

RF Measurement Concepts

F. Caspers¹ and P. Kowina²

¹CERN, Geneva, Switzerland

²GSI, Darmstadt, Germany

Abstract

For the characterization of components, systems and signals in the radiofrequency (RF) and microwave ranges, several dedicated instruments are in use. In this article the fundamentals of the RF signal techniques are discussed. The key element in these front ends is the Schottky diode which can be used either as a RF mixer or as a single sampler. The spectrum analyser has become an absolutely indispensable tool for RF signal analysis. Here the front end is the RF mixer as the RF section of modern spectrum analysers has a rather complex architecture. The reasons for this complexity and certain working principles as well as limitations are discussed. In addition, an overview of the development of scalar and vector signal analysers is given. For the determination of the noise temperature of a one-port and the noise figure of a two-port, basic concepts and relations are shown as well as a brief discussion of commonly used noise-measurement techniques. In a further part of this article the operating principles of network analysers are shown. A distinction can be made between scalar and vector network analysers 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 analysers require sophisticated calibration procedures, which are now indispensable for many measurement applications. Non-linear network analysis completes tests of most characteristic amplifier features. The Smith chart is a very valuable and important tool that facilitates interpretation of S-parameter measurements. The last part of this article gives a brief overview of how to use the chart. Its definition as well as an introduction on how to navigate inside the chart are illustrated. Useful examples show the broad possibilities for use of the chart in a variety of applications. This article is a compilation of previous CAS proceedings where information being in the scope of the intermediate-level RF course was extracted.

1 INTRODUCTION TO SIGNAL RECEIVING TECHNIQUES

In the early days of radiofrequency (RF) engineering the available instrumentation for measurements was rather limited. Besides elements acting on the heat developed by RF power (bimetal contacts and resistors with very high temperature coefficient) only point/contact diodes, and to some extent vacuum tubes, were available as signal detectors. For several decades the slotted measurement line, see Section 10.1, was the most used instrument for measuring impedances and complex reflection coefficients. Around 1960 the tedious work with such coaxial and waveguide measurement lines became considerably simplified with the availability of the vector network analyser. At the same time the first sampling oscilloscopes with 1 GHz bandwidth arrived on the market. This was possible due to progress in solid-state (semiconductor) technology and advances in microwave elements (microstrip lines). Reliable, stable and easily controllable microwave sources are the backbone of spectrum and network analysers as well as

Table 1: Overview of dB key values and their conversion into power and voltage ratios

	Power ratio	Voltage ratio
−20 dB	0.01	0.1
−10 dB	0.1	0.32
−3 dB	0.50	0.71
−1 dB	0.74	0.89
0 dB	1	1
1 dB	1.26	1.12
3 dB	2.00	1.41
10 dB	10	3.16
20 dB	100	10
$n \cdot 10$ dB	10^n	$10^{n/2}$

sensitive (low-noise) receivers. The following sections will concentrate on signal receiving devices such as spectrum analysers. An overview of network analysis will be given later on in Section 7.

2 BASIC DEFINITIONS, ELEMENTS AND CONCEPTS

Before discussing several measurement devices, a brief overview of the most important components in such devices and some basic concepts are presented.

2.1 Decibel

Since the unit decibel (dB) is frequently used in RF engineering, a short introduction and definition of terms is given here. The decibel is the unit used to express relative differences in signal power. It is expressed as the base-10 logarithm of the ratio of the powers of two signals:

$$P \text{ [dB]} = 10 \cdot \log(P/P_0). \quad (1)$$

It is also common to express the signal amplitude in dB. Since power is proportional to the square of a signal's amplitude, the voltage in dB is expressed as follows:

$$V \text{ [dB]} = 20 \cdot \log(V/V_0). \quad (2)$$

In Eqs. (1) and (2), P_0 and V_0 are the reference power and voltage, respectively. A given value in dB is the same for power ratios as for voltage ratios. It is important to note that there are no 'power dB' or 'voltage dB' as dB values always express a ratio. Conversely, the absolute power and voltage can be obtained from dB values by

$$P = P_0 \cdot 10^{\frac{P \text{ [dB]}}{10}}, \quad (3)$$

$$V = V_0 \cdot 10^{\frac{V \text{ [dB]}}{20}}. \quad (4)$$

The advantage of logarithms as the unit of measurement is twofold:

- i) signal power tends to span several orders of magnitude; and
- ii) signal attenuation losses and gains can be expressed in terms of subtraction and addition.

Table 1 helps to indicate the order of magnitude associated with dB.

Frequently dB values are expressed using a special reference level and not SI units. Strictly speaking, the reference value should be included in parentheses when giving a dB value, e.g. +3 dB (1 W)

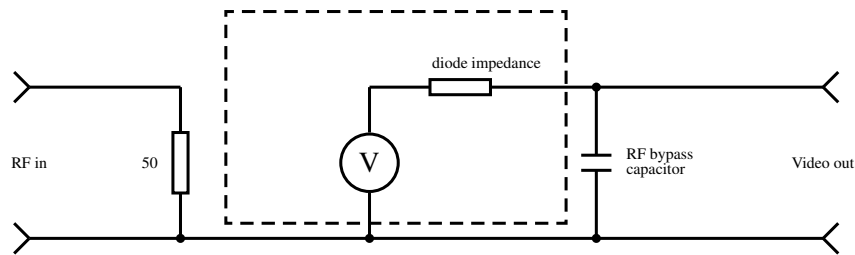


Fig. 1: The equivalent circuit of a diode



Fig. 2: A commonly used Schottky diode. The RF input of this detector diode is on the left and the video output on the right (figure courtesy Agilent).

indicates 3 dB at $P_0 = 1$ W, thus 2 W. However, it is more common to add some typical reference values as letters after the unit, for instance, dBm defines dB using a reference level of $P_0 = 1$ mW. Thus, 0 dBm correspond to -30 dBW, where dBW indicates a reference level of $P_0 = 1$ W. Often a reference impedance of 50Ω is assumed. Other common units are:

- i) dBmV for small voltages with $V_0 = 1$ mV; and
- ii) dBmV/m for the electric field strength radiated from an antenna with reference field strength $E_0 = 1$ mV/m.

2.2 The RF diode

One of the most important elements inside all sophisticated measurement devices is the fast RF diode or Schottky diode. The basic metal–semiconductor junction has an intrinsically very fast switching time of well below a picosecond, provided that the geometric size and hence the junction capacitance of the diode is small enough. However, this unavoidable and voltage-dependent junction capacity will lead to limitations of the maximum operating frequency. The equivalent circuit of such a diode is depicted in Fig. 1 and an example of a commonly used Schottky diode can be seen in Fig. 2. One of the most important properties of any diode is its characteristic, which is the relation of current as a function of voltage [1]. This relation is depicted graphically for two diodes in Fig. 3. As can be seen, the diode is a non-ideal commutator (on the contrary to that shown in Fig. 4) for small signals. Note that it is not possible to apply big signals, since this kind of diode would burn out. Although there exist rather large power versions of Schottky diodes which can stand more than 9 kV and several tens of amperes, they are not suitable in microwave applications due to their large junction capacity.

The region where the output voltage is proportional to the input power is called the square-law region (Fig. 5). In this region the input power is proportional to the square of the input voltage and the

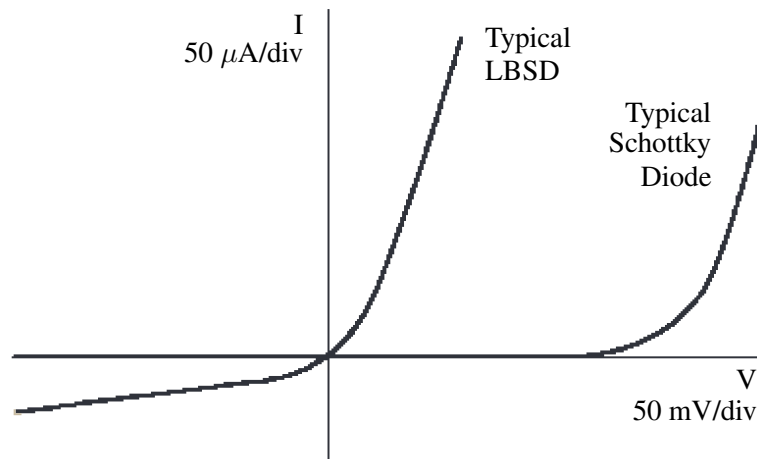


Fig. 3: Current as a function of voltage for different diode types (LBSD = low barrier Schottky diode)

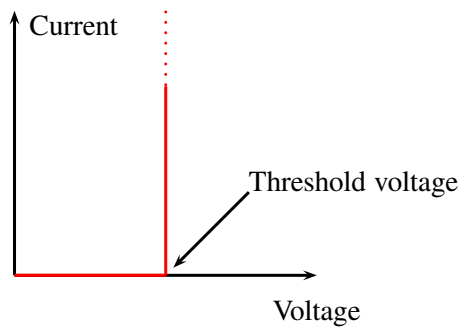


Fig. 4: The current–voltage relation of an ideal commutator with threshold voltage

output signal is proportional to the input power, hence the name square-law region.

The transition between the linear region and the square-law region is typically between -10 and -20 dB (Fig. 5). For a more precise description, see [2].

There are fundamental limitations when using diodes as detectors. The output signal of a diode (essentially DC or modulated DC if the RF is amplitude modulated) does not contain phase information. In addition, the sensitivity of a diode restricts the input level range to about -60 dBm at best, which is not sufficient for many applications.

The minimum detectable power level of a RF diode is specified by the ‘tangential sensitivity’, which typically amounts to -50 to -55 dBm for 10 MHz video bandwidth at the detector output [3].

To avoid these limitations, another method of operating such diodes is needed. This method is described in the next section.

2.3 Mixer

For the detection of very small RF signals a device that has a linear response over the full range (from 0 dBm (= 1 mW) down to thermal noise = -174 dBm/Hz = $4 \cdot 10^{-21}$ W/Hz) is preferred. A RF mixer provides these features using one, two or four diodes in different configurations (Fig. 6). A mixer is

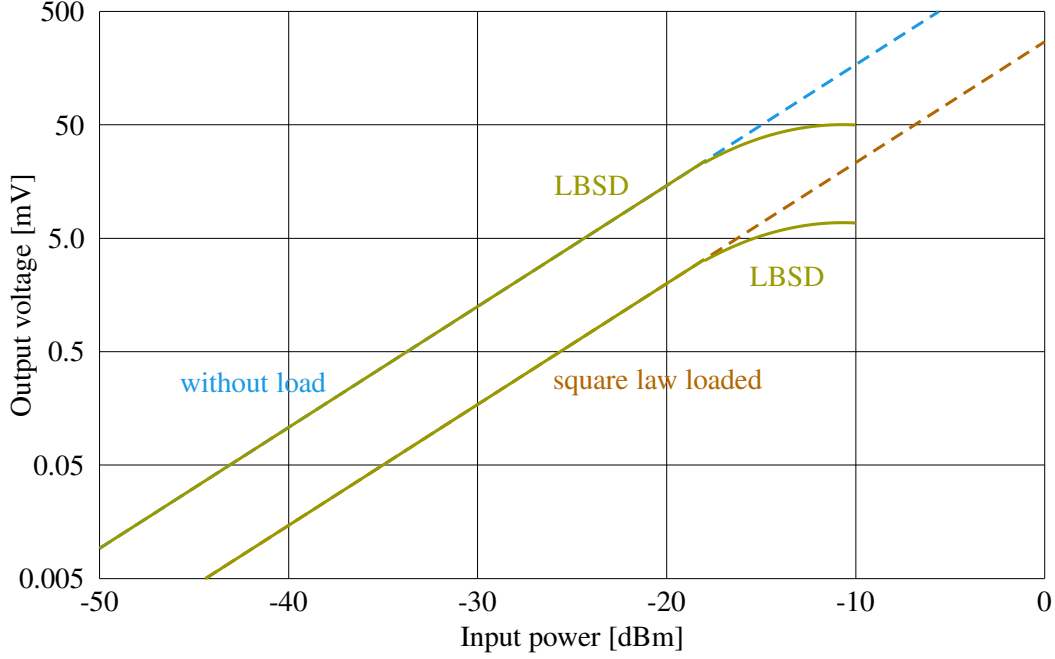


Fig. 5: Relation between input power and output voltage

essentially a multiplier with a very high dynamic range implementing the function

$$f_1(t) \cdot f_2(t) \text{ with } f_1(t) = \text{RF signal} \text{ and } f_2(t) = \text{LO signal} \quad (5)$$

or, more explicitly for two signals with amplitude a_i and frequency f_i ($i = 1, 2$),

$$a_1 \cos(2\pi f_1 t + \varphi) \cdot a_2 \cos(2\pi f_2 t) = \frac{1}{2} a_1 a_2 [\cos((f_1 + f_2)t + \varphi) + \cos((f_1 - f_2)t + \varphi)]. \quad (6)$$

Thus, we obtain a response at the intermediate-frequency (IF) port that is at the sum and difference frequencies of the local oscillator (LO = f_1) and RF (= f_2) signals.

Examples of different mixer configurations are shown in Fig. 6.

As can be seen from Fig. 6, the mixer uses diodes to multiply the two ingoing signals. These diodes function as a switch, opening different circuits with the frequency of the LO signal (Fig. 7).

The response of a mixer in the time domain is depicted in Fig. 8. The output signal is always in the ‘linear range’ provided that the mixer is not in saturation with respect to the RF input signal. Note that for the LO signal the mixer should always be in saturation to make sure that the diodes work as a nearly ideal switch. The phase of the RF signal is conserved in the output signal available from the RF output.

2.4 Amplifier

A linear amplifier augments the input signal by a factor which is usually indicated in decibels (dB). The ratio between the output and the input signals is called the transfer function and its magnitude – the voltage gain G – is measured in dB and given as

$$G[\text{dB}] = 20 \cdot \frac{V_{\text{RFout}}}{V_{\text{RFin}}} \text{ or } \frac{V_{\text{RFout}}}{V_{\text{RFin}}} = 20 \cdot \log G[\text{lin}]. \quad (7)$$

The circuit symbol of an amplifier is shown in Fig. 9 together with its S-matrix.

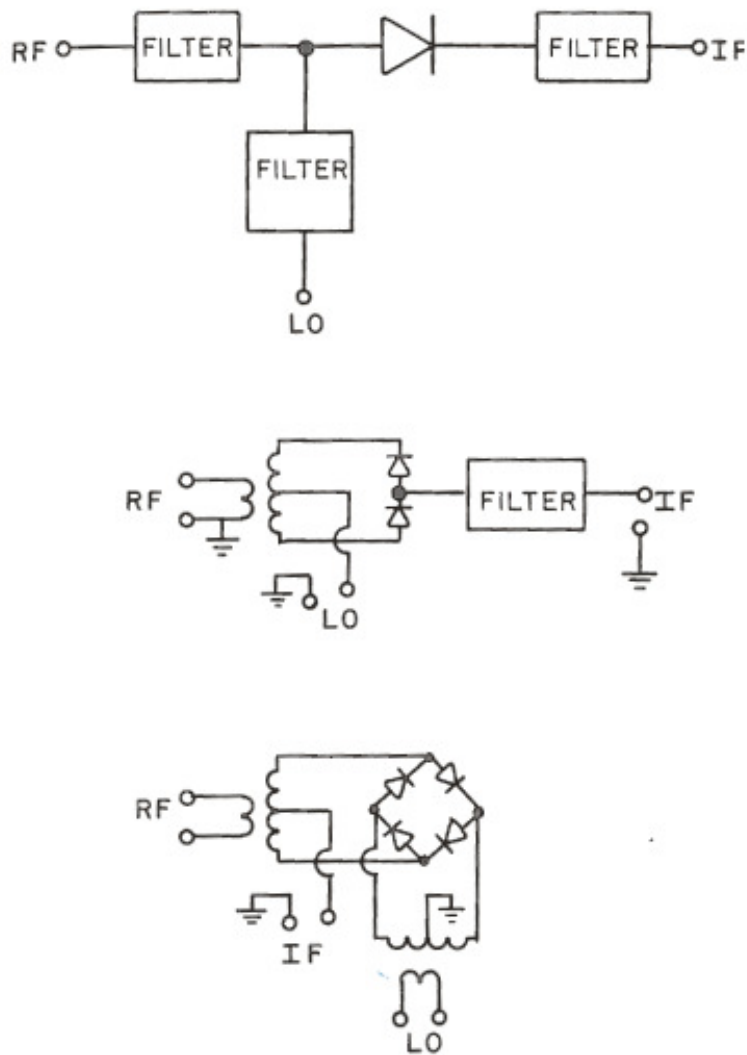


Fig. 6: Examples of different mixer configurations

The bandwidth of an amplifier specifies the frequency range where it is usually operated, see Fig. 10. This frequency range is defined by the -3 dB points¹ with respect to its maximum or nominal transmission gain.

In an ideal amplifier the output signal would be proportional to the input signal. However, a real amplifier is non-linear, such that for larger signals the transfer characteristic deviates from its linear properties valid for small-signal amplification. When increasing the output power of an amplifier, a point is reached where the small-signal gain becomes reduced by 1 dB (Fig. 11). This output power level defines the 1 dB compression point, which is an important measure of quality for any amplifier (low level as well as high power).

The transfer characteristic of an amplifier can be described in terms which are commonly used for RF engineering, i.e. the S-matrix, see Section 7. As implicitly contained in the S-matrix, the amplitude and phase information of any spectral component are preserved when passing through an ideal amplifier. For a real amplifier the element $G = S_{21}$ (transmission from port 1 to port 2) is not a constant but a complex function of frequency. Also, the elements S_{11} and S_{22} are not 0 in reality.

¹The -3 dB points are the points left and right of a reference value (e.g. a local maximum of a curve) that are 3 dB lower than the reference.

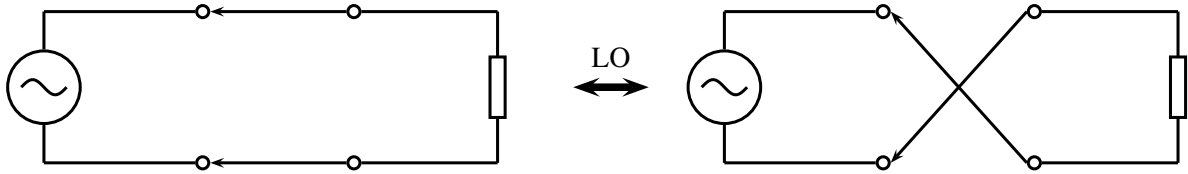


Fig. 7: Two circuit configurations interchanging with the frequency of the LO where the switches represent the diodes.

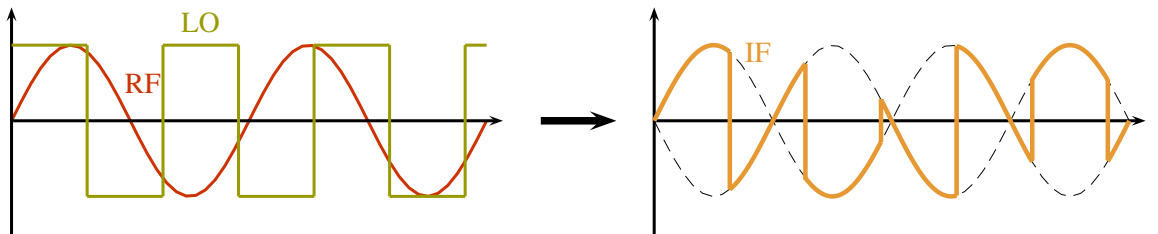


Fig. 8: Time-domain response of a mixer

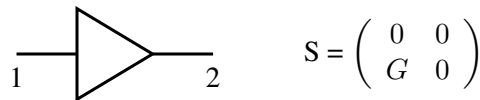


Fig. 9: Circuit symbol and S-matrix of an ideal amplifier

2.5 Interception points of non-linear devices

Important characteristics of non-linear devices are the interception points. Here only a brief overview will be given. For further information, the reader is referred to [4].

One of the most relevant interception points is the interception point of third order (IP3 point). Its importance derives from its straightforward determination, plotting the input versus the output powers in logarithmic scale (Fig. 11). The IP3 point is usually not measured directly, but extrapolated from measurement data at much smaller power levels in order to avoid overload and damage of the device under test (DUT). If two signals ($f_1, f_2 > f_1$) which are closely spaced by Δf in frequency are simultaneously applied to the DUT, the intermodulation products appear at $+\Delta f$ above f_2 and at $-\Delta f$ below f_1 . This method is usually called third-order intermodulation (TOI). An example of automatized TOI measurement is shown in Fig. 12.

The transfer functions of weakly non-linear devices can be approximated by Taylor expansion. Using n higher order terms on one hand and plotting them together with an ideal linear device in logarithmic scale leads to two lines with different slopes ($x^n \xrightarrow{\log} n \cdot \log x$). Their intersection point is

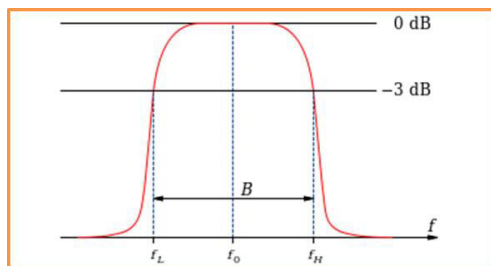


Fig. 10: Definition of the bandwidth

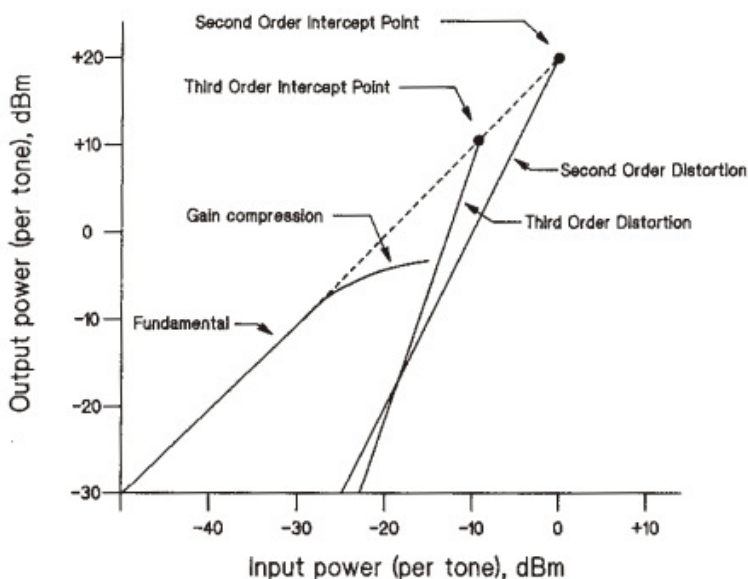


Fig. 11: Example for the 1 dB compression point [4]

the intercept point of n th order. These points provide important information concerning the quality of non-linear devices.

In this context, the aforementioned 1 dB compression point of an amplifier is the intercept point of first order. For the method of measurements of the 1 dB compression point, see Section 9.10.

Similar characterization techniques can also be applied with mixers, which, with respect to the LO signal, cannot be considered weakly non-linear devices.

2.6 The superheterodyne concept

The word superheterodyne is composed of three parts: super (Latin: over), $\epsilon\tau\epsilon\rho\omega$ (hetero, Greek: different) and $\delta\nu\nu\alpha\mu\sigma$ (dynamis, Greek: force) and can be translated as two forces superimposed². Different abbreviations exist for the superheterodyne concept. In the USA it is often referred to by the simple word ‘heterodyne’ and in Germany one can find the terms ‘super’ or ‘superhet’. The ‘weak’ incident signal is subjected to non-linear superposition (i.e. mixing or multiplication) with a ‘strong’ sine wave from a LO. At the mixer output we then get the sum and difference frequencies of the signal and LO. The LO signal can be tuned such that the output signal is always at the same frequency or in a very narrow frequency band. Therefore, a fixed-frequency bandpass with excellent transfer characteristics can be used, which is cheaper and easier than a variable bandpass with the same performance. A well-known application of

²The direct translation (roughly) would be: another force becomes superimposed.

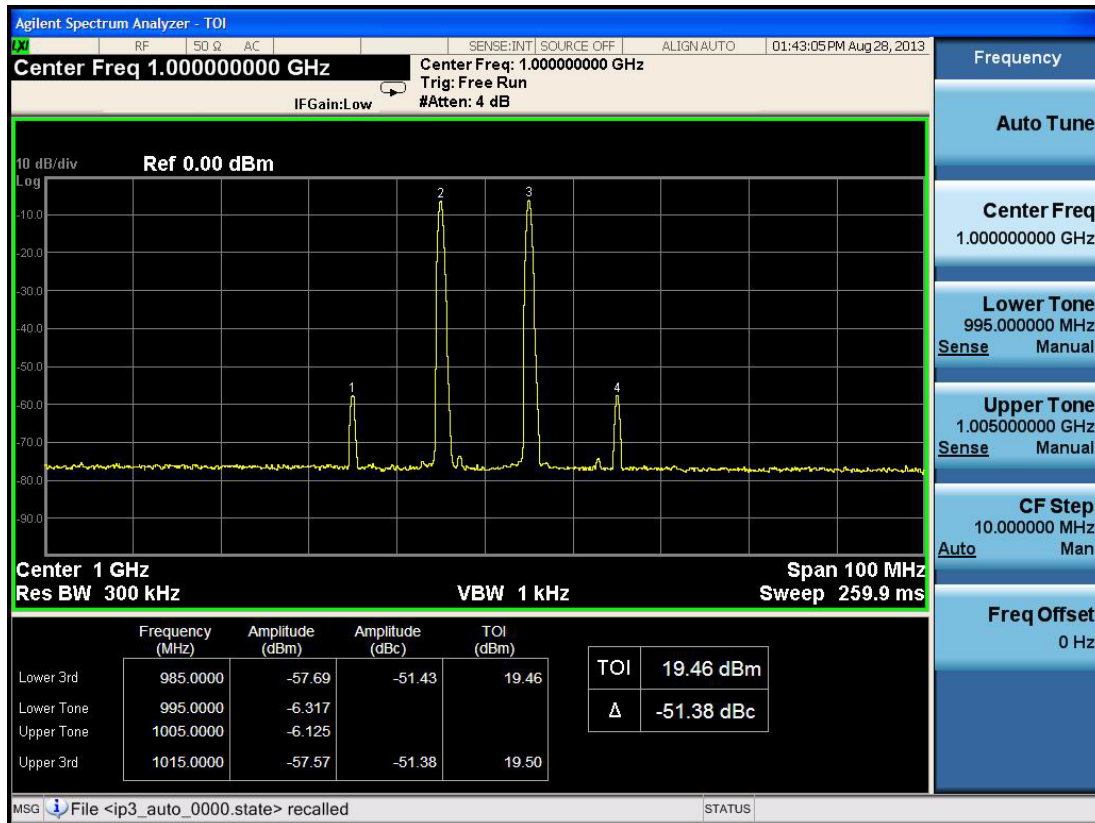


Fig. 12: An example of automatized TOI measurement

this principle is any simple radio receiver (Fig. 13).

3 SPECTRUM ANALYSER

RF spectrum analysers can be found in virtually every control room of a modern particle accelerator. They are used for many aspects of beam diagnostics including Schottky signal acquisition and RF observation. A spectrum analyser is in principle very similar to a common superheterodyne broadcast receiver, except for the requirements of choice of functions and change of parameters. It sweeps automatically through a specified frequency range, which corresponds to an automatic turning of a knob on a radio. The signal is then displayed in the amplitude/frequency plane. Thirty years ago, instruments were set manually and had some sort of analogue or cathode ray tube (CRT) display. Nowