate $z^* = (x + jy)^* = x - jy$  	$(z^*)^* = z$ $(z_1 + z_2)^* = z_1^* + z_2^*$; $(z_1 - z_2)^* = z_1^* - z_2^*$ $(z_1 \cdot z_2)^* = z_1^* \cdot z_2^*$; $(z_1/z_2)^* = z_1^*/z_2^*$ $z + z^* = 2x$; $z - z^* = 2jy$ $\lvert z^*\rvert = z$; $z \cdot z^* = \lvert z\rvert^2$; $\arg(z^*) = -\arg(z)$ In technical literature, $\bar{z}$ is also commonly used instead of z^*!
Real part $\mathrm{Re}\{z\} = \mathrm{Re}\{x + jy\} = x$ Alternatively: $\mathrm{Re}\{z\} = \frac{1}{2} \cdot (z + z^*)$	$\mathrm{Re}\{z^*\} = x$ $\mathrm{Re}\{z_1 + z_2\} = x_1 + x_2$; $\mathrm{Re}\{z_1 - z_2\} = x_1 - x_2$ $\mathrm{Re}\{z_1 \cdot z_2\} = x_1 x_2 - y_1 y_2$; $\mathrm{Re}\left\{\dfrac{z_1}{z_2}\right\} = \dfrac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2}$ $\mathrm{Re}\{z^n\} = \lvert z\rvert^n \cos(n\varphi)$ where $\varphi = \arg(z)$ $\mathrm{Re}\{\sqrt[n]{z}\} = \sqrt[n]{\lvert z\rvert}\cos\left(\dfrac{\varphi + 2k\pi}{n}\right)$ where $\varphi = \arg(z)$
Imaginary part $\mathrm{Im}\{z\} = \mathrm{Im}\{x + jy\} = y$ Alternatively: $\mathrm{Im}\{z\} = \dfrac{1}{2j}(z - z^*)$	$\mathrm{Im}\{z^*\} = -y$ $\mathrm{Im}\{z_1 + z_2\} = y_1 + y_2$; $\mathrm{Im}\{z_1 - z_2\} = y_1 - y_2$ $\mathrm{Im}\{z_1 \cdot z_2\} = x_1 y_2 + y_1 x_2$; $\mathrm{Im}\left\{\dfrac{z_1}{z_2}\right\} = \dfrac{x_2 y_1 - x_1 y_2}{x_2^2 + y_2^2}$ $\mathrm{Im}\{z^n\} = \lvert z\rvert^n \sin(n\varphi)$ where $\varphi = \arg(z)$ $\mathrm{Im}\{\sqrt[n]{z}\} = \sqrt[n]{\lvert z\rvert}\sin\left(\dfrac{\varphi + 2k\pi}{n}\right)$ where $\varphi = \arg(z)$

Table A1.1.2 Additional operations on complex numbers.

A1.2 Matrix operations

Definition of matrices and vectors

Matrices are just a rectangular scheme of numbers. These numbers are known as the "elements" of the matrix. A matrix can even have complex elements, meaning each element is a complex number. A typical example for this is the scattering parameter matrix introduced in section 1.2. Matrices with only a single column are known as "vectors". Matrices and vectors are written in bold to distinguish them from scalar numbers. Capital letters are commonly used for matrices and lower-case letters are commonly used for vectors.

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = (a_{ik}) \qquad \textit{(A1.2.1)}$$

The dimension $\mathrm{Dim}(\mathbf{A}) = m \times n$ characterizes the number of rows m and number of columns n of matrix **A**. The matrix element in row i and column k is notated as a_{ik}.

Special matrices

It is advantageous to define the following two matrices.

Zero matrix:
$$\mathbf{0} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \qquad \textit{(A1.2.2)}$$

Unit matrix:
$$\mathbf{1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \textit{(A1.2.3)}$$

Operations on matrices

The mathematical operations like addition, multiplication etc. follow well-defined rules. These rules (axioms) define how to calculate each element of the resulting matrix. All basic arithmetic operations ($+$, $-$, $\cdot$, and $\div$) on matrices are defined this way. Note that division is done by multiplication with the inverse matrix. To undergo a mathematical operation certain preconditions must be fulfilled by the matrices involved. The following tables list selected axioms and elementary rules for operations on matrices.

Operation	Precondition	Example, rules for calculations
$\mathbf{A} = \mathbf{B}$	$\mathrm{Dim}(\mathbf{A}) = \mathrm{Dim}(\mathbf{B})$	Example for $\mathrm{Dim}(\mathbf{A}) = 2 \times 2$: $\mathbf{A} = \mathbf{B} \Leftrightarrow a_{11} = b_{11} \wedge a_{12} = b_{12}$ $\wedge\ a_{21} = b_{21} \wedge a_{22} = b_{22}$ In words, the corresponding matrix elements are equal.
$\mathbf{A} + \mathbf{B}$	$\mathrm{Dim}(\mathbf{A}) = \mathrm{Dim}(\mathbf{B})$ Result has same dimension as addends	Example for $\mathrm{Dim}(\mathbf{A}) = 2 \times 2$: $\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}$ In words, the corresponding matrix elements are added. We have: $\mathbf{A} + \mathbf{0} = \mathbf{A}$ $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ $(\mathbf{A} + \mathbf{B}) + \mathbf{D} = \mathbf{A} + (\mathbf{B} + \mathbf{D})$ $(\mathbf{A} + \mathbf{B}) + \mathbf{D} = \mathbf{A} + \mathbf{B} + \mathbf{D}$
$\mathbf{A} - \mathbf{B}$	$\mathrm{Dim}(\mathbf{A}) = \mathrm{Dim}(\mathbf{B})$	Example and computation rules are analogous to $\mathbf{A} + \mathbf{B}$
$\eta\mathbf{A}$	$\mathrm{Dim}(\eta) = 1$ In words η is a scalar quantity and the result has same dimension as matrix $\mathbf{A}$	Example for $\mathrm{Dim}(\mathbf{A}) = 2 \times 2$: $\eta A = \begin{pmatrix} \eta a_{11} & \eta a_{12} \\ \eta a_{21} & \eta a_{22} \end{pmatrix}$ In words, each element is multiplied by the scalar factor. We have: $1\mathbf{A} = \mathbf{A}$ $(\eta\mathbf{A}) = \mathbf{A} \cdot \eta$ $(\eta\mathbf{A}) \cdot \mathbf{B} = \mathbf{A} \cdot (\eta\mathbf{B}) = \eta(\mathbf{A} \cdot \mathbf{B})$ $(\eta + \upsilon)\mathbf{A} = \eta\mathbf{A} + \upsilon\mathbf{A}$ where η, υ scalar
$\mathbf{A} \cdot \mathbf{B}$	$\mathrm{Dim}(\mathbf{A}) = \mathrm{m} \times \mathrm{p}$ $\mathrm{Dim}(\mathbf{B}) = \mathrm{p} \times \mathrm{n}$ Result: $\mathrm{Dim}(\mathbf{C}) = \mathrm{m} \times \mathrm{n}$	Example for $\mathrm{Dim}(\mathbf{A}) = 2 \times 2$: $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} =$ $\begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}$ The following computation rules generally apply: $1\mathbf{A} = \mathbf{A}$ $\mathbf{AB} \neq \mathbf{BA}$ $(\mathbf{A} \cdot \mathbf{B}) \cdot \mathbf{D} = \mathbf{A} \cdot (\mathbf{B} \cdot \mathbf{D}) = \mathbf{A} \cdot \mathbf{B} \cdot \mathbf{D}$

Table A1.2.1 *Basic operations on matrices.*

Name, *Definition*, Computation

Determinant

Def.: The determinant $\det(\mathbf{A})$ *of a quadratic matrix* $(\mathrm{Dim}(\mathbf{A}) = n \times n)$ *is a scalar number. It corresponds to the volume of the body that is spanned by the column vectors of* $\mathbf{A}$ *in n-dimensional space.*

$$n = 2: \quad \det(\mathbf{A}) = \det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{21}a_{12}$$

$n > 2$: Row or column expansion is necessary; see, e.g. [Di87], p. 497 ff.

Inversion

Definition: $\mathbf{A}\mathbf{A}^{-1} = 1$ *Existence:* $\det(\mathbf{A}) \neq 0$ and $\mathrm{Dim}(\mathbf{A}) = n \times n$

$$n = 2: \quad \mathbf{A}^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = \frac{1}{\det(\mathbf{A})}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

$n > 2$: Gauss-Jordan algorithm must be used; see, e.g. [BS91], p. 739 ff.

Transposition

Def.: $\mathbf{A}^{\mathrm{T}} = (a_{ik})^{\mathrm{T}} = (a_{ki})$ In short: "Interchange rows and columns"

Conjugation

Def.: $\mathbf{A}^{*} = (a_{ik})^{*} = a_{ik}^{*}$ In short: "All of the elements are conjugated individually"

Note: In technical literature, $\overline{\mathbf{A}}$ is sometimes used instead of $\mathbf{A}^{*}$!

