CAS: Network Analysis (Part 2)

F. Caspers
CERN, Geneva, Switzerland

Abstract
The network analyzer has become an absolutely indispensable tool for RF signal analysis. In this lecture, the operating principles of this instrument will be shown. A distinction can be made between scalar and vector network analyzers and their methods of measuring the transmission or reflection coefficients are explained. As digital signal processing has become cheap and easily available over the last 30 years, these instruments have become extremely versatile and powerful. Fourier transformation permits time domain measurements and allows the removal of undesired parts of the signal trace in the time domain by gating. Network analyzers require sophisticated calibration procedures, which are now indispensable for many measurement applications. Nonlinear network analysis has gained increasing interest, in particular for the design of structures with nonlinear properties.

1 MOTIVATION

One of the most common measurement tasks in the field of RF engineering is the analysis of circuits and electrical networks. Such networks can be a simple one port (two pole), containing only a few passive components (resistors, inductances and capacitors) or they may be complex units, consisting of active or nonlinear components with several in- and output ports.

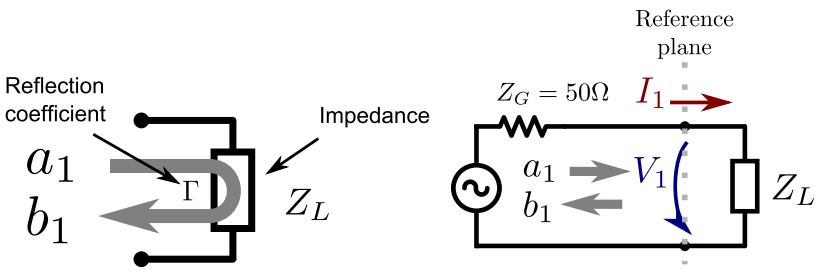
A network analyzer is one of the most versatile and valuable piece of measurement equipment used in an RF laboratory or particle accelerator control room. The best commercially available network analyzers can cover a frequency range of nine orders of magnitude (from a few Hz to several GHz) with up to 0.1 Hz resolution. By exciting the device under test (DUT) with a well defined input in terms of frequency and amplitude and recording the response of the network. For each frequency a complex number is determined (reflection / and or transmission). In microwave engineering the properties of a DUT are usually described as scattering parameters (S-parameters).

In this paper, scalar and vector network analyzers are introduced and measurement techniques for the determination of S-parameters of networks are discussed. S-parameters are basically defined only for linear networks. In the real world, many DUTs are at least weakly nonlinear (e.g. active elements such as amplifiers or mixers). For analysis of these devices certain approximations or extensions of the definitions are required. Practical tools for this analysis include power sweeps or harmonic analysis which is sometimes referred to as X-parameter measurement. Another interesting application is the determination of the beam transfer function (BTF), where the DUT is a circulating particle beam in an accelerator. This paper presents an introduction to the basic principles of this type of instruments, the definition of the related terminology and in particular for measurements in the microwave laboratory and in particle accelerators.

2 S-PARAMETER

2.1 1 Ports

In RF-engineering, **wave quantities** are preferred over currents or voltages for the characterization of RF circuits. We can distinguish between incident (a) and reflected wave, (b). The incident wave travels from a source to the DUT – the reflected one in the opposite direction. The fundamental reason for this terminology is the fact that in RF engineering the linear geometrical dimensions are often larger



(a) incident (a_1) and reflected (b_1) wave. (b) Relation of a_1 and b_1 to V_1 and I_1.

Fig. 1: Wave quantities of a 1 port (with 2 poles) and impedance Z_L

than 10 % of the corresponding free space wavelength. This also requires the definition of a reference plane to which the measurement is referring to. Without this reference plane, in particular the phase of the reflection coefficient is undefined, which renders vectorial measurements impossible. Of course a mathematically correct description of the DUT in terms of voltage and current is still possible and will also return correct results but working with wave quantities turns out to be much more convenient in practice. In particular both methods – if correctly applied – have no fundamental limitation e.g. S-parameters can be used to very low frequencies and voltage and current descriptions can be used to very high frequencies. Both methods are over the full frequency range completely equivalent, the results are mutually convertible. This fact can also be expressed in different terms, namely S-parameters can be converted into impedances and vice versa.

The interface of the DUT to the outside world is one or more **pole pairs**, which are commonly referred to as **ports**. A device with one pair of poles (as in Fig. 1a) is described as one port, where one incident (a_1) and one reflected (b_1) wave can propagate simultaneously. The index of the wave quantities represents the number of the port.

The wave quantities can be determined from the voltage and current at the port. They are related to each other by Equ. 1, where V_1 and I_1 represent the voltage and current respectively at the port as depicted in Fig. 1b. Z_0 is an arbitrary reference impedance (often, but not necessarily always the characteristic impedance $Z_0 = Z_G = 50\Omega$ of the system).

$$a_1 = \frac{V_1 + I_1 Z_0}{2\sqrt{Z_0}} \qquad b_1 = \frac{V_1 - I_1 Z_0}{2\sqrt{Z_0}} \tag{1}$$

The wave quantities have the dimension of $\sqrt{W}$ (see the lecture "basic concepts" in this proceedings). This normalization is important for the conversation of energy. The power which is travelling towards the DUT can be calculated by $P_{inc} = |a|^2$. The reflected power by $|b|^2$. It is important to note that this definition is mainly used in the US – by the European notation, the incident power is usually calculated by $P_{inc} = 0.5|a|^2$. These conventions have no impact on the calculation of S-parameters and only need to be considered when the absolute power is of interest.

The reflection coefficient Γ represents the ratio of the incident wave to the reflected wave on a specific port. It is defined in Equ. 2.

$$\Gamma = \frac{b_1}{a_1} \tag{2}$$

By substitution with Equ. 1 we can find a relation between the complex impedance Z_L of a one

port and its complex reflection coefficient Γ (Equ. 3).

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{3}$$

There are some particular cases which are worth noting:

DUT	Z_L	Γ
Open circuit	∞	+1
Short circuit	0	-1
Matched load	Z_0	0
Load	$Z_0/2$	-1/3
Load	$2Z_0$	1/3

Table 1: Key numbers for the reflection coefficient.

2.2 2 Ports

Looking at electrical networks with two ports (e.g. attenuators, amplifiers), we find more quantities to be measured. Besides the reflection coefficients on either port, the transmission in forward and reverse direction can also be characterized. We now require the definition of scattering parameters (S-parameters) for two ports. The idea is to describe how the incident energy on one port is scattered by the network and exits through the other ports. All the possible signal paths through a two port are shown in Fig. 2. A two port has four complex and frequency dependent scattering parameters:

$$S_{11} = \frac{b_1}{a_1} \qquad S_{12} = \frac{b_1}{a_2} \qquad S_{21} = \frac{b_2}{a_1} \qquad S_{22} = \frac{b_2}{a_2} \tag{4}$$

S_{11} and S_{22} are equal to the reflection coefficient Γ on the respective ports – but **only** with the condition

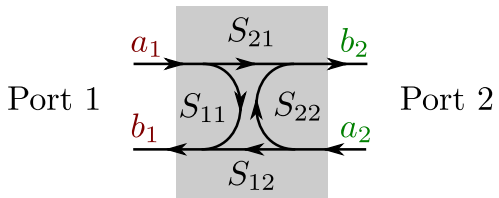


Fig. 2: All possible S-parameters of a two port network.

that the corresponding other port is terminated with the characteristic impedance. S_{21} and S_{12} are the forward and reverse transmission coefficient respectively. The first index of the S-parameter defines at which port the outgoing wave is observed, the second index defines at which port the network is excited. This leads to the counterintuitive appearing situation, that for forward transmission, the corresponding S-parameter is S_{21} and not S_{12}. The S-parameters are measured by exactly the same definition. The internal source of the network analyzer excites an incident wave on port one, namely a_1. Now b_1 and b_2, the outgoing waves from the DUT can be measured, which allows the determination of S_{11} and S_{21} (provided that port one and port two are terminated with their characteristic impedance).

It is very important to **terminate all ports of the DUT with the respective characteristic impedance**. In many situations this is Z_0 but there are cases where the characteristic impedance is different on port one and port two, such as a transformer with a turns ratio of two, leading to an impedance transformation by a factor of four. In this case the characteristic impedance would be for port one 50Ω and for port two 12.5Ω.

The termination prevents unwanted reflections and makes sure the DUT is only excited by a single incident wave. For practical S-parameter measurements this implies that any port of the DUT needs to be connected to a matched load corresponding to the characteristic impedance of this port. This rule includes in particular the port connected to the VNA measurement output, or in other works the generator impedance must also match the impedance of the DUT for the port under consideration. As a practical example one can not measure this in a straightforward manner – unless a special calibration procedure is used – with a 50Ω network analyzer and a DUT with 25Ω characteristic impedance. But modern VNAs permit, in a special calibration procedure, the modification of their characteristic impedance to any value (within a reasonable range from $> 5\Omega$ to $< 500\Omega$) and adapt it to the requirements of the DUT.

The S-parameters are an intrinsic property of the DUT and not a function of the incident power used in the measurement (condition of linearity). Obviously the S-parameter measured shall be independent of the instrumentation used for the measurement.

Once all n^2 S-parameter for a n port network are measured, the properties of this network can be described by a set of linear equations. For incident waves a_1 and a_2 of arbitrary phase and magnitude on a two port, the outgoing or scattered waves b_1 and b_2 can be determined using Equ. 5.

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

These equations can be written in matrix format, for convenience.

$$\begin{aligned} \vec{b} &= \mathbf{S}\vec{a} \\ \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &= \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \end{aligned} \tag{6}$$

The S-matrix is a linear model of the DUT. Its diagonal elements represent the reflection coefficients of each port. The remaining elements characterize all possible signal transmission paths between the ports. S-parameters are in general complex and a function of frequency. The set of linear equations, given by the S-matrix must be solved for a single frequency at a time. S-parameters are usually acquired for a certain frequency span at a number N of discrete frequency steps. With N data points, the system of equations has to be solved N - times. A discussion of general properties of the S-matrix is in the contribution "S-Matrix" in this proceedings.

3 Scalar network analysis

In a scalar network analyzer, only the amplitude of the signal is measured (reflected or transmitted) and the phase is not available. Consequently only the absolute value (the magnitude) of the complex S-parameters can be obtained. These kind of devices are generally less expensive than vector network analyzers. A very simple measurement setup as used more than 50 years ago is shown in Fig. 3.

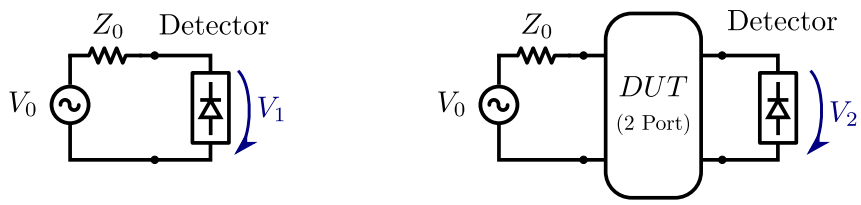


Fig. 3: A simple measurement setup for the scalar transmission coefficient ($|S_{21}|$).

The measurement is done twice, the first time (Fig. 3 left) without a DUT to measure the power of the incident signal (V_1). Then the DUT is inserted (Fig. 3 right), V_2 is measured and the magnitude of

the transmission coefficient can be calculated by

$$|S_{21}| \propto \frac{V_2}{V_1} \tag{7}$$

For obtaining a reading in decibel, a logarithmic amplifier is sometimes used following the detector. It has a logarithmic transfer function ($V_{out} = \log V_{in}$) and permits to show a large dynamic range on a dB scale. Also the mathematical operation of a division, required for normalization in Equ. 7, then transforms into the easier procedure of a subtraction.

The detector can be any kind of device converting the input RF signal into a DC voltage, which is "more or less"[1] proportional to RF power. There are basically three possibilities to achieve this:

Rectifier A very fast Schottky diode and a lowpass filter is used to convert the input RF signal to a DC voltage. Operating the diode in its square law region ($P_{in} < -10dBm$) results in an output voltage proportional to the RF power. Advantage: Cheap, fast response (depending on f_{max} of the output filter), Disadvantage: Commercially available RF power meters, based on Schottky diodes, can operate from -60 dBm (limited by tangential sensitivity), up to about 30 dBm (damage level). The nonlinearity of the output signal vs. input power is compensated by electronic means (look up table) in this kind of device. Coaxial RF Schottky detectors are usually limited to about 100 GHz, essentially determined by the connector available for this frequency range. Usually an input matching network has to be used to adjust the input impedance of the Schottky diode to $Z_0 = 50\Omega$. Waveguide type Schottky detectors can operate up to about one THz, however here the frequency limitation comes in via the size of the waveguides available. For higher frequencies above one THz, antenna type structures are used, a good overview is given in [3]. For special applications, Schottky detectors can be used up to optical frequencies.

Thermal measurement There are several types of detectors based on heating effects for the measurement of RF power. When using a Bolometer (Thermistor or Barretter) the high temperature coefficient of the thermal conductivity of certain metal or metal alloys is used. By a nonlinear calibration the dissipated heat is calculated from the DC wise measured change in temperature ΔT. Barretters use the positive temperature coefficient of metals like tungsten and platin. Thermistors consist of a metal oxide with a strong negative temperature coefficient. Another class of RF power meters based on heating is the thermo element, which takes advantage of the thermoelectrical coefficient of a junction between two different metals. A well known example is the SbBi junction which has a temperature coefficient of about $10^{-4}V/K$ which is one of the highest values available for this kind of detector. Even better values can be achieved using semiconductor - metal junctions, where thermoelectric coefficients of $250\mu V/K$ can be obtained. For further detail see [4].

Mixer Multiplying two sinusoidal signals with different frequencies results in signals at the sum and difference frequency at the multipliers output. This can technically be used to convert a band of high frequency signals to a much lower intermediate frequency (IF). Now all the measurements can be carried out at this IF. Mixers are known for their very high linearity and thus excellent dynamic input range. A mixer is the most important component of modern network analyzers, spectrum analyzers and virtually all RF communication systems.

For a more detailed discussion of RF detector systems is given in part 1 of this course: "Signal receiving techniques".

3.1 Levelling

If measurements are done over a wide frequency range the signal strength V_0 of the source has to stay constant. This usually requires an active feedback loop (**Levelling**), keeping V_0 constant and independent of frequency. Every feedback loop requires a measured process variable that has to be controlled to a

[1] With the term more or less, it should be stressed, that many detectors have a nonlinear relation between input power and output voltage.

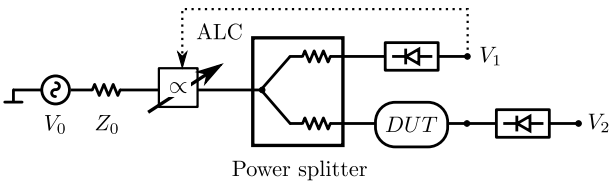


Fig. 4: Simplified circuit diagram of a typical automatic level control.

certain set point. This is the output signal level V_1. A resistive power divider may be used to provide this reference signal, while keeping its in- and outputs matched to $Z_0 = 50\Omega$ (Fig. 4). In this example, the test signal arriving at the DUT is reduced by 6 dB due to the insertion loss of the resistive power divider. The feedback loop ensures that the signal going to the DUT is constant and has a known power level over a wide frequency range.

For characterization of linear DUTs, only the ratio V_2/V_1 is required, independent of the absolute value of V_0. Theoretically one may carry out S-parameter measurements with an unleveled generator, but in practice the levelling has a large number of advantages, in particular for measurements on weakly nonlinear elements such as amplifiers.

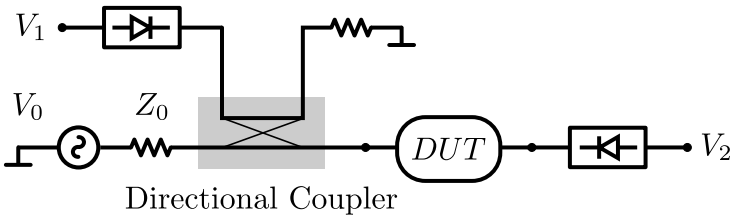


Fig. 5: Feedback loop of a typical automatic level control (ALC).

3.2 Directional couplers

By changing the resistive power divider to a directional coupler, the insertion loss can be reduced to much smaller values. This is shown in Fig. 5. V_1 is an attenuated (by the coupling factor) replica of the forward travelling wave which is only used for leveling and as reference. Typical directional couplers, used for this purpose, have a coupling of -20 dB and a transmission attenuation in the main branch of less than 0.3 dB. In contrast to the resistive power splitter, the directional coupler always has a more limited frequency range, which can lead to other problems.

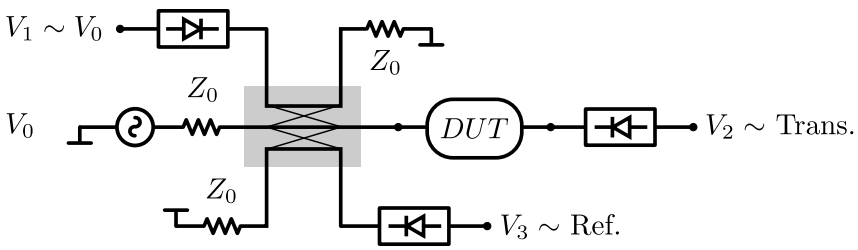


Fig. 6: Dual directional coupler in a network analyzer

Modern Network Analyzers (both scalar and vectorial versions) can measure the forward transmission coefficient, as well as the reflection coefficient of a DUT simultaneously, without the need to change any connections manually. Each port of the instrument has a dual directional coupler that provides a replica of the incident and reflected wave from the DUT. This is shown in Fig. 6. All those directional couplers in combination with switches and attenuators are commonly called a test set. Di-

rectional couplers are described in other lectures e.g. "S-Parameters" in this course. 20 or 30 years ago, network analyzers consisted of separate building blocks like S-parameters test set, frequency generator, display and controller unit. All these elements had to be connected by lots of external cables. Modern instruments have all those building blocks in a single frame, including advanced computer controls with digital data in- and output facilities.

From Fig. 6, the reflection and transmission coefficient are defined by:

$$|S_{11}| \propto \frac{V_3}{V_1} \qquad |S_{21}| \propto \frac{V_2}{V_1} \tag{8}$$

From the ratio of the reflected wave to the incident (S_{11}), quantities such as the standing wave ratio (SWR), reflection coefficient, impedance, admittance as well as return loss of the DUT can be determined. From the ratio of the transmitted wave to the incident (S_{21}), the gain or insertion loss, the transmission coefficient, the insertion phase and group delay of the DUT can be found.

One point to note is that scalar network analyzers generally modulate their test signal at around 20 kHz. This modulation is a simple on / off switch (AM modulation) and helps to prevent DC drifts of the measurement diodes. Instead of acquiring a DC signal which suffers from flicker noise and drifts, now only the 20 kHz component is used for further evaluation.

3.3 Transmission measurements using a Spectrum analyzer (tracking generator method)

The term $|S_{21}|$ can be measured by means of a spectrum analyzer. This instrument incorporates a sophisticated receiver part that is not much different to the receiver in a commercial network analyzer (superhetrodyne structure). At a fixed frequency we just need an external generator and use the spectrum analyzer as an detector with appropriate setting to read the power at this frequency. The spectrum analyzer is basically an advanced radio receiver operating in the superhetrodyne mode. For details see lecture on spectrum analyzers. The reading of the signal intensity from the screen can be done in linear or logarithmic format. Obviously the signal of the generator and the center frequency of the spectrum analyzer can be adjusted manually in a step by step mode in order to carry out a measurement of the DUT on several frequency points. Using an internal or externally synchronized tracking generator, this method is considerably simplified. Many commercial spectrum analyzer contain (as an option) this tracking generator, which is frequency locked but not phase locked to the sweep shown in the display.

4 VECTOR MEASUREMENTS

A vector network analyzer (VNA) is able to measure the magnitude **and phase** of a complex S-parameter. There are different possible hardware configuration for the implementation of such an instrument. Such as the six port reflectometer, certain RF bridge methods or superhetrodyne RF network analyzers.

4.1 Six port reflectometer

An interesting way to acquire the complex S_{11} with multiple detectors is the six port reflectometer. The historical origin of this is the RF measurement line. The RF measurement line is essentially a coaxial line with a slot in the longitudinal direction, where a small electric field probe can be moved along the direction of propagation. The output of this probe is connected to a diode and subsequently to a micro volt meter which permits the measurement of the field intensity of the standing wave pattern as a function of position. Having acquired voltage readings on four different positions and using a simple algorithm, the complex reflection coefficient of the DUT can be acquired. This procedure was the only available method for complex reflection coefficient measurements until about 1965. The six port reflectometer is essentially a static replica of this concept, using four diodes at fixed positions.

It was introduced in the early 1970s by G. Engen and C. Hoer as a simple and reasonably accurate power measurement setup. Shortly after its initial publication, the principle was expanded to the charac-

terization of voltage, current, impedance, and phase i.e. complex quantities required for vector network analysis. As this concept uses scalar diode detectors not operating in the mixer mode, its usable dynamic range is limited to about 60 dB.

The four diode detectors at ports 3 . . . 6 sample the magnitude of the RF signal at different positions along a transmission line. If the DUT is not matched perfectly, the reflected power will give rise to a standing wave pattern inside the measurement line which can be observed through the power meters. In case the DUT is perfectly matched, we have only a forward travelling wave and all the power meters give the same reading. With the 4 readings of the power meters, the same information is acquired as with the measurement line, using a moveable probe. For the practical implementation of this device, normally a

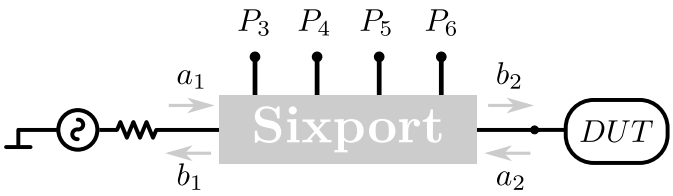


Fig. 7: Block diagram of a S_{11} measurement setup with a six port reflectometer.

transmission line is not used for practical reasons. A typical implementation can be seen in Fig. 8 on the right.

This circuit allows an easier calculation of forward and backward wave. It has been shown, that we need at least three power measurements under ideal conditions, to acquire the complex reflection coefficient S_{11}. The additional power sensor is used to increase the accuracy of the result [6].

It uses 3dB, 90 degree power splitters, which are readily available as microwave components. Their characteristics are shown in Fig. 8 on the left. A signal injected in port 1 of one power splitter is equally split in amplitude and shows up at port 3 and port 4. With the output signals 90 degrees out of phase. As a reminder, the S-parameters of a 90 degree hybrid are shown on the left of Fig. 8.

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & j & 0 \\ 1 & 0 & 0 & j \\ j & 0 & 0 & 1 \\ 0 & j & 1 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}$$

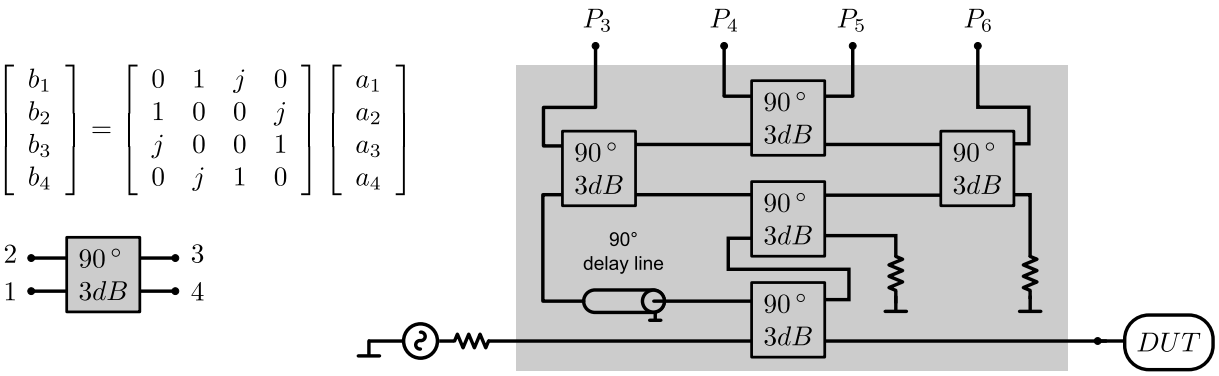


Fig. 8: The insides of a 6 port reflectometer.

When this kind of six port analyzers had first been proposed in the 1970s, the limited availability of computer power for the complex calculations was seen as a disadvantage of the concept. Nowadays this issue has been solved with the availability of powerful digital signal processors.

The concept has found its market share for network analysis (reflection and transmission measurements) and also communication receivers operating at very high frequencies (>300GHz) where conventional down-mixing gets difficult and power detectors are more easily available [6].

4.2 Bridge type methods

For a given single lumped element (R, C, L) with a known impedance Z, the 2-port S-parameters of the element can be calculated by Equ. 9.

$$S = \begin{pmatrix} \frac{Z}{2Z_0+Z} & \frac{2Z_0}{2Z_0+Z} \\ \frac{2Z_0}{2Z_0+Z} & \frac{Z}{2Z_0+Z} \end{pmatrix} \tag{9}$$

Obviously we can invert Equ. 9 and also obtain from some measured S-parameters (reflected or transmission) the unknown impedance Z of the DUT. As caveat this procedure is only applicable for lumped elements i.e. the maximum linear dimension is shorter than $\lambda/10$.

In analogy to bridge measurements for impedance determination (Wheatstone bridge) there is a very similar technique for measuring S_{21} or S_{11} respectively at a fixed frequency. The corresponding setup is shown in Fig. 9. For the transmission measurement, a calibrated attenuator and phase shifter are set such that the reading of the detector becomes minimal or zero (zero tuning). The S_{21} of the DUT then corresponds to the setting of the attenuator and the phase shifter. Basically, if the DUT and the reference

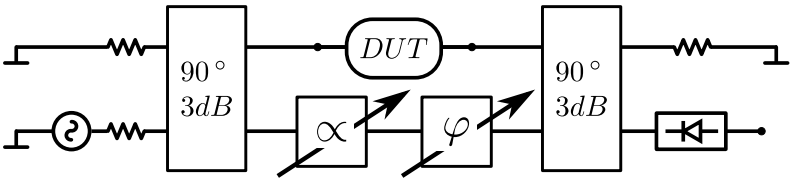


Fig. 9: Measurement bridge for determining S_{21}

path have the same complex transmission coefficient, both incident waves will cancel completely at the detector output of the 3 dB power combiner. The reading of the diode detector then vanishes and from the attenuation and phase shift settings of the reference path, the DUT parameters can be extracted.

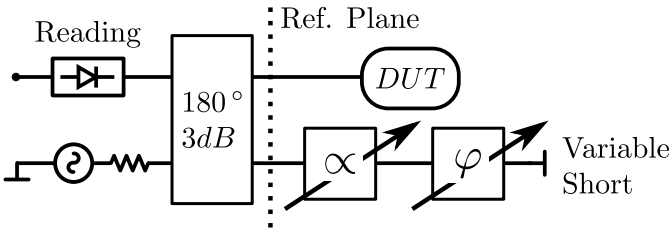


Fig. 10: Measurement bridge for determining S_{11}

The setup for reflection factor measurements using the cancellation method is shown in Fig. 10. It uses the same operating principle as described above for the transmission. The reference path consists of a short circuit, which reflects all of the microwave power. With a calibrated attenuator and phase shifter, the reflection factor of the reference path is matched to the reflection factor of the DUT.

Certain vector volt meter based instruments use this method in an automated way for determination of S_{21} and S_{11}.

5 Modern Vector Network Analyzers (VNA)

In the previous chapter, a short overview of the historical evolution of RF impedance and network measurement techniques has been given. Now it is just a small step, using previously defined tools, to discuss the operating concept of modern commercial vector network analyzers (superhet VNA).

A modern network analyzer contains the generator which produces the signal seen by the DUT. This signal is usually generated by a synthesizer type oscillator and is adjustable in very fine steps over a

large frequency range, in a programmable manner. Since all modern VNAs operate with analog or digital downmixing, the generation of a tracking LO frequency is also needed. This tracking LO is typically generated by PLL circuits and represents essentially a second oscillator following the main frequency with a specified offset.

Digital signal processing is used to adjust the observation bandwidth (IF bandwidth) over a very wide range (from 1 Hz to up to 20 MHz in certain instruments). The vectorial nature of the signal is preserved, both the phase and magnitude are acquired. More on the internal signal processing can be found in [1, 2]. Note that similar to the spectrum analyzer, the sweep time and resolution bandwidth can not be adjusted independently. A modern four port vector network analyzer is shown in Fig. 11.

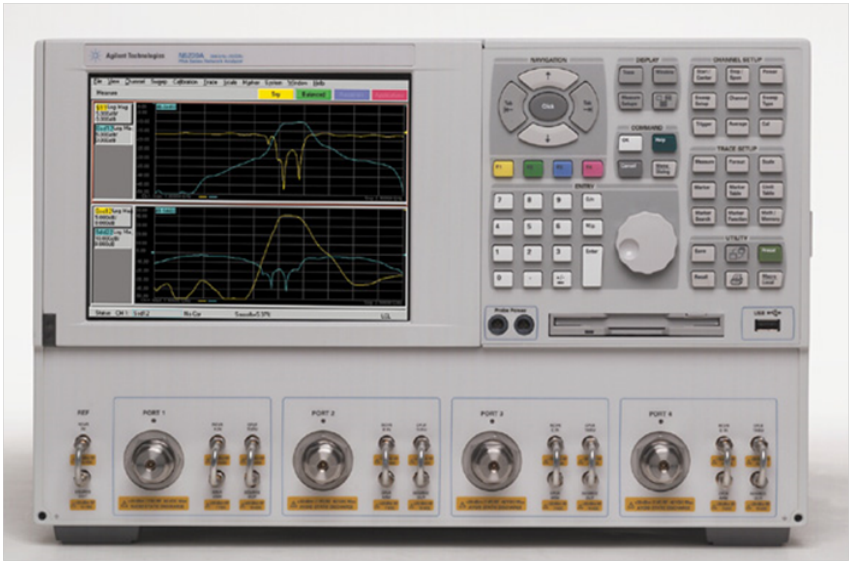


Fig. 11: A modern 4 port VNA.

Alhough complete network analysis of any N port can be done with a two port device, a four port unit can be very convenient for certain measurement tasks. It permits a quick analysis e.g. of a directional coupler or a three port circulator without the need for swapping cables.

5.1 Time domain transformation (synthetic pulse)

For any linear system, a data trace in the frequency domain can be converted to the time domain by a fast Fourier transformation[2] and vice versa. This is the basis of the synthetic pulse technique, available on many modern VNAs. It was commercially introduced by HP in the 1980s for network analyzer applications.

It renders the VNA even more versatile, allowing to display the impulse (gaussian) and also step response of the DUT and to carry out time domain reflectometry (TDR) measurements. Typical applications of this measurement are:

1. Finding and localizing discontinuities (faults) in transmission lines.
2. Separating the scattering properties of sections of complicated RF networks by time domain gating.
3. Echo cancellation (in multipath environments).
4. Synthetic pulse time domain reflectometry using waveguide modes has been used for obstacle detection in the LHC beampipe.

[2]More precisely: By a discrete fourier transformation. The FFT is just an optimized form of this, exploiting symmetries in a clever way. This saves computation time. However both algorithms will produce the same result for equal input data.

The only constraint of the applicability of the synthetic pulse measurement technique is, that the DUT must be a **linear** and **time invariant** (LTI) system.

A measurement example is shown in Fig. 12. A transmission line with a given length and some perturbation is connected to the calibrated VNA. The real part of the Fourier transformed reflection coefficient ($S_{11}(\omega)$) is shown vs time. The VNA permits the display of either the synthetic step (Fig. 12a) or impulse response (Fig. 12b). The step is simply obtained by integration over the impulse response.

The incident synthetic pulse is scattered from the discontinuity and also from the open end of the cable. The travel time for the pulse can be read on the horizontal axis on the time domain display. In this example we measure a delay of $t_d = 22ns$ until the open end of the cable becomes visible. This time accounts for the impulse travelling towards the open end **and back** thus the factor $1/2$ has to be taken into account when calculating line length.

$$l = \frac{c}{\sqrt{\varepsilon}} \cdot \frac{1}{2} t_D \tag{10}$$

The mechanical line length is given by Equ. 10. In this example we have $\epsilon = 2.3$ which returns a line length of $l = 2.2m$. The same method can be applied for obtaining the position of the irregularity (deformation, bad connector) of the cable. Nearly all VNAs with time domain option permit the designation of the velocity factor ($1/\sqrt{\varepsilon}$ for a homogeneously filled transmission line) and thus convert travel time to mechanical distance of the display.

Note that the step response shown in Fig. 12a returns the local reflection factor vs. time. Along the cable it amounts to $\Gamma = 0$, except for the position of the irregularity, indicating a well matched 50Ω transmission line. At the end we notice a positive step to $\Gamma = 1$ indicating an open circuit (see Tab. 1).

The fact that in the pulse response display (Fig. 12b) the reflected pulse from the open end does not reach unit amplitude is related to the impact of the cable attenuation of the transmission line used in this example (semi rigid coaxial cable around 2m long). The amplitude of this reflection from the open end indicates the attenuation over twice the electrical length of the cable at the center frequency ($f_{\text{max}} = 3GHz$, $f_{\text{center}} = 1.5GHz$).

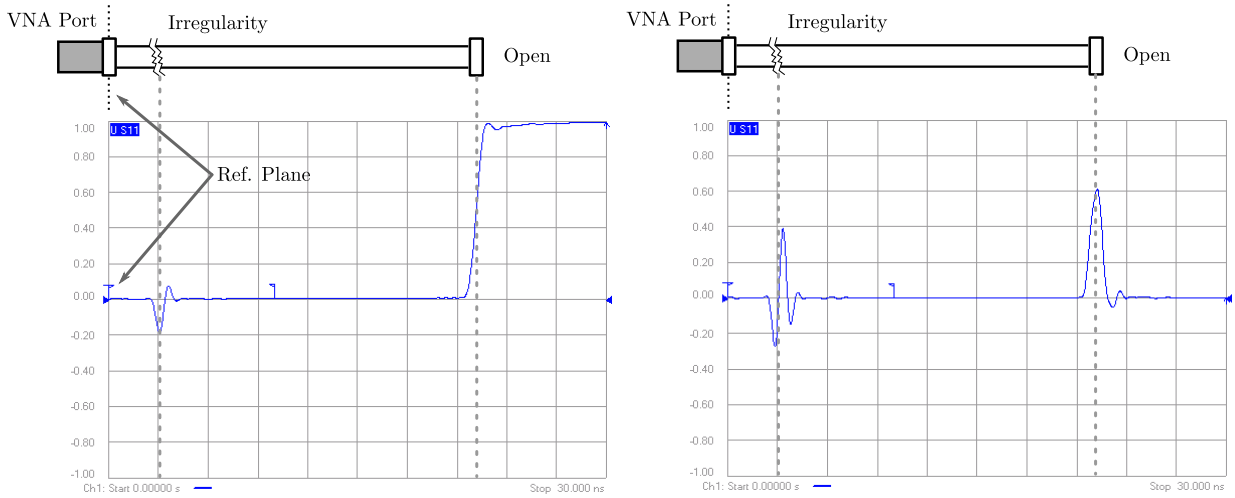
Time domain		Frequency Domain
T_{max} (time span)	$\leftrightarrow$	Δf (frequency resolution)
Δt (time resolution)	$\leftrightarrow$	f_{max} (frequency span)

Table 2: Most important characteristics of the FFT

For practical application of instrumentation using the synthetic pulse technology, certain basic properties of the discrete Fourier transform should be kept in mind. They are shown in Tab. 2. For example: A long cable needs to be tested. Due to the long time window required in order to assure that all multiple reflections have decayed to zero, a close spacing of the frequency samples has to be adjusted. This is simply related to the fact, that the length of the time value is related to $1/\Delta f$.

On the other hand if a bad connector or a cable damage needs to be located along the transmission line, a high resolution in time is required. Thus the VNA has to measure over a wide frequency span (f_{max}). Obviously we would like to always use both, a high frequency span and a close spacing of the samples in the frequency domain but there are practical limitations. Namely the number of data points available. Usually in modern instruments the number of data points available amounts to 60000 and depending on the application, compromises have to be accepted and a preference according to the criteria mentioned above given.

When using the time domain option of the vector-network-analyser there are two basic modes available: The “low-pass” mode and the “band-pass” mode. Both of these modes are briefly discussed in the following sections.



(a) Step response

(b) Impulse response

Fig. 12: Synthetic pulse measurement with a VNA. The measured frequency data is converted by an iDFT to the time domain. Now the synthetic impulse response of the cable can be plotted over time. The reflections of the incident pulse on an irregularity and at the end of the cable can be seen clearly. By measuring the delay in between, the position of the irregularity and the electrical length of the cable can be calculated.

5.2 "Low-pass mode"

In the low pass mode, the basic discrete Fourier transformation algorithm is applied. This returns certain constraints on the frequency domain measurement of the DUT (Fig. 13a). The iDFT demands that the starting frequency must always be 0 Hz (DC) and only equidistant frequency steps are allowed from there. Since most VNAs can not measure very low frequencies, the data points from DC to the minimum measurable frequency are extrapolated mathematically. Data points for negative frequencies are derived from the measured samples on the corresponding positive frequency by complex conjugation. Compared to the band-pass mode, this effectively doubles the number of data points available for the calculation of the time trace. For this particular symmetry, the discrete Fourier transformation returns a purely real valued time trace. The practical TDR measurement routine usually goes as follow:

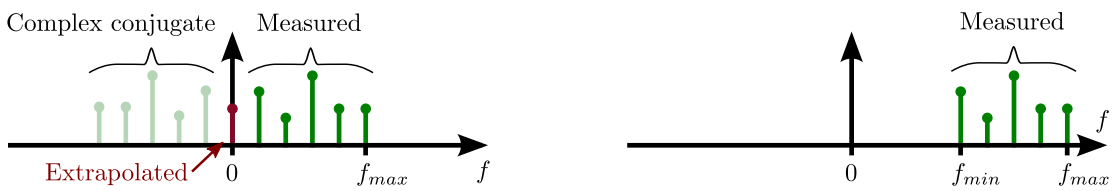
1. The DUT is connected, the port and type of measurement is selected (transmission or reflection).
2. The frequency range of interest and the number of datapoints is entered (this relates to the timedomain by Tab. 2)
3. After pushing the soft-key, "set frequency Low-pass[3]", the instrument will work out the exact frequencies where it has to sample the DUT.
4. Once the sampling points are defined, the VNA has to be calibrated (open, short, load for reflection measurements).

In the Low-pass mode, the trace appearing on the screen for TDA and TDT is basically equivalent to what is shown on a real time or sampling oscilloscope see Chapter 5.6.

5.3 "Band-pass mode"

In the band-pass mode (Fig. 13b) the spectral lines (frequency domain data points) need no longer be equidistant to DC but just within the frequency range of interest (e.g. from $f_{min} = 1.2GHz$ to $f_{max} =$

[3]This soft-key may appear with slightly different naming, depending on the definitions of the manufacturer.



(a) Low-pass mode (b) Band-pass mode

Fig. 13: Sampling of frequency points in the different operating modes.

$1.5GHz$). The start and stop frequency of the VNA can be chosen arbitrarily, which returns a high degree of flexibility and is especially suited for measuring devices with a limited range of operating frequencies (example: waveguide mode reflectometry).

As already mentioned, the band-pass mode is the equivalent of a narrow-band TDR (and also time domain transmission TDT) using the synthetic pulse technique. It permits the display of an impulse response only, since no extrapolated information on a DC component is available. In the measurements, the position and size of perturbations along a transmission line (including waveguides) can be nicely identified. But their characterization in terms of capacitive, inductive or resistive properties is possible but not straightforward ([5]).

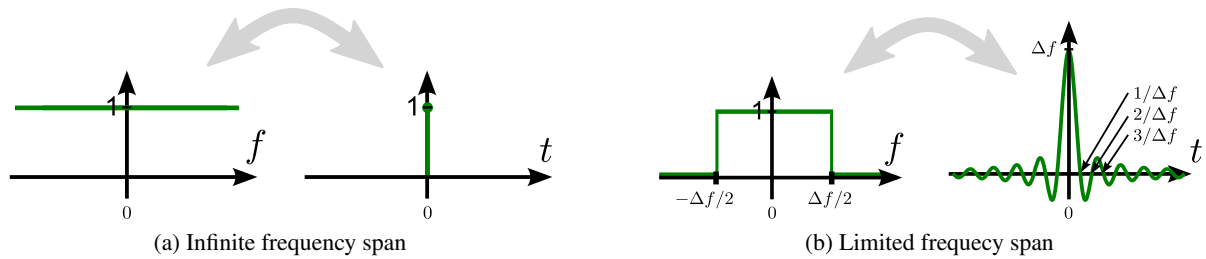
For the display there are several formats available which are different from the low-pass mode. In contrast to the low-pass mode, where we only have a real time domain trace (no imaginary part), we get complex time domain data in the band-pass mode. However, the magnitude of the response can be displayed. In band-pass mode, the corresponding time-domain response for the same bandwidth is twice as long as in low-pass mode.

In general, complex signals are obtained for the band-pass mode in the time domain. These complex signals are equivalent to the I and Q components (I = in phase and Q = quadrature) often found in complex signal treatment terminology. They can be directly displayed using soft-keys “real” or “imaginary” in the format menu. The real part is equivalent to what one would see on a fast scope i.e. an RF signal with a Gaussian envelope. The imaginary part looks very similar, except that the carrier (center frequency) is shifted by 90 degree. We can also display the magnitude of the measured signal, this equates to “modulus of the complex envelope $\sqrt{\mathrm{re}^2(t) + \mathrm{imag}^2(t)}$ of a carrier modulated signal” which is most frequently used. If a very high dynamic range in the time domain is needed, a logarithmic display format can be used. Note that the time domain mode can also be applied for CW excitation from the VNA but then to analyse a slowly time variant response of the DUT. More on the general properties and mathematical backgrounds of the low- and bandpass modes can be found in [2, 5, 7].

5.4 Windowing

As the VNA is only able to sample a limited frequency spectrum, starting at f_{min} and ending at f_{max}, we start off with a spectrum, clipped by a rectangular envelope. How this effects the time domain data, calculated through the iDFT can be seen in Fig. 14.

An infinite spectrum with constant density (shown in Fig. 14a), leads to a dirac pulse function in the time domain. The dirac pulse contains by definition all frequency components with equal power. In Fig. 14b the spectrum is limited, for example by the maximum measurable frequency of the VNA or by some user settings. This can be expressed by multiplication of the ideal spectrum with a rectangular function. The iDFT of a rectangular function with the width Δf, leads to a sinc function (also sometimes



(a) Infinite frequency span (b) Limited frequecy span

Fig. 14: The limited frequency span Δf of the VNA leads to distortions in the time domain synthetic pulse measurement. The ideal response is convoluted with a sinc function. Its characteristics depend on Δf.

denoted as si - function) in the t-domain. This relation is shown in Equ. 11 and graphically in Fig. 14.

$$\begin{aligned} \text{Frequency domain} \quad &\Longleftrightarrow \quad \text{Time domain} \\ \text{rect}\left(\frac{f}{\Delta f}\right) \quad &\Longleftrightarrow \quad \frac{\sin\left(\Delta f \pi t\right)}{\pi t} = \Delta f \cdot \text{sinc}\left(\Delta f \pi t\right) \end{aligned} \tag{11}$$

To alleviate the influence of the rectangular clipping of the spectrum, different kinds of weighting functions are in use. They smoothly reduce the amplitude of the spectrum at f_{min} and f_{max} (band-pass mode) but for f_{max} only in the low-pass mode. This helps to reduce strong side-lobes (ringing) in the time domain. However the price to be paid for the reduced sidelobes is a wider main lobe (pulse length), thus reducing the time resolution and the ability to distinguish between two closely spaced impulses. For the user, a reasonable trading between those two parameters has to be done, depending on the requirements of the particular measurement. The effect of some windows on main and sidelobes can be seen in Fig. 15.

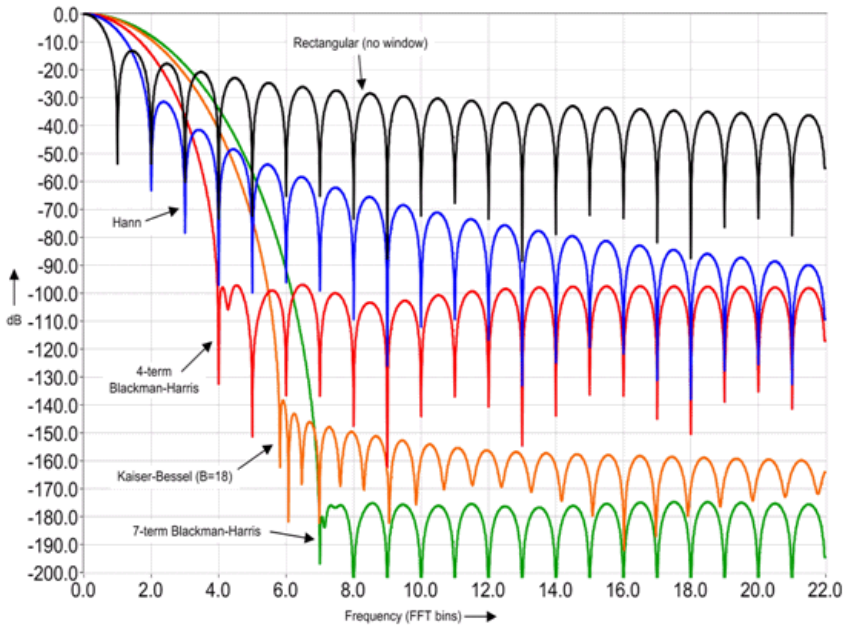


Fig. 15: Typical window functions to suppress strong-side lobes.

5.5 Gating

The gating function of the VNA can be used to eliminate undesired parts of the time domain signal, provided they are reasonably well situated in the time domain trace. As an example, the already mentioned cable, connecting the DUT to the VNA port is assumed to have an internal irregularity at a certain position.

By suitable selection of a time domain gate (marked in Fig. 16 at the range of about $t = 18ns$ to $t = 26ns$) we are able to analyze the selected pulse only and are not bothered by multiple reflections and other perturbations. Thus selecting and applying a gate, the time domain trace outside the gate is set to zero, eliminating undesired distortions. For transmission measurements, usually the **first** arriving pulse in the time domain is selected, thus suppressing the effect of all multiple reflections and related signals. For reflection measurements, the first or also later pulse in the time domain pulse may be selected. Once the gate has been applied, the transformation back to the frequency plane can be easily carried out and we see in the display the S-parameters (frequency domain) of the time gated signal.

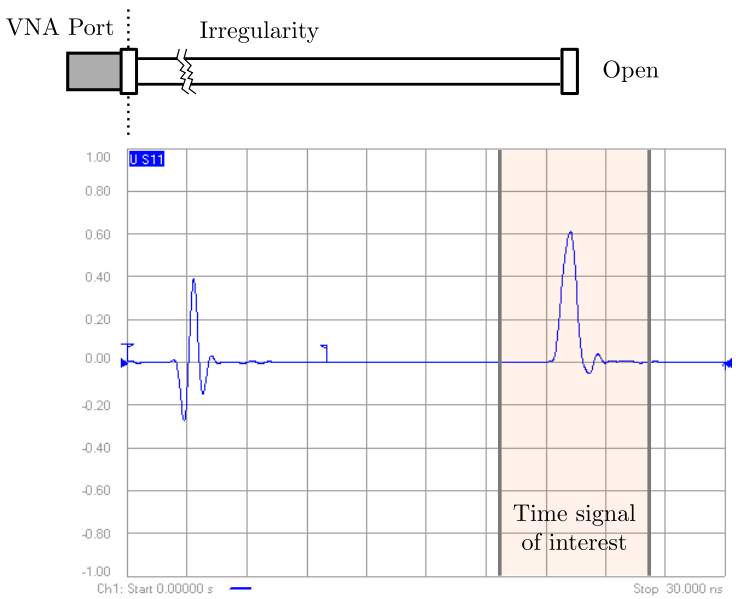


Fig. 16: Only the signal in a certain time window is of interest. After selection, the FFT of this window will be calculated. Here the real values of the synthetic impulse response are shown in a linear scale.

However, when using the gating function, it should be always kept in mind, that this operation is **non linear**. This implies, it may generate additional frequency components which were not present in the original signal. As a general practical rule, the gate should not cut into a signal trace different from zero.

A collection of measurement examples of simple DUTs are shown in Fig. 17. In all cases depicted there, the VNA is setup to measure the step response. The traces from top to bottom show:

1. Matched load ($Z = Z_C$). As is Γ is equal to zero, the response is zero over the whole time.
2. Moderate (resistive) mismatch ($Z = 2Z_C$ e.g. 100Ω in a 50Ω system). For the first 200 ps, only the well matched cable is seen, then the mismatch with its associated reflection coefficient.
3. Capacitor. The TDR sees the capacitor in the first moment as a short and terminates with an exponential function in an open circuit, as the capacitor is charged.
4. Inductor. For the TDR, the inductor appears at $t = 0$ as an open and terminates with an exponential function in a short, as the inductor starts conducting.

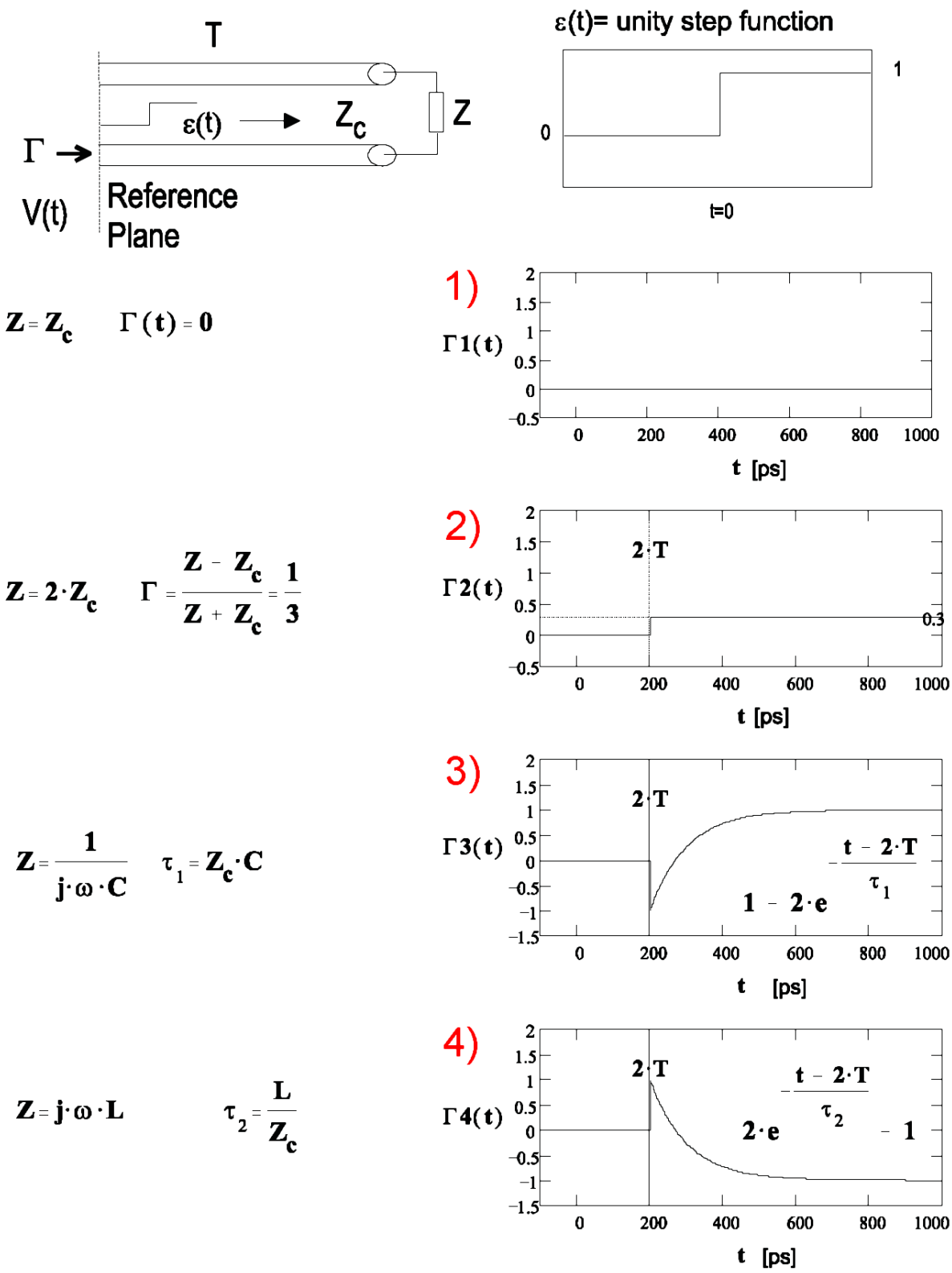


Fig. 17: Examples of an arbitary impedance, measured in TDR.

5.6 Comparison to real time domain measurements

There is a wide range of application for these synthetic pulse time domain techniques. A VNA in the time-domain low-pass step mode has a very similar range of applications as a sampling scope. However it must always be kept in mind, that carrying out a measurement in the frequency domain and then going via iDFT or similar into the time-domain implies strict linearity of the DUT. Thus a transient on a non-linear system such as the onset of oscillations on some microwave oscillator with active elements after turn-on of the supply voltage would not return meaningful results when using the synthetic pulse method. In other words, for highly nonlinear and time variant DUTs, real pulse measurements are still indispensable. Like in most of the airtraffic radar systems, where we have linear but time variant conditions.

The dynamic range of a typical sampling scope is limited to about 60 dB to 80 dB with a maximum input signal of 1 Volt and a noise floor around 0.1 mV to 1 mV (typical microwave scope). The VNA can easily go beyond 100 dB for the same maximum level of the input signal of about +10 dBm (some VNAs allow 20 dBm). Both instruments are using basically the same kind of detector, either a balanced mixer (4 diodes) or the sampling head (2 or 4 diodes), but the essential difference is the noise floor and the average signal power arriving at the receiver. In the case of the VNA we have a CW signal with bandwidth of a few Hz and thus can obtain with appropriate filtering a very good signal to noise ratio[4].

With the sampling scope we acquire data over a short time with a rather low repetition rate (typically around 100 kHz to several MHz) and all the thermal noise power is spread over the full frequency range (typically 20 GHz - 50 GHz bandwidth). With this low average signal power (around a micro-Watt) the signal spectral density is orders of magnitude lower than in the case of the VNA (it acquires signals continuously) and this finally explains the large difference in dynamic range (even without gain switching).

In the case of the VNA a CW signal with bandwidth of a few Hz is used, thus with appropriate filtering, a very good signal to noise ratio can be obtained. Also the VNA permits in contrast to the sampling scope, to tailor a wide range of spectra, which would be very tedious with a sampling scope.

A more detailed discussion about time domain reflectometry with vector network analyzers can be found in [7].

6 Calibration methods

Since even the hardware of a modern VNA is not perfect, we have to take into account that its internal source is not matched perfectly to 50Ω over the complete frequency range, also its internal directional couplers have a finite directivity, since there exists no ideal (infinite directivity) in reality. Furthermore we have to eliminate the effect of the frequency dependent attenuation of the coaxial cable, connecting the DUT to the ports of the VNA.

There are several calibration procedures to eliminate all or some of the deficiencies mentioned above. The easiest is the "response calibration" often used for transmission measurement, rarely for reflection. It consist essentially in connecting, for a transmission measurement, the two ends of the test cables to each other and store the amplitude and phase response for the complete frequency range in some memory. This reference trace may consist of several hundred or several thousand (according to the parameters selected) complex data points. DUT measurement data is normalized to the stored reference trace. In other words, for each frequency point, the measure of attenuation in dB of the test cable is subtracted (Equ. 12). Which is equivalent to a division in linear quantities. The phase is processed accordingly.

$$S_{21\text{resp. cal.}} = \frac{S_{21\text{measured}}}{S_{21\text{reference}}} \tag{12}$$

[4]Remember that thermal noise is proportional to measurement bandwidth. Its density at room temperature is -174 dBm/Hz

This very simple calibration procedure eliminates essentially the frequency dependent losses and phase transfer functions of the test cables only. But the cable and generator mismatch and finite directivity impact is still present.

A more sophisticated and widely applied calibration technique for reflection measurements is the open, short, match technique. This technique covers the three independent error sources mentioned above: Finite directivity, generator mismatch and cable transfer function.

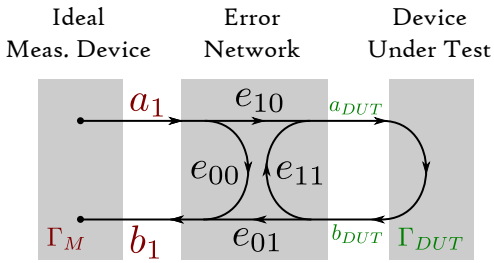


Fig. 18: Error model in a VNA. The parameters e_{xx} of the error network are determined by the calibration procedure and used to determine the real result (Γ_{DUT}) from the measured result (Γ_M).

The VNA uses a internal error model, shown in Fig. 18. All the measured raw data, the instrument "sees" (Γ_M) are affected by certain systematic errors, modelled via the four parameters of the error network (e_{10}, e_{00}, e_{01}, e_{11}). It is assumed that they are in general complex and frequency dependent with $e_{10} = e_{01}$. Since it is possible with suitable calibration methods, to make these parameters available, the true value of the DUT (Γ_{DUT}) can be calculated accordingly. In simple terms we need to carry out three independent measurements for each frequency point, in order to be able to solve three coupled equations with three complex unknowns.

These error terms represent the above mentioned effects as shown in Tab. 3.

Error term	Interpretation
e_{10}	reflection tracking
e_{00}	directivity
e_{11}	test port match

Table 3: Interpretation of the error terms.

We determine the unknowns of the error network by using three different, but known calibration DUTs. These calibration DUTs do not need to be perfect, only the electromagnetic properties need to be known with great precision. The tabulated complex and frequency dependent S-parameters of the calibration standards are provided by the manufacturer of this calibration hardware (they are often referred to as calibration kit), in electronic format.

They usually represent an open circuit, a short circuit and a match. This way the VNA can determine the frequency dependent error model, which may be altered if different test cables are used and correct further measurements accordingly. Now the "reference plane" is moved to the end of the test cables. Only the networks behind the reference plane will be taken into account for the measurement.

For example, shown in Fig. 19 is a measurement of S_{11} for a high quality 50Ω - termination with and without calibration. For an ideal termination, no reflection should be present. The effect of the calibration in this case improves the measurement by 20 dB. In case of a short, the non calibrated S_{11} response is typically a fraction of a dB up to a few dB below the 0 dB line (same for the open), after calibration this error reduces to range of several milli - dB.

So far we have discussed the "response calibration" and the complete one port calibration. In order

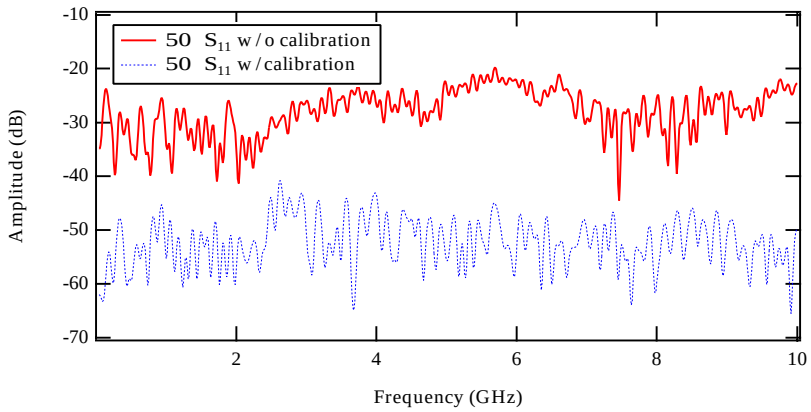


Fig. 19: S_{11} measurement of a 50Ω - termination with and without calibration. The calibration provides 20 dB improvement over this frequency band.

to carry out completely error corrected transmission measurements, we need the "full 2 port calibration". In this case, the error model must be expanded to include errors at the receiving port, requiring calibration of each port using 3 known loads in reflection. Also for transmission, we need 2 standards i.e. the response measurement and the "isolation measurement" which however may be omitted.

For measurements on devices with standard connectors, calibration standards such as a termination, open and short circuit are available (shown in Fig. 20a). As already mentioned for the calibration procedure for the reflection, the tabulated values, representing the electromagnetic properties of the calibration standards must be already programmed in the VNA or loaded in the instrument.

Obviously, the tabulated parameters of the calibration kit do not have an infinite frequency resolution. The instrument applies an interpolation procedure if the selected frequency points are not exactly at the tabulated values of the calibration kit.

The calibration technique described above, is widely used and well established in the measurement field. However it has one important disadvantage: It is tedious and time consuming, in particular for a calibration of a multiport VNA. Already for a full 2-port calibration, eight calibration measurements are needed in order to satisfy the requirements for an eight term error model. The manual connection and de-connection of the calibration standards is time consuming, prone to errors and may be boring. The situation becomes even worse when carrying out a full 4-port calibration (32 connections and de-connections of standards). For this reason, the electronic calibration kit method has been invented and became very popular. In this case, each port is connected via a cable to the electronic calibration box (shown in Fig. 20b), which switches the different standards automatically by communicating with the VNA. With this method, a full 4-port calibration takes less than a minute. Again, like in the manual calibration method, the standards do not need to be perfect but well known, reproducible (switching) and stable. In this chapter, only a short introduction to calibration methods has been given and more details can be found in [1, 2].

7 Nonlinear analysis

7.1 1 dB compression point

A single tone source is connected to the input of an amplifier and its intensity gradually increased vs. time. Monitoring the output of this amplifier, we notice a proportional dependency between in- and output power for small signal levels. This proportionality is referred to as linear gain factor. For higher input signal levels, this will not hold anymore since the amplifier is not a perfectly linear system. A fraction of the output power will appear at different frequencies, which are harmonics of the input signal. These

(a) Manual (open, short, match) (b) Electronic

Fig. 20: Typical calibration kits for a VNA.

are typically the second and third harmonic and the distortion is referred to as harmonic distortion. This is called harmonic distortion. In parallel we can observe a compression of the gain for the fundamental signal. The actual gain falls off below the small signal gain (Fig. 21). When this deviation amounts to 1 dB, we have reached the 1 dB compression point.

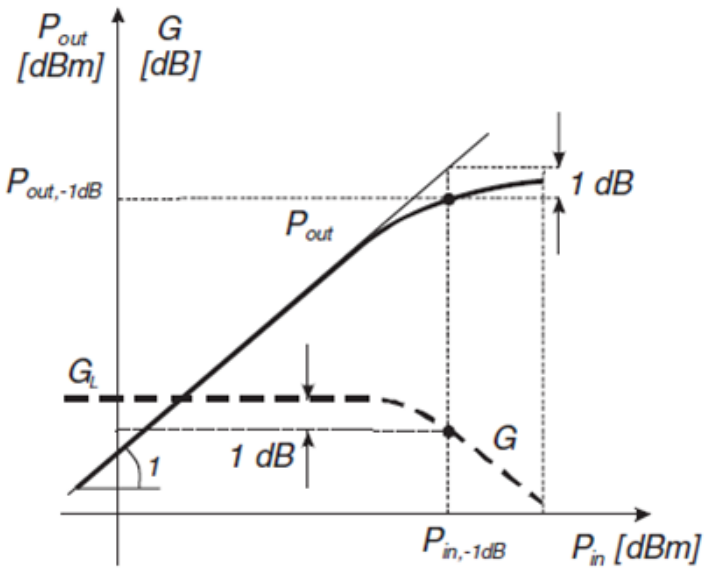


Fig. 21: Definition of the 1 dB compression point for an amplifier: Input power, where the output power falls below 1 dB from its (linearly) predicted value.

This compression is an important figure of merit, used to characterize the linearity of a system, in particular the performance of small signal and power amplifiers. This 1 dB compression point can easily be measured with most VNAs by setting it to CW (continuous wave) mode i.e. choosing a single frequency and running a power sweep. In the power sweep mode, the instrument displays exactly the situation shown in Fig. 21.

7.2 X-parameters

For the development of active electronic systems such as mixers, small signal and power amplifiers, nonlinear analysis is becoming an increasingly important tool. Apart from the general definition "nonlinear analysis", the term X-parameters has recently become quite popular.

X-parameters[5] are a superset of classical S-parameters and provide the necessary mathematical framework to measure, model and simulate nonlinear systems [8]. They can show particular characteristics of nonlinear systems, like the generation of harmonics or intermodulation distortion, which S-parameters are – strictly speaking – not able to do. For small input signals they converge to classical small signal S-parameters.

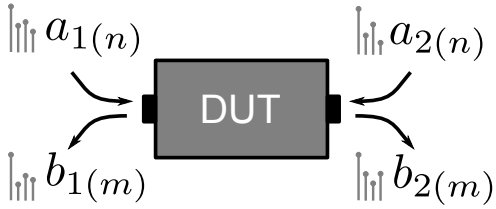


Fig. 22: Power waves on a two port.

X-parameters characterize the amplitudes and relative phase of harmonics generated by components under large input power levels to all ports. They are able to correctly characterize impedance mismatches and frequency mixing behavior to allow accurate simulation of cascaded nonlinear X-parameter blocks, such as amplifiers and mixers.

Transmission S-parameters are essentially complex but linear gain factors. As non - linearities give rise to all kinds of harmonic generation, non - linear analysis and the related X-parameters are gaining increasing interest. This is also linked to the fact that with the availability of cheap computing power, non - linear analysis can be carried out fairly easily and in a semi - automatic manner on modern instruments. It should be known that parallel, non - linear circuit and field simulations are becoming available and finding an increased number of applications.

While S-parameters are a **linear** describing function in the sense of

$$b_1, b_2, \ldots = f\left(a_1, a_2, \ldots\right) \tag{13}$$

Nonlinear analysis is an expansion of this linear concept. A second subscript i is introduced, defining the index of the i-th harmonic in brackets.

$$b_{1(i)}, b_{2(i)}, \ldots = f\left(a_{1(i=1\ldots n)}, a_{2(i=1\ldots n)}, \ldots\right) \tag{14}$$

In a practical measurement, we need to excite the DUT with a realistic[6] test signal (containing several harmonics) at each port consecutively. In every instance, the magnitude and phase of the fundamental and of several harmonics is measured on all ports of the DUT (shown in Fig. 22).

As this complicates the measurement (and calibration) routine, there are not many "large signal network analyzers", capable of nonlinear analysis, available these days. This will gradually change in the future, since nonlinear analysis is both, for measurement and simulation, a vital tool, for proper system design.

[5] X-parameters is a registered trademark of Agilent Technologies.

[6] Meaning, the test signal should look like the signal the DUT will see in its actual application.

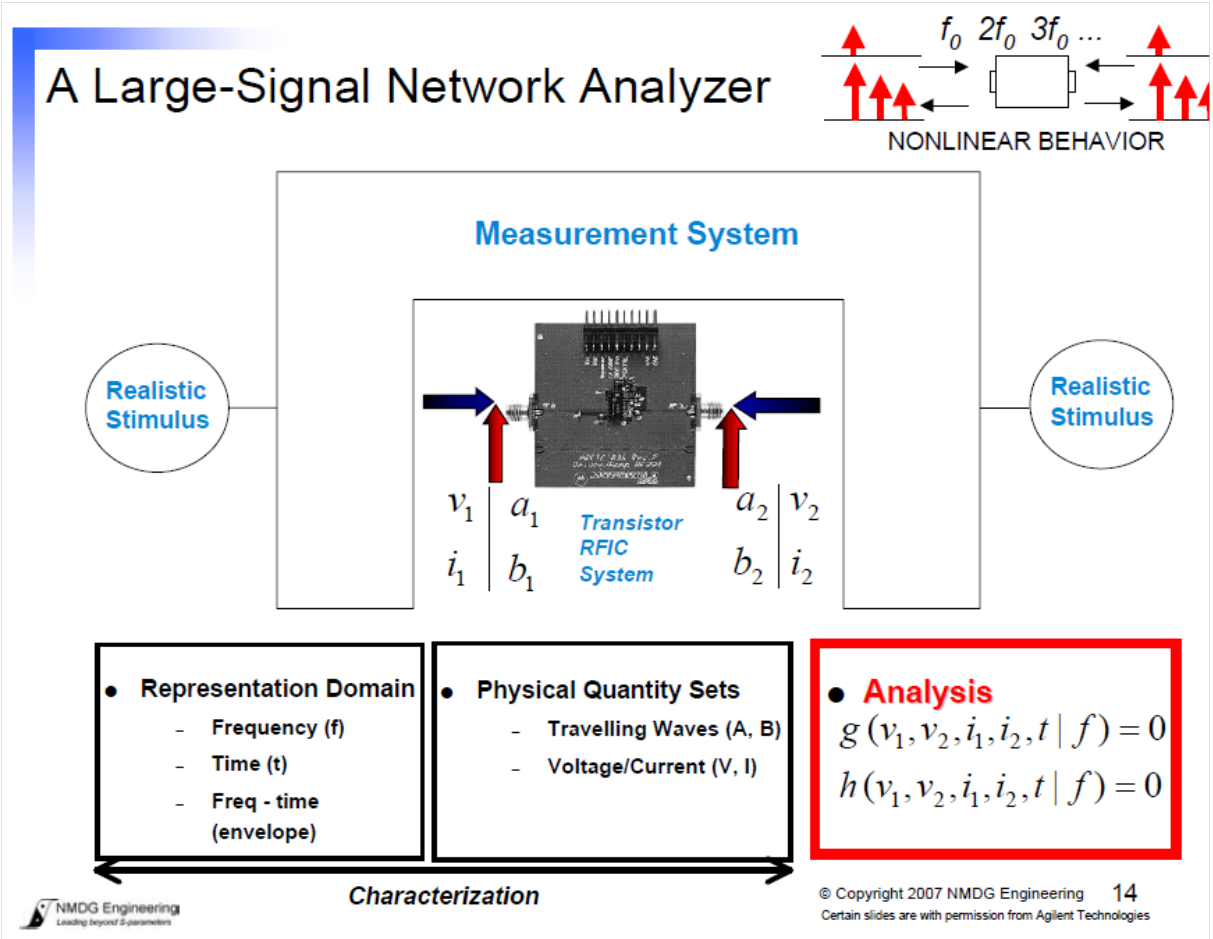


Fig. 23: Block diagram of a large-signal network analyzer (for non-linear DUTs).

8 Summary

- Network analyzer technology has gone through an impressive evolution over the last 40 years
- The availability of fast ADCs, powerful digital signal processors and cheap computer power made the VNA a very versatile and precise instrument with comfortable data output and control functions.
- Vector network analyzers are nowadays available up to 300 GHz (with external frequency converter units). Even for optical frequencies, vector signal analysis concepts have been investigated.
- Scalar network analyzers can reach frequencies of 1 THz and higher in a few cases with a different architecture (optical network analyzers).
- For frequencies up to 10 GHz, scalar network analyzers are a lower cost and moderate performance alternative to VNA's, at the expense of reduced performance.
- Spectrum- and VNA-functions are sometimes available in the same instrument (spectrum network analyzers).

References

[1] M. Hiebel, Fundamentals of Vector Network Analysis.

[2] Agilent AN 1287-1, Understanding the Fundamental Principles of Vector Network Analysis

[3] G. Beziuk, A. Grobelny, E. F. Plinski, J. S. Witkowski, THz Optical Mixer

[4] M. Thumm, W. Wiesbeck, S. Kern, Hochfrequenzmesstechnik, Verfahren und Messsysteme

[5] Anritsu Application note, Time Domain Measurements Using Vector Network Analyzers

[6] A. Koelpin, G. Vinci, B. Laemmle, D. Kissinger, R. Weigel, The Six-Port in Modern Society, IEEE microwave magazine, Dec 2010

[7] Agilent AN 1287-12, Time Domain Analysis Using a Network Analyzer.

[8] J. Verspecht, D. Root, Polyharmonic Distortion Modeling.

[9] G. H. Bryant, Principles of Microwave Measurements.

[10] F. Caspers, Proc. CERN Accelerator School, RF Engineering for Particle Accelerators, Oxford, UK (1991).