Adjunction (or transjunction)

Def.: $\mathbf{A}^{\mathrm{H}} = (\mathbf{A}^{*})^{\mathrm{T}}$ In short: "Conjugation and transposition of a matrix"

Note: In technical literature, $\mathbf{A}^{\mathrm{Adj}}$ is also used instead of $\mathbf{A}^{\mathrm{H}}$.

Diagonal

Def.: $\mathrm{diag}(\mathbf{A}) = [a_{11}, a_{22}, \ldots, a_{nn}]^{\mathrm{T}}$ In short: "Vector of diagonal elements"

Table A1.2.2 *Selected arithmetic operations on matrices.*

A2 Important symbols and quantities

Symbol	Meaning	Unit	Definition
$\mathbf{a}$	Vector of the incident waves[1)]	$\sqrt{\mathrm{W}}$	(1.2-7)
a	Phasor of the incident wave[1)]	$\sqrt{\mathrm{W}}$	(1.2-1)
$a(t)$	Time-domain representation of the incident wave	$\sqrt{\mathrm{W}}$	(5.1-7)
d_{CMR}	Common-mode rejection ratio	dB	(4.12-3)
a_{IM2}	Second order intermodulation suppression	dBc	(7.5-2)
a_{IM3}	Third order intermodulation suppression	dBc	(7.5-3)
a_{k2}	Second order harmonic suppression	dBc	Fig. 7.4.1
a_{k3}	Third order harmonic suppression	dBc	Fig. 7.4.1
$\mathrm{a_p}$	Power gain figure	dB	(3.6-2)
$\arg(\ )$	Argument (angle) of a complex number	rad or °	(A1.1-10)
$a_{\mathrm{s,ii}}$	Reflection	dB	(4.3-1)
$a_{\mathrm{s,ik}}$	Transmission (where $i \neq k$)	dB	(4.4-1)
a_{sl}	Side lobe suppression	dB	Fig. 9.2.2
a_{Γ}	Reflection	dB	(3.6-5)
$\mathbf{b}$	Vector of the reflected waves[1)]	$\sqrt{\mathrm{W}}$	(1.2-7)
b	Phasor of the reflected wave[1)]	$\sqrt{\mathrm{W}}$	(1.2-1)
$b(t)$	Time-domain representation of the reflected wave	$\sqrt{\mathrm{W}}$	(5.1-8)
B_{IBRR}	Relative impulse width	1	(5.2-17)
c_0	Speed of light in a vacuum	$\mathrm{m/s}$	(9.1-1)
c_{T}	T-check parameter	1	(3.5-3)
D	Directivity	$\sqrt{\mathrm{W}}/\sqrt{\mathrm{W}}$	(2.2-6)
d_{Conv}	Conversion loss	dB	(8.1-14)
$\det(\ )$	Determinant of a matrix	–/–	Section A1.2
$\mathrm{DIM}(\ )$	Dimension of a matrix	1	Section A1.2

1) All phasors and complex vectors of the wave quantities a and b are scaled in such a manner that their magnitude or the magnitude of their elements is an RMS value.

Symbol	Meaning	Unit	Definition
$d_{\mathrm{LO-IF}}$	LO-IF isolation	dB	(8.1-11)
$d_{\mathrm{LO-RF}}$	LO-RF isolation	dB	(8.1-13)
d_{p}	Power attenuation figure	dB	(3.6-2)
$d_{\mathrm{RF-IF}}$	RF-IF isolation	dB	(8.1-12)
$dy(x)/dx$	Derivative of the function $y(x)$ with respect to x	–/–	–/–
d_{Γ}	Return loss	dB	Section 3.6.2
e_{ik}	Terms of the error model	$\sqrt{\mathrm{W}}/\sqrt{\mathrm{W}}$	Section 3.4
$e_{\mathrm{ik,eff}}$	Effective system data	1	Section 3.5.2
f	Frequency	Hz = 1/s	–/–
F	Periodicity of the spectrum	Hz	Fig. 5.3.2 Fig. 5.3.4
f_{IF}	Intermediate frequency	Hz	Section 8.1.1
f_{LO}	Local oscillator frequency	Hz	Section 8.1.1
f_{RF}	Radio frequency	Hz	Section 8.1.1
f_{c}	Waveguide cutoff frequency	Hz	Section 3.3.2
f_{d}	Aperture for a group delay measurement	Hz	(4.5-2)
G_{i}	Antenna gain factor, referred to an isotropic radiator	1	(9.2-2)
g_{i}	Antenna gain, referred to an isotropic radiator	dBi	(9.2-3)
$i(t)$	Time-domain representation of the current	A	(5.1-9)
$\mathrm{Im}\{\ \}$	Imaginary part	–/–	(A1.1-6)
j	Imaginary unit	1	(A1.1-1/4)
k	Stability factor	1	(4.7-3)
$\lg(x)$	Base 10 logarithm of x (decimal logarithm)	–/–	–/–
$\mathrm{ld}(x)$	Base 2 logarithm of x (logarithm dualis)	–/–	–/–
$\ln(x)$	Base e logarithm of x (natural logarithm)	–/–	–/–
L_{IM2}	Level of the second order intermodulation product	dBm	Fig. 7.5.1

Symbol	Meaning	Unit	Definition
L_{IM3}	Level of the third order intermodulation product	dBm	Fig. 7.5.1
L_{in}	Input level	dBm	Section 7.4 ff
$L_{\mathrm{IP2\,IN}}$	Input level at the second order intercept point IP2	dBm	Fig. 7.5.2
$L_{\mathrm{IP2\,OUT}}$	Output level at the second order intercept point IP2	dBm	Fig. 7.5.2
$L_{\mathrm{IP3\,IN}}$	Input level at the third order intercept point IP3	dBm	Fig. 7.5.2
$L_{\mathrm{IP3\,OUT}}$	Output level at the third order intercept point IP3	dBm	Fig. 7.5.2
$L_{\mathrm{IPk2\,IN}}$	Input level at the second order harmonic intercept point IPk2	dBm	Fig. 7.4.2
$L_{\mathrm{IPk2\,OUT}}$	Output level at the second order harmonic intercept point IPk2	dBm	Fig. 7.4.2
$L_{\mathrm{IPk3\,IN}}$	Input level at the third order harmonic intercept point IPk3	dBm	Fig. 7.4.2
$L_{\mathrm{IPk3\,OUT}}$	Output level at the third order harmonic intercept point IPk3	dBm	Fig. 7.4.2
L_{k2}	Level of the second harmonic	dBm	Fig. 7.4.1
L_{k3}	Level of the third harmonic	dBm	Fig. 7.4.1
L_{out}	Output level (of the first harmonic)	dBm	Section 7.4 ff
L_{p}	Level (power level)	dBm	(3.6-3)
M	Measured value (on linear scale)	$\sqrt{\mathrm{W}}/\sqrt{\mathrm{W}}$	Chapters 2, 3
M	Number of points after complex conjugated complement for $f < 0$	1	(5.2-5)
N	Number of points	1	Chapter 5
N	Number of ports	1	Fig. 2.1.1
P_{DC}	DC power	W	Section 4.15
P_{RF}	Effective RF power	W	(4.15-2)
R	Reflection tracking	$\sqrt{\mathrm{W}}/\sqrt{\mathrm{W}}$	(2.2-4)
$\mathrm{Re}\{\ \}$	Real part	–/–	(A1.1-5)
$\mathbf{S}$	Scattering parameter matrix (S-matrix for short)	$\sqrt{\mathrm{W}}/\sqrt{\mathrm{W}}$	(1.2-7)

Symbol	Meaning	Unit	Definition
S	Test port match	$\sqrt{\text{W}}/\sqrt{\text{W}}$	Fig. 2.2.6
$\mathbf{S}_{\text{G}}$	S-matrix for error two port G	$\sqrt{\text{W}}/\sqrt{\text{W}}$	(3.4-11)
$\mathbf{S}_{\text{H}}$	S-matrix for error two port H	$\sqrt{\text{W}}/\sqrt{\text{W}}$	(3.4-12)
$\text{si}(t)$	si function	1	(5.2-16)
$SWR(f)$	Standing wave ratio (or voltage standing wave ratio)	V/V	(4.3-2)
$SWR(t)$	Time-domain display of standing wave ratio	V/V	(5.3-7)
t	Time	s	-/-
t_{d}	Arbitrary time step size	s	Section 5.2.2
T	Period of time-domain representation	s	(5.2-14) Fig. 5.2.7
$u(t)$	Time-domain representation of voltage	V	(5.1-9)
$U_{\text{S}}(t)$	Source voltage	V	(5.1-9)
$v(t)$	Time gate (representation in time-domain)	1	Section 5.4
$W(f)$	Window function	1	Section 5.2.2
z	Normalized impedance	1	Section 1.2
Z_0	Reference impedance	Ω	Section 1.2
Z_{F0}	Free-space impedance	Ω	(9.1-2)
Z_{C}	Characteristic impedance of a transmission line (e. g. an air line)	Ω	(3.3-6)
Z_{S}	Source impedance	Ω	Section 7.8
Z_{Ce}	Even-mode characteristic impedance	Ω	Section 2.3.2
Z_{Co}	Odd-mode characteristic impedance	Ω	Section 2.3.2
Z_{Cc}	Common-mode characteristic impedance	Ω	Section 4.11
Z_{Cd}	Differential-mode characteristic impedance	Ω	Section 4.11
Δf	Frequency step size	Hz	Fig. 5.3.2 Fig. 5.3.4
Δt	Time resolution	s	(5.3-2) (5.3-4)
Γ	Phasor of the reflection coefficient (Reflection coefficient for short)	$\sqrt{\text{W}}/\sqrt{\text{W}}$	(1.2-1)

Symbol	Meaning	Unit	Definition
Γ_{DUT}	Reflection coefficient of the DUT	$\sqrt{\mathrm{W}}/\sqrt{\mathrm{W}}$	(2.2-1)
$\Gamma_{\mathrm{H}}(f)$	Frequency response of the reflection coefficient	$\sqrt{\mathrm{W}}/\sqrt{\mathrm{W}}$	(5.1-15)
$\Gamma_{\mathrm{h}}(t)$	Impulse response of the reflection coefficient For a continuous time axis: Unit 1/s, For a discrete time axis: Unit 1	1/s or 1	(5.1-5)
$\Gamma_{\Theta}(t)$	Step response of the reflection coefficient	$\sqrt{\mathrm{W}}/\sqrt{\mathrm{W}}$	(5.1-6)
$\delta(t)$	Dirac impulse	1/s	(5.1-1)
ε_{r}	Permittivity	1	–/–
η_{PAE}	Power added efficiency	W/W	(4.15-3)
κ	Aperture stepwidth	1	Fig. 4.5.2
λ_0	Free-space wavelength	m	(9.1-1)
λ_{c}	Cutoff wavelength of a waveguide	m	(3.3-7)
λ_{W}	Waveguide wavelength	m	(3.3-9)
μ_1, μ_2	Stability factors	1	(4.7-1) (4.7-2)
μ_{r}	Relative permeability	1	–/–
σ	Radar cross section	m^2	(9.3-1)
$\sigma(t)$	Unit step	1	Fig. 5.1.2
$\Theta(t)$	Step response	–/–	Fig. 5.1.2
$\mathrm{h}(t)$	Impulse response	–/–	Fig. 5.1.1
τ_{G}	Group delay	1/Hz = s	(4.5-1)
τ_{P}	Phase delay	1/Hz = s	(4.6-1)
$\overline{x}$	Linear mean value of the quantity x	–/–	(3.5-11)
Ξ	Radar cross section	dBsm	(9.3-2)
$\vartheta_{3\mathrm{dB}}$, $\varphi_{3\mathrm{dB}}$	Half power beam width	°	Fig. 9.2.2
$\partial y(\mathbf{x})/\partial x_1$	Partial derivative of the function $y(\mathbf{x})$ with respect to x_1	–/–	–/–
$\int y(x)\,dx$	Integral of $y(\mathbf{x})$ with respect to x	–/–	(5.2-7)
$\lvert\ \rvert$	Magnitude of a vector or a complex number	–/–	(A1.1-9)

Table A2 *Important symbols and quantities.*

A3 Circuit symbols used

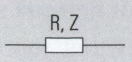
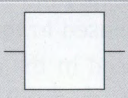
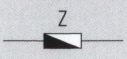
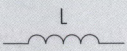
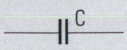
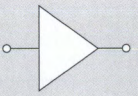
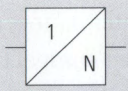
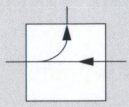
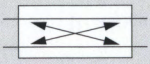
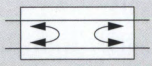
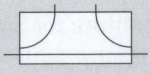
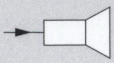
One-port devices		Two-, three- and four-port devices	
	Resistor (or complex impedance)		Two-port device (general)
	Nonlinear impedance		Attenuator pad
	Semiconductor diode		Lowpass filter
	Inductance		Bandpass filter
	Capacitor		Highpass filter
	Safety fuse		Amplifier
	DC constant voltage source		Frequency multiplier
	DC constant current source		Directional element
	Oscillator		Directional coupler (forward coupler)
	Crystal oscillator		Directional coupler (reverse coupler)
	Power meter, receiver (possibly narrowband)		Dual directional coupler
	Antenna		Three-port circulator
	Horn feed (antenna shape)		Mixer

Table A3 Circuit symbols used.

A4 Acronyms and abbreviations

The following list covers the acronyms and abbreviations used in this book as well as those used in the more general context of network analyzers. The → icon used in the column "section" denotes that you will find a definition of the term within the named section. Otherwise, the term is simply related to the context of the section.

Acronym, abbreviation	Meaning	Section
A	Attenuator standard	→ 3.3.1
AM	Amplitude modulation	–/–
ASCII	American standard code for information interchange	2.9.3
AUT	Antenna under test	9.1
BW	Bandwidth	4.12
CATR	Compact antenna test range	→ 9.1
CCR	Compensated compact test range	→ 9.1
CMRR	Common-mode rejection ratio	→ 4.12
COAX	COAXial	3.3.1
CPW	Coplanar waveguide	→ 3.3.4
CW	Continuous wave	→ 2.8.2
CZT	Chirp z-transform	→ 5.2.4
DC	Direct current	→ 4.15
DFT	Discrete Fourier transform	5.2
DIN	Deutsches Institut für Normung e.V. (German Institute for Standardization)	–/–
DKD	Deutscher Kalibrierdienst (German Calibration Service)	→ 3.6.4
DLL	Dynamic link library	2.9.3
DSP	Digital signal processing	→ 2.6
DTF	Distance to fault	6.1
DUT	Device under test	→ 2.1
EA	European co-operation for Accreditation	3.6.4
EHF	Extremely high frequencies (f = 30 GHz to 300 GHz)	–/–
EIRP	Effective isotropic radiated power	9.2

ELF	Extremely low frequencies ($f \leq 300$ Hz)	–/–
EMI	Electromagnetic interference	–/–
ERP	Effective radiated power	9.2
FCC	Federal Communications Commission	–/–
FET	Field effect transistor	–/–
FFT	Fast Fourier transform	5.2
FM	Frequency modulation	–/–
FTC	Continuous Fourier transform	→ 5.1.4
GCPW	Grounded coplanar waveguide	→ 3.3.4
GPIB	General purpose interface bus	→ 2.9.2
GPS	Global positioning system	4.14
GSG	Ground-Signal-Ground	→ 3.3.4
HF	High frequencies (f = 3 MHz to 30 MHz)	–/–
HPA	High power amplifier	→ 7.6
HPBW	Half power beam width	→ 9.2
IDFT	Inverse discrete Fourier transform	→ 5.2
IDT	Interdigital transducer	→ 6.2
IEEE	Institute of Electrical and Electronic Engineers	–/–
IF	Intermediate frequency	→ 8.1.1
IFFT	Inverse fast Fourier transform	→ 5.2.4
IFTC	Inverse continuous Fourier transform	→ 5.1.4
IMB	Imbalance	→ 4.12
IMD	Intermodulation distortion	→ 7.5.1
IP	Internet protocol	2.9.2
IP2	2nd order intercept	→ 7.5.1
IP3	3rd order intercept	→ 7.5.1
ISAR	Inverse synthetic aperture radar	→ 9.3
ISO	International Standardization Organization	–/–
IVI	Interchangeable virtual instruments	→ 2.9.3
L	Line standard	→ 3.3.1
LF	Low frequencies (f = 30 kHz to 300 kHz)	–/–
LHCP	Left-hand circular polarization	9.1
LO	Local oscillator	→ 8.1.1
LSB	Lower sideband	→ 8.1.1
LTI	Linear time invariant	→ 5.1

LVDS	Low voltage differential signaling	→ 2.9.3
LXI	LAN eXtensions for Instrumentation	→ 2.9.3
M	Match standard	→ 3.3.1
MESFET	Metal semiconductor FET	–/–
MF	Medium frequencies (f = 300 kHz to 3 MHz)	–/–
MM	Millimeter wave	8.5
MMIC	Microwave monolithic integrated circuit	–/–
MW	Microwave	–/–
N	Symmetrical network standard	→ 3.3.1
NCO	Numerically controlled oscillator	→ 2.6
NDE	Nondestructive evaluation	→ 6.3
NIST	National Institute of Standards and Technology of the USA	→ 3.3
NMI	National Metrology Institute	→ 3.3
NOI	n^{th} order intercept	→ 7.5.1
NPL	National Physical Laboratory	→ 3.3
O	Open standard	→ 3.3.1
OCXO	Oven controlled temperature crystal oscillator	→ 2.5
OSM	Open short match calibration technique	→ 3.4.1
OSML	Open short match line calibration technique	→ 6.4
PAE	Power added efficiency	→ 4.15
PCB	Printed circuit board	→ 3.3.3
PCMW	Planar circuit mounted in waveguide	8.5
PLL	Phase locked loop	→ 2.6
PTB	Physikalisch-Technische Bundesanstalt (German national metrology institute)	→ 3.6.4
PTFE	Polytetrafluorethylene (Teflon)	3.4.5
PTP	Precision timing protocol	→ 2.9.3
R	Reflect standard	→ 3.3.1
R&S®	Rohde&Schwarz	–/–
RADAR	Radio detection and ranging	9.3
RCS	Radar cross section	→ 9.3
RF	Radio frequency	→ 8.1.1
RHCP	Right-hand circular polarization	9.1
RMS	Root mean square	–/–
Rx	Receive (antenna), receiver	→ 9.3

S	Short standard	→ 3.3.1
S/N	Signal-to-noise ratio	2.8.2
SAR	Synthetic aperture radar	9.3
SAW	Surface acoustic waves	→ 6.2
SCPI	Standard Commands for Programmable Instruments	→ 2.9.3
SCR	Single-reflector compact test range	→ 9.1
SHF	Super high frequencies (f = 3 GHz to 30 GHz)	–/–
SI	Système International d'Unités (International System of Units)	→ 3.6.1
SMD	Surface-mount device	→ 3.3.3
SMT	Surface-mount technology	3.3.3
SNR	Signal-to-noise ratio	2.8.2
SOI	Second order intercept	→ 7.5.1
SSRAM	Synchronous static RAM	2.7
SWR	Standing wave ratio (= VSWR)	→ 4.3
T	Through (Thru) standard	→ 3.3.1
TCXO	Temperature compensated crystal oscillator	→ 2.5
TDR	Time domain reflectometry	→ 5.1.2
TNA	Through network attenuation calibration procedure	→ 3.4.2
TOI	Third order intercept	→ 7.5.1
TOM	Through open match calibration procedure	→ 3.4.2
TOM-X	Through open match crosstalk calibration procedure	→ 3.4.4
TOSM	Through open short match calibration procedure	→ 3.4.3
TRL	Through reflect line calibration procedure	→ 3.4.2
TRM	Through reflect match calibration procedure	→ 3.4.2
Tx	Transmit (antenna), transmitter	→ 9.3
U	Unknown through standard	→ 3.3.1
UHF	Ultra high frequencies (f = 300 MHz to 3 GHz)	–/–
ULF	Ultra low frequencies (f = 300 Hz to 3 kHz)	–/–
UOSM	Unknown through open short match calibration procedure	→ 3.4.2
USB	Universal serial bus	→ 2.9.2
USB	Upper sideband	→ 8.8.1
UUT	Unit under test (= DUT)	2.1
VARICAP	VARIable CAPacitance diode	–/–
VCO	Voltage controlled oscillator	→ 2.5

VDE	Verband der Elektrotechnik, Elektronik und Informationstechnik e.V. (Association for Electrical, Electronic & Information Technologies)	–/–
VDI	Verein Deutscher Ingenieure e.V. (Association of German Engineers)	–/–
VHF	Very high frequencies (f = 30 MHz to 300 MHz)	–/–
VLF	Very low frequencies (f = 3 kHz to 30 kHz)	–/–
VNA	Vector network analyzer	→ 1.3
VNC	Virtual network computing	→ 2.9.2
VPN	Virtual private network	→ 2.9.2
VSWR	Voltage standing wave ratio (= SWR)	→ 4.3
WLAN	Wireless LAN	2.9.2
XTAL	Crystal	2.5
YIG	Yttrium-Iron-Garnet	→ 2.5
YTO	YIG tuned oscillator	2.5

A5 Sources of figures

Fig. 1.1.3 ISAR image of a model of a Boeing 747, from [1EZ52]

Fig. 2.2.8 Measurement uncertainty as a function of directivity, test port match and measured reflection loss, from [RS04], p. 222

Fig. 2.9.2 The unified instrument model in the SCPI standard from Jochen Wolle's "Standards for remote control of devices". Training presentation, July 12, 2002, Rohde & Schwarz

Fig. 2.9.3 The VISA I/O library and the SCPI standard; for source, see Fig. 2.9.2

Fig. 3.3.15 Test fixture for RF substrates (R&S®) from ROHDE & SCHWARZ Messgerätebau GmbH

Fig. 3.5.2 Sample evaluation using the T-checker program, from [1EZ43]

Fig. 3.5.5 Transmission (black) and reflection (blue) with losses neglected, from [1EZ48]

Fig. 3.5.8 Measured quantity |M| vs. frequency, from [RJ03]

Fig. 3.5.11 Measured quantity |M| vs. frequency, from [RJ03]

Fig. 5.5.1 Double-logarithmic display of the function $1/f$; from [1EZ44]

Fig. 6.4.2 Measurement result for verification in the time domain; from [RJ03]

Fig. 6.4.3 Example of the locus curve for the effective directivity, measured using a very narrow time gate; from [RJ03]

Fig. 6.4.4 Example of the locus curve for the effective source match, measured using a very narrow time gate; from [RJ03]

Fig. 6.4.6 Juroshek configuration from [RJ03]

Fig. 7.5.1 Output spectrum of a nonlinear circuit with two-tone input signal with $f_1 < f_2$; from [Ra00], p. 112

Fig. 9.1.4 Compensated compact test range at the Munich University of Applied Sciences (system type: CCR 20/17 from EADS Astrium); from [Fa05]

Fig. 9.1.5 Test of a multi beam/shaped beam reflector antenna in the 12 GHz range (compensated compact test range: Munich University of Applied Sciences, antenna: Kathrein); from [Fa05]

Fig. 9.2.2 Radiation diagram of a 12 GHz offset parabolic antenna with high gain without hardgating (red) and with hardgating (blue); from [HF99]

Fig. 9.3.2 ISAR image of the model of a Boeing 747, from [1EZ52]

The author would like to thank these sources for granting permission to reproduce these figures.

A6 Bibliography

The first two letters of the author's last name are used to form the short identifier [Xx99]. In case of works that have multiple authors, the first letters of the last names of the first two authors are used and the two letters are written in capitals [XX99]. This is followed by the last two digits of the year of publication. Application notes are identified using the standard Rohde&Schwarz identifier. They always begin with a number so that readers can tell whether the bibliographic data associated with an identifier can be found in section A.6.1 or A.6.2.

A6.1 Selected application notes[1) 2)]

[1EF46] GANZERT, JOHANNES: Using MATLAB® for Remote Control and Data Capture with R&S Spectrum and Network Analyzers, version 0E, Dec. 2001[3)]

[1EF47] GANZERT, JOHANNES: Remote Control of R&S Spectrum and Network Analyzers via LAN, version 2E, Jun. 2003[3)]

[1EZ29] BEDNORZ, THILO: Measurement Uncertainties for Vector Network Analysis, version 1E, Oct. 1996

[1EZ43] OSTWALD, OLAF: T-Check Accuracy Test for Vector Network Analyzers utilizing a Tee-junction, version 0E, Jun. 1998

[1EZ44] OSTWALD, OLAF: Time Domain Measurements using Vector Network Analyzer ZVR, version 1D, May 1998

[1EZ45] SIMON, JOCHEN: Virtual embedding networks for ZVR network analyzers, version 0E, Sep. 1998

[1EZ48] OSTWALD, OLAF: Measurement accuracy of the ZVK vector network analyzer, version 0E, Jan. 2001

[1EZ51] JÄGER, HARALD: Measurement Method for Determining the Equivalent Reflection Coefficient of Directional Couplers and Power Splitters, version 1E, Aug. 2002

[1EZ52] GLEISSNER, ALBERT and VOKURKA, J.: Antenna Measurements, RCS Measurements and Measurements on Pulsed Signals with Vector Network Analyzers R&S ZVM, R&S ZVK, version 0E, Aug. 2004

1) As of 12/2006
2) These application notes can be downloaded from the Rohde&Schwarz website at http://www.rohde-schwarz.com/.
3) Sample programs or software associated with this application note.

[1EZ53] SIMON, JOCHEN: Measuring Balanced Components with Vector Network Analyzer ZVB, version 0E, Sept. 2004

A6.2 Books, scientific publications and standards

[BE95] BOCKELMAN, DAVID and EISENSTADT, WILLIAM: Combined Differential and Common-Mode Scattering Parameters: Theory and Simulation. IEEE Transactions on Microwave Theory and Techniques, Vol. 43, No. 7, Jul. 1995, pp. 1530-1539

[BS91] BRONSTEIN, I. N. and SEMENDJAJEW. K. A.: Taschenbuch der Mathematik [translation: Pocket mathematics reference]. B. G. Teubner, Stuttgart, Leipzig; published by Harri Deutsch, Thun, Frankfurt/Main, 25th edition, 1991, ISBN 3-8154-2000-8 (Teubner), ISBN 3-87144-492-8 (in German)

[BS98] BACHMAIR, HANS and STUMPER, ULRICH: Aktuelle Probleme bei der Weitergabe von HF-Meßgrößen [translation: Current problems in transferring RF measured quantities]. Presentations at the 139th PTB seminar on May 13, 1998, PTB-E58, Braunschweig, June 1998, pp. 53-84, ISBN 3-89701-173-5

[Ca88] CATTERMOLE, KENNETH W.: Signale und Wellen - mathematische Grundlagen und technische Anwendungen [translation: Signals and waves - Mathematical fundamentals and technical applications]. VCH Verlagsgesellschaft, Weinheim, 1st edition, 1988, ISBN 3-527-26612-7

[CE05] CHO, CHOONGEOL and EISENSTADT, WILLIAM R.: IIP3 Estimation From the Gain Compression Curve. IEEE Transactions on Microwave Theory and Techniques, Vol. 53, No. 4, Apr. 2005, pp. 1197-1202

[CT65] COOLEY, JAMES W. and TUKEY, JOHN W.: An Algorithm for the Machine Calculation of Complex Fourier Series. Mathematics of Computations, Vol. 19, Apr. 1965, pp. 297-301

[Di87] DIRSCHMID, HANS JÖRG: Mathematische Grundlagen der Elektrotechnik [translation: Mathematical principles of electrical engineering]. Vieweg&Sohn Braunschweig/Wiesbaden, 2nd edition, 1987, ISBN 3-528-13034-2

[EA00] EUROPEAN CO-OPERATION FOR ACCREDITATION: Guidelines on the Evaluation of Vector Network Analyzers (VNA). Publication Reference EA-10/12, May 2000

[EH79] ENGEN, GLENN and HOER, CLETUS: "Thru-Reflect-Line" - An Improved Technique for Calibrating the Dual Six-Port Automatic Network Analyzer. IEEE Transactions on Microwave Theory and Techniques, Vol. 27, No. 12, Dec. 1979, pp. 987-993

[ES91] EUL, HERMANN-JOSEF and SCHIEK, BURKHARD: A Generalized Theory and New Calibration Procedures for Network Analyzer Self-Calibration. IEEE Transactions on Microwave Theory and Techniques, Vol. 39, No. 4, Apr. 1991, pp. 724-731

[ES92] EDWARDS, MARION LEE and SINSKY, JEFFREY: A New Criterion for Linear 2-Port Stability Using a Single Geometrically Derived Parameter. IEEE Transactions on Microwave Theory and Techniques, Vol. 40, No. 12, Dec. 1992, pp. 2303-2311

[Fa05] FASOLD DIETMAR: Description of Antenna Test Facility. Laboratory for Satellite Communications, Munich University of Applied Science, Munich, 2005

[FP92] FERRERO, ANDREA and PISANI, UMBERTO: Two-Port Network Analyzer Calibration Using an Unknown "Thru". IEEE Microwave and Guided Wave Letters, Vol. 2, No. 12, Dec. 1992, pp. 505-507

[GB04] GASSELING, TONY and BARATAUD, DENIS and MONS SÉBASTIEN and NEBUS, JEAN-MICHEL and VILLOTTE, JEAN PIERRE and OBREGON, JUAN and QUERE, RAYMOND: Hot Small-Signal S-Parameter Measurements of Power Transistors Operating Under Large-Signal Conditions in a Load-Pull Environment for the Study of Nonlinear Parametric Interactions. IEEE Transactions on Microwave Theory and Techniques Vol. 52, No. 3, Mar. 2004, pp. 805-812

[HF98] HARTMANN, J. and FASOLD, DIETMAR: A Flexible Hardgating System as a Diagnostic Tool in Single and Double Reflector Compact Ranges. Proceedings of AMTA 98, 25. to 29. Oct. 1998, Montreal, Canada

[HF99] HARTMANN, J. and FASOLD, DIETMAR: Identification and Suppression of Measurement Errors in Compact Ranges

by Application of an Improved Hardgating System. Proceedings of 22nd ESTEC Antenna Workshop of Antenna Measurements, 11. to 12. May 1999, Noordwijk, The Netherlands

[HS04] HELD, WERNER and SIMON, JOCHEN: Verfahren zum Kalibrieren von Netzwerkanalysatoren [translation: Technique for calibrating network analyzers]. Published patent application, German Patent and Trade Mark Office, DE 10235221A1, Feb. 2004

[IS93] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: Guide to the Expression of Uncertainty in Measurement. First edition 1993, corrected and reprinted 1995, International Organization for Standardization Geneva, Switzerland

[Ju97] JUROSHEK, JOHN R.: A Direct Calibration Method for Measuring Equivalent Source Mismatch. Microwave Journal, Oct. 1997, pp. 106–118

[KS71] KRUPPA, W and SODOMSKY, K.F.: An Explicit Solution for the Scattering Parameters of a Linear Two-Port Measured with an Imperfect Test Set. IEEE Transactions on Microwave Theory and Techniques Vol. 19 No. 1, Jan. 1971, pp. 122–123

[Ku01] KURZROK, RICHARD M.: Design Idea – Coax-to-Waveguide Adapters Meet Needs of Communications Equipment, Applied Microwave and Wireless, Vol. 13, Feb. 2001, pp. 98–100

[Kü73] KÜPFMÜLLER, KARL: Die Systemtheorie der elektrischen Nachrichtenübertragung [translation: System theory in electrical transmission of information]. Hirzel Verlag, Stuttgart, 4th edition, 1974, ISBN 3-7776-0248-5

[LX05] LXI CONSORTIUM: LXI STANDARD. LXI Consortium Inc., Revision 1.0, 23. Sept. 2005

[Ma77] MARKO, HANS: Methoden der Systemtheorie [translation: Methods in system theory]. Springer Verlag, Berlin, Heidelberg, New York, 1st edition, 1977, ISBN 3-540-08106-2

[MG92] MEINKE, H. H. and GUNDLACH, FRIEDRICH-WILHELM, publishers LANGE, KLAUS and LÖCHERER, KARL-HEINZ: Taschenbuch der Hochfrequenztechnik [translation: Pocket reference for radio engineering]. Student edition, vols. 1–3, Springer Verlag, Berlin, Heidelberg, New

York, 5th edition, 1992, ISBN 3-540-54714-2 (vol. 1), ISBN 3-540-54715-0 (vol. 2), ISBN 3-540-54716-9 (vol. 3)

[MS66] MACKENZIE, T. E. and SANDERSON, A. E.: Some Fundamental Design Principles for the Development of Precision Coaxial Standards and Components. IEEE Transactions on Microwave Theory and Techniques, Vol. 14, No. 1, Jan. 1966, pp. 29-39

[Pi00] PIEPER, JOHN M.: Automatic Measurement Control. Rohde & Schwarz GmbH & Co. KG, Munich, 1st edition, 2000

[PT04] PHYSIKALISCH-TECHNISCHE BUNDESANSTALT: Website of the German national metrology institute (German abbreviation: PTB) in Braunschweig, Germany. Work group 2.22, RF measurement technology, access on May 18, 2004, responsible official under Press Act: Jürgen Rühaak, http://www.ptb.de

[Ra00] RAUSCHER, CHRISTOPH: Grundlagen der Spektrumanalyse [translation: Fundamentals of spectrum analysis]. Rohde & Schwarz GmbH & Co. KG, Munich, 1st edition, 2000

[Re74] REHNMARK, STIG: On the Calibration Process of Automatic Network Analyzer Systems. IEEE Transactions on Microwave Theory and Techniques, Vol. 22, No. 4, Apr. 1974, pp. 457-458

[RJ03] REICHEL, THOMAS and JÄGER, HARALD: New Verification & Calibration Method for Vector Network Analyzers. 33rd European Microwave Week 6.-10. Oct. 2003, Munich

[Ro62] ROLLET, J.M.: Stability and Power Gain Invariants of Linear Two-ports. IRE Transactions and Circuit Theory, Vol. CT-9, Mar. 1962, pp. 29-32

[RS04] Rohde & Schwarz: Test & Measurement Products Catalog 2004/2005. Rohde & Schwarz GmbH & Co. KG, Munich, 2004

[RS05] ROLFES ILONA and SCHIEK, BURKHARD: Multiport Method for the Measurement of the Scattering Parameters of N-Ports. IEEE Transactions on Microwave Theory and Techniques, Vol. 53, No. 6, Jun. 2005, pp. 1990-1996

[RS69] RABINER, LAWRENCE R. and SCHAFER, RONALD W. and RADER, C. M.: The Chirp z-Transform Algorithm. IEEE Transactions on Audio and Electroacoustics, Vol. AU-17, No. 2, Jun. 1969, pp. 86-92

[Sc84] SCHIEK, BURKHARD: Meßsysteme der Hochfrequenztechnik [translation: Measurement systems in radio engineering]. Hüthig Verlag, Heidelberg, 1st edition, 1984, ISBN 3-7785-1045-2

[SG03] SCHIEK, BURKHARD and GRONEFELD, ANDREAS: Verfahren zum Kalibrieren eines zwei Messtore und vier Messstellen aufweisenden vektoriellen Netzwerkanalysators nach dem 15-Term-Fehlermodell [translation: Technique for calibration of a vector network analyzer with two test ports and four receivers based on the 15-term error model]. Patent at the German Patent and Trade Mark Office, DE 197 43 712 C2, dated October 2, 2003

[Ta95] TAYLOR, BARRY N.: NIST Special Publication 811: Guide for the Use of the International System of Units (SI). National Institute of Standards and Technology (NIST), Gaithersburg USA, Apr. 1995

[Vi94] VIDKJAER, JENS: Network Analyzer Uncertainty Computations for Small-Signal Model Extractions. Technical University of Denmark, Electromagnetics Institute, Lyngby, Denmark, R549, Feb. 1994

[WK99] WOHLGEMUTH, O. and KREMS, T. and REUTER, R. and RODWELL, M.J.W. and HAYDL, W. and SCHLECHTWEG, M.: Integrated Directional Coupler for 90 and 180 GHz. IEEE Microwave and Guided Wave Letters, Vol. 9, No. 8, Aug. 1999, pp. 308–310

[ZB93] ZINKE, OTTO and BRUNSWIG, HEINRICH, publishers VLCEK, ANTON and HARTNAGEL, HANS LUDWIG: Hochfrequenztechnik [translation: Radio engineering]. Vols. 1–2, Springer Verlag, Berlin, Heidelberg, New York, vol. 1: 5th edition, 1995, ISBN 3-540-58070-0, vol. 2: 4th edition, 1993, ISBN 3-540-55084-4

A7 Index

Symbols
10-term error model 119
12-term error model 119
15-term error model 123
1 dB compression point 301, 344
1 dB gain compression point 301
3-term error model (OSM technique) 111
7-term error model 113

A
accredited 155
active mixer 346
active test port 22
active test set 48
actual level 289
adapter 124
air line 97
alternating mode 55
ambiguity range 233
amount of substanc 152
amplitude imbalance 200
analyzer with an N-port switching matrix 76
anechoic chamber 364
antenna 360
antenna gain 368
aperture 170
aperture stepwidth 170
arbitrary gate shape 266
arbitrary mode 296
array antenna 360
Attenuator standard (A, attenuator) 98
auto length 175
automatic calibration equipment 90
automatic gain control (AGC) 51
automatic level control (ALC) 288
available power 332
azimuth angle 366

B
balanced transmission line 188
bandpass mode 241
Beatty standard 135
biasing 48
biasing current 346
bias T 48
bistatic radar 369
Bohman window, steep falloff 265
boosted source 328
bulk propagation mode 276
burst signal 243

C
cable loss 195
calibration 88
calibration certificate 155
calibration chain 155
calibration kit 89
calibration standard 88
Cartesian diagram 64
causal 215
center frequency 61
channel 60
channel settings 60
characteristic data 88
characterization 89
chopper mode 55
coefficients 89
combined standard measurement uncertainty 145
combiner 300
command 74
common-mode 188
common-mode power gain 201
common-mode rejection ratio 201
compact antenna test range 364
compensated compact test range 364
complex 65
compression effects 86

computer 23
conditionally stable 177
continuous Fourier transform 225
controller 289
controller parameter 290
conversion loss 343
convolution 216
coplanar calibration standard 105
coplanar waveguide (CPW) 106
coupling coefficient 24
cross-polarization 364
cross modulation 316

D
data entry key 58
dB magnitude 65
DC block 48
DC measurement input 211, 303
DC power substitution method 156
DC supply 48
decimal fraction 153
decimal multiple 153
deembedding 183
desired/actual comparison 289
desired level 289
device response 74
diagram 64
- Cartesian diagram 64
- inverted Smith chart 64
- polar diagram 64
- radiation diagram 367
- Smith chart 14, 64
differential-mode 188
differential-mode power gain 201
differential tone 316
digital control 290
digital interface 70
digital signal processing (DSP) 51
Dirac impulse 215
direct electromagnetic coupling 276

direct generator and receiver access 298
directional coupler 36
directional element 23
directivity 26
- effective directivity 137
distance resolution 250
Dolph-Chebyshev window, arbitrary side lobes 265
down-conversion 340
dual-reflector compact test range 364
dual directional coupler 41

E
effective directivity 137
effective load match 142
effective reflection tracking 142
effective source match 140
effective system data 87, 136
efficiency 210
electrical length 173
electric current 152
elevation angle 366
embedding 178
equal-level far-end crosstalk loss 195
equilibrium state 222
equivalent focus length 364
error
- linear error 86
- nonlinear error 86
error handling 74
error network 87
error terms 32, 87, 88
error two-port 32
estimated value 143
Euler's formula for complex numbers 375
experimental standard deviation 144
exponential products of the base unit 153
external test set 328
external trigger 202
external trigger input 70

F
factory system error correction 56
far-field distance 361
far-field region 361
feed 364
female (f) 158
Field effect transistors (FETs) 306
firmware 23
first harmonic 305
format 65
Fourier transform
- continuous Fourier transform 225
- inverse continuous Fourier transform (IFTC) 225
- inverse fast Fourier transform (IFFT) 244
free-space impedance 362
free-space propagation 360
frequency extension 354
frequency extension module 354
frequency response 225
frequency sample 227
frequency step size 227
frequency sweep 60
- linear frequency sweep 60
- logarithmic frequency sweep 60
fringing capacitance 92
full N-port analyzer 76
full verification 136
fundamental constant 152
fundamental frequency 305

G
generator 22
generator step attenuators 46
GPIB (General Purpose Interface Bus) 71
group delay 65
GSG probe 106

H
half power beam width 368
Hamming window, low first side lobe 264

Hann window, normal profile 264
hardgating 365
harmonic 305
- first harmonic 305
- second harmonic 305
harmonic grid 227, 246
harmonic level 305
- higher-order harmonic level 306
harmonic mixer 346
harmonic mixing 53
harmonic suppressions 306
harmonic tuning 335
HELP key 58
heterodyne principle 18, 50
high-speed oscilloscope 217
higher-order harmonic level 306
higher order mixing products 341
highest side lobe amplitude 238
high power amplifier (HPA) 328
homodyne principle 18
hysteresis effect 50

I
I/Q demodulation 51
I/Q mixer 346
IEC bus/IEEE bus 71
IF bandwidth (measurement bandwidth) 62
image reception 340
image rejection mixer 346
imaginary part 65
imbalance (IMB) 200
impedance step 223
impulse response 215
incident wave 13
inherent spurious response 342
inphase component 52
input/output far-end crosstalk loss 194
intercept
- second order intercept (SOI) 317
- third order intercept (TOI) 317

intercept point 344
- n-th order intercept point 317
- second harmonic intercept point 307
- third harmonic intercept point 308
interface
- digital interface 70
- GPIB (General Purpose Interface Bus) 71
- IEC bus/IEEE bus 71
- LAN interface 71
- USB interface 72
intermediate frequency 50
intermediate frequency signal 338
intermodulation product 315, 316
intermodulation suppression 317
International System of Units 151
inverse chirp z-transform (CZT) 244
inverse continuous Fourier transform (IFTC) 225
inverse fast Fourier transform (IFFT) 244
inverse synthetic aperture radar 371
inverted case 339
inverted Smith chart 64
inverting conversion 340
isolation 142
isopropyl alcohol 84
isotropic radiator 368
IVI standard (Interchangeable Virtual Instrument) 75

J
Josephson constant 153

L
LAN interface 71
length 152
length offset 182
level 154
limit lines 70
linear average 143
linear error 86
linear frequency sweep 60
linear magnitude 65

linear network 214
Line standard (L) 97
LO-IF feedthrough 342
LO-IF isolation 342
LO-RF feedthrough 342
LO-RF isolation 343
local oscillator 338
logarithmic frequency sweep 60
LO level 345
lossy transmission line 263
lower sideband 340
lowpass mode 241
LRL technique 117
LTI system 214
luminous intensity 152
LXI standard (LAN eXtensions for instrumentation) 76

M
magnitude
- dB magnitude 65
- linear magnitude 65
main reflector 364
male (m) 158
marker 69
mass 152
Match (M) 93
Match (M) (or load) 101
matching point 14
Maximum flatness gate 266
maximum impulse level 345
measured quantity 63
measured value 25, 56
measurement accuracy test 133
measurement bandwidth 62
measurement receiver 24
measurement uncertainty 80
MEAS WIZARD 58
menubar 58
microstrip calibration standards 102
microwave monolithic integrated circuit 183

mismatch 137
mismatched transmission line 262
mixed-mode S-matrix 189
mixer
- active mixer 346
- harmonic mixer 346
- I/Q mixer 346
- image rejection mixer 346
- mixer burnout limit 345
- mixer constant 338
- mixer mode 346
- reference mixer 348
- subharmonic mixer 346
mixer burnout limit 345
mixer constant 338
mixer mode 346
modal decomposition 189
mono-static radar 369
more capacitive 97
more inductive 97
multipath propagation 256
multiple reflection 29, 221
multiple source concept 295
multiplier 338

N
N+1 receiver analyzer 76
N-port 16
N-port switching matrix 78
n-th order intercept point 317
national and international standards 155
navigation keys 58
near-end crosstalk loss 193
near-field transformation 362
near field 362
network analyzer 12
- scalar network analyzer 17
- vector network analyzer 17
noise figure 84
noninsertable device 126

noninverted case 339
noninverting conversion 340
nonlinear error 86
normal gate 266
normalization 128
number 153
number of points 61
numerically controlled oscillator (NCO) 51

O
offset short 101
on-wafer measurements 106
one-path two-port calibration 128
one-port device 14
Open (O) 92
open-circuit point 14
open site testing 362
order 316
original meter 152
oscillator
- numerically controlled oscillator (NCO) 51
- oven controlled temperature crystal oscillator (OCXO) 49
- temperature-compensated crystal oscillator (TCXO) 49
- voltage-controlled oscillator (VCO) 50
- Yttrium-Iron-Garnet (YIG) oscillator 50
OSM technique 111
outer conductor mating plane 88
oven controlled temperature crystal oscillator (OCXO) 49
overlap 342
overtone 305

P
partial derivative 145
partial measurement 54
partial trigger 209
passband ripple 255
passive test port 347
passive test set 49
phase 65
phase delay 173

phase imbalance 200
phase locked loop (PLL) 49, 52
PI controller 289
pin depth gage 84
planarity 108
plane wavefronts 361
point trigger 208
polar diagram 64
positioner 360
power added efficiency 210
power attenuation figure 153
power calibration 290
power divider 301
power gain figure 153
power level 154
power matching 332
power sweep 60, 294
preamplifier 85
precision resistor 211
prefix 153
primary standards 155
printed circuit boards (PCBs) 102
probe 106
processing chain 55
pulsed operating mode 208
pyramidal absorber 364

Q
quadrature component 52
quantum effect 152
query 74
quiet zone 365

R
radar cross section 370
radar cross section measurement 369
radar range equation 370
radiation diagram 367
radiation pattern 366
radio frequency signal 338

radio transmission 361
random measurement error 80
raw measured values 88
raw measurement data 23
raw numerical measured values 55
raw system data 87
RC parallel connection 261
RC series connection 260
real part 65
receiver 23, 24
receiver power calibration 293
receiver step attenuator 46
reciprocity 361
recognition 132
rectangle window, no profiling 264
rectangular distribution 145
reference impedance 14
reference mixer 348
reference plane 88
reference receiver 24
reflected wave 13
reflection 154
reflection coefficient 14
reflection tracking 25
reflect standard (R) 97
relative impulse width 239
repeatability 81
response 215, 216
- impulse response 215
- step response 216
return loss 154
reverse coupler 301
RF-IF feedthrough 342
RF-IF isolation 343
RF level 345
RF trigger 203
RL parallel connection 260
RL series connection 259
rotary joint 360

S
S-matrix 16
S-parameter 15
S-parameter matrix 16
sample time 227
sampling 228
scalar network analyzer 17
scattering parameter 15
SCPI standard (standard commands for programmable instruments) 73
second harmonic 305
second harmonic intercept point 307
second IEC bus interface 71
second order intercept (SOI) 317
self-calibration techniques 97
self-induced voltage 202
sensitivity coefficient 146
setting command 74
setup keys 58
shape factor 85
Short (S) 91, 100
- offset short 101
short-circuit point 14
Shunt resistor 223
SI
- SI base quantity 152
- SI base unit 151
- SI derived unit 152
- SI prefix 153
side-lobe suppression 368
sideband
- lower sideband 340
- upper sideband 340
sideband separation 346
side lobe 368
- highest side lobe amplitude 238
- side-lobe suppression 255
- side lobe, ringing 235
signal feeding 360
signal retransmission 360

single-ended port 189
single-ended S-parameter 189
single-reflector compact test range 364
single-tone signal 305
sliding match 94, 102
Smith chart 14, 64
softkeys 57
software trigger 203
source match 115
source power 62
source power calibration 291
span 61
spectrum, periodical 228
spherical wavefront 361
spurious response 342
standard deviation 143
standing wave ratio 166
start frequency 61
steady state error 289
steep edges gate 266
steepest edges gate 266
step generator 217
step response 216
stimulus 22, 215
stop frequency 61
stub 333
subharmonic mixer 346
subreflector 364
substrate 102
support keys 58
sweep range 61
sweep time 63
sweep type 60
SWR, VSWR 65, 166
symmetrical coupled conductors 188
symmetrical network 98
systematic measurement errors 80
system data
- effective system data 87, 136
- raw system data 87

system error correction 87
system of units 151

T
T-check 133
T-checker 133
T-check parameter 134
T-junction 223
Taylor series 304
temperature-compensated crystal oscillator (TCXO) 49
test fixture 103, 160
test port 22, 158
- active test port 22
- passive test port 347
test port match 29
test set 22
- active test set 48
- external test set 328
- passive test set 49
thermal equilibrium 81
thermal noise 84
thermodynamic temperature 152
third harmonic intercept point 308
third order intercept (TOI) 317
three-port directional coupler 37
three-port network 16
Through (T) 96, 102
time 152
time-domain 214
time-domain analysis 214
time-domain option 245
time-domain samples 227
time constant 220
time domain reflectometry 217
time invariance 214
time resolution 229
timeslot 208
time sweep 60
title 350
TNA technique 117

TOM technique 116
top view 365
torque wrench 84
TOSM technique 122
trace 63
traceable 155
trace generation 57
trace mathematics 68
trace settings 63
transmission 154
transmission coefficient 24
transmission line
- air line 97
- balanced transmission line 188
- lossy transmission line 263
- mismatched transmission line 262
- stub 333
- terminated with R 261
- twisted pair transmission line 188
transmission loss 154
trigger delay 202, 208
triple transit signal 276
TRL technique 117
TRM technique 117
true-differential-stimulus 337
true value 143
tuner 333
twisted pair transmission line 188
two-port device 15
two-tone signal 315

U
unconditionally stable 176
UNDO 58
uniform distribution (rectangular distribution) 145
unit step 216
unknown through 98
unwrapped phase 65
up-conversion 340
upper sideband 340

USB hub 72
USB interface 72
user calibration 56
user interface 57

V
vector network analyzer 17
velocity factor 270
virtual impedance transformer 182
virtual transformation network 178
VISA I/O library 75
voltage-controlled oscillator (VCO) 50
voltage standing wave ratio 65, 166
von Klitzing constant 153
VSWR bridge 34

W
wafer prober 106
wave
- bulk propagation mode 276
- incident wave 13
- plane wavefront 361
- reflected wave 13
wavefront 361
waveguide 99
- waveguide calibration standard 99
- waveguide short 100
waveguide calibration standard 99
waveguide short 100
wave quantity 13
window function 231

Y
Yttrium-Iron-Garnet (YIG) oscillator 50

A8 Current network analyzer product lines from Rohde&Schwarz[1)]

Due to increasing integration, DUTs with more than two ports are now becoming more and more important. Balanced DUTs that have more than two ports are also increasingly common. Accordingly, the Rohde&Schwarz network analyzer family includes models with up to eight test ports. These instruments are compact although they offer a wide range of functions. Of course, the complexity increases along with the number of functions. An innovative operating concept has been developed to simplify their operation nevertheless. It includes:

- A measurement and calibration wizard to handle complex operating sequences when so desired
- Simple handling of balanced measurements with incorporation of true-differential-measurements depending on the instrument type
- An undo key to reverse previous operating steps
- A context-sensitive help function that includes the relevant remote-control commands (among other things)
- Special operating modes for mixer and harmonic measurements. Proper configuration for the required power calibration is organized by a special wizard for each operating mode, which considerably simplifies operation of the instruments.

And much more.

The network analyzer families R&S®ZVA/R&S®ZVT and R&S®ZNB/R&S®ZNC are available with different options.

The following pages discuss these models in somewhat greater detail. If you have any questions, please contact your nearest Rohde&Schwarz sales office. If you need data sheets and other product information, please visit the company website at www.rohde-schwarz.com.

Subject to technical change at any time.

1) As of September 2011

Brief overview of the network analyzer family

	R&S® ZVA	R&S® ZVT	R&S® ZNB
Multiple source concept	•	•	•
Hot S-parameter measurement	•	•	
True differential measurement mode	•	•	
Mixer measurement mode	•	•	•
Harmonic measurement mode	•	•	•
Arbitrary mode	•	•	
Virtual differential measurements	•	•	•
Digital automatic level control	•	•	•
Use of USB power sensors as additional receivers	•	•	•
Generator step attenuators	•		1)
Receiver step attenuators	•		•
Remote control also possible via LAN (alternative to GPIB)	•	•	•
Access to user interface via LAN	•	•	•
Second GPIB interface or alternatively USB/GPIB adapter	•	•	•
Active test sets	•	•	•
Direct generator and receiver access	•	•	
OCXO frequency reference	•	•	•
Embedding/deembedding	•	•	•
Time-domain transformation	•	•	•
Iterative power calibration with selectable accuracy limit	•	•	•
In addition to stability factor k, support for factors μ_1, μ_2	•	•	•
Besides the TOSM and OSM calibration procedures, multiple procedures using the 7-term model and calibration with waveguides	•	•	•
Calibration unit	•	•	•
Pulsed measurements	•	•	
DC measurement inputs	•	•	•

- This function is available with the basic configuration or as an option. Certain functions require at least three test ports. In some cases, this requirement can be satisfied using an external generator.

1) No option "Generator step attenuators" offered but option "Extended power range" available.

The R&S®ZVA vector network analyzers: Reference instruments for the most demanding applications

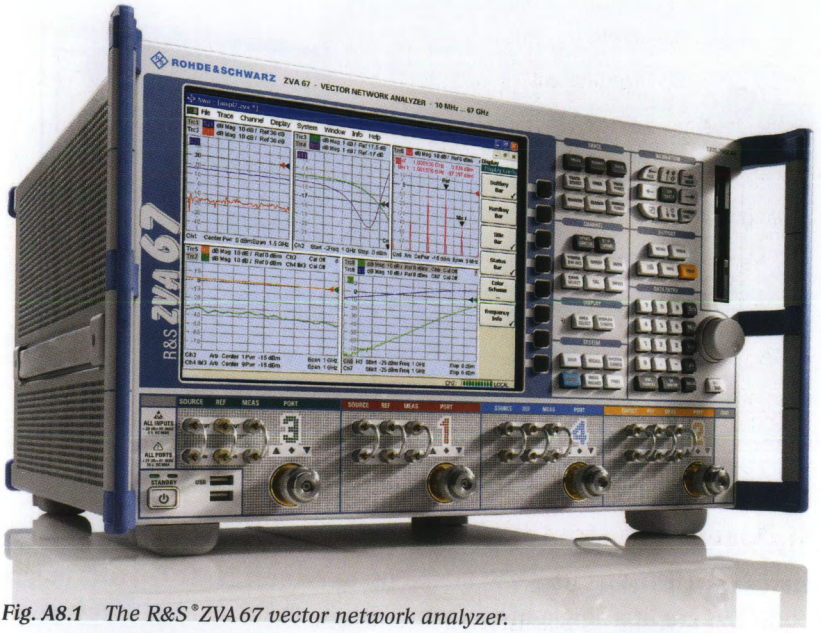


Fig. A8.1 The R&S®ZVA 67 vector network analyzer.

The R&S®ZVA family is the high-end product line in the Rohde&Schwarz network analyzer portfolio. These instruments are available as two- and four-port models with various frequency ranges. The R&S®ZVA 67, for example, covers the frequency range from 10 MHz to 67 GHz. In addition to excellent specifications, the R&S®ZVA models feature powerful, in part unique measurement capabilities, such as:

- Arbitrary frequency conversion measurements for testing standard and sophisticated mixers, amplifiers and T/R modules
- Pulse profile measurements with 12.5 ns time resolution
- True differential stimulation and measurement
- Complex mixer measurements with easy calibration
- Reliable, fast delay testing of converters with internal LO
- Converter delay measurements over wide distances with two VNAs
- All 4-port units have at least two integrated sources. The R&S®ZVA 67 is even equipped with four sources for convenient and fast testing of amplifiers, mixers and double-converting T/R modules

The R&S®ZVT multiport network analyzers: Network analysis with up to eight test ports



Fig. A8.2 The R&S®ZVT8 multiport network analyzer.

The R&S®ZVT is the first network analyzer to combine up to eight test ports (8 GHz model R&S®ZVT8) and up to six test ports (20 GHz model R&S®ZVT20) in a single instrument. With its large number of test ports, it is the perfect instrument for investigating balanced networks, multipath filters and T/R modules. Its special hardware concept with one internal source per two ports (i.e. up to 4 internal sources) also allows it to generate a true-differential stimulus and to test double-converting devices quickly and flexibly with just one unit. The analyzer delivers outstanding performance in terms of stability, reproducibility, accuracy, sweep time, output power, receiver sensitivity and dynamic range. With its measurement capabilities (all functions of the R&S®ZVA are available for the R&S®ZVT as well), its flexibly configurable test set and varied interfaces, the R&S®ZVT is ideal for applications in complex test systems. 8-port and 6-port calibration units enable automatic calibration with a single keystroke.

Network analysis with millimeter-wave converters: Unprecedented convenience up to 500 GHz

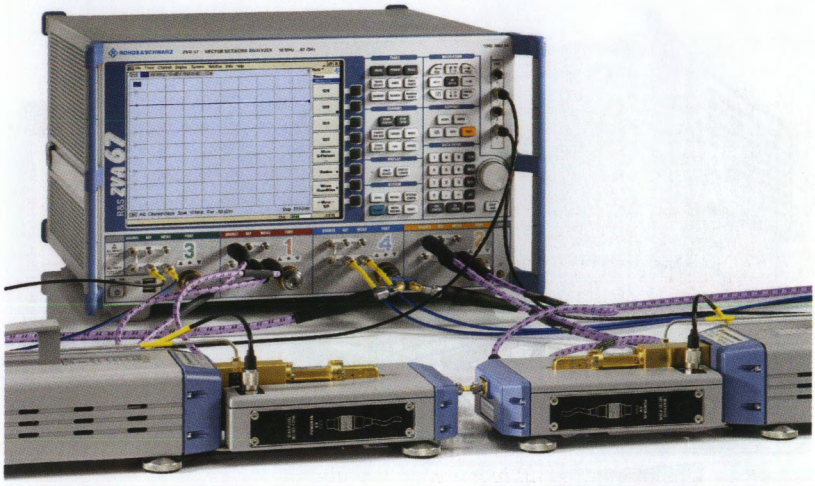


Fig. A8.3 The R&S®ZVA with millimeter-wave converters.

Rohde&Schwarz offers millimeter-wave converters for all common waveguide bands for use with the R&S®ZVA and R&S®ZVT families of network analyzers. The analyzers' and converters' hardware and software are perfectly matched to each other for the most accurate, highest dynamic range measurements and unparalleled ease of operation. The output power can be adjusted on all of the converters, even via the analyzer in some models, making it possible to perform power calibration or the power sweeps required for compression measurements on amplifiers. The R&S®ZVA110 provides complete coverage for the entire range from 10 MHz to 110 GHz.

The R&S®ZNB and R&S®ZNC vector network analyzers: Breaking new ground in productivity



Fig. A8.3 The R&S®ZNB vector network analyzer.

The R&S®ZNB and R&S®ZNC are the first of a new generation of Rohde&Schwarz network analyzers that were developed with an emphasis on both excellent RF performance and user ergonomics. The front of the instrument is dominated by a large, high-resolution touchscreen that clearly displays measurement activity and permits users to work ergonomically over long periods of time. The touchscreen not only displays a large number of traces which can be easily rearranged using drag&drop. It also allows users to quickly select and use softtools which were previously unavailable in this form in network analysis. These softtools simplify operation so that an instrument can typically be configured for a measurement in only three steps. For even more convenience, users can select the display language, such as Russian or Chinese, so that each user feels right at home. The analyzers' outstanding performance (for example, the R&S®ZNB has a dynamic range of more than 140 dB and requires only 250 ns/test point) and longterm stability make the R&S®ZNB and R&S®ZNC analyzers the first choice for both the lab and in production. The R&S®ZNC is designed for the measurement of passive dual-port components up to 3 GHz and the R&S®ZNB is ideal for 2- and 4-port measurements on complex components up to 4.5 GHz, 8.5 GHz, 20 GHz or 40 GHz (2 ports only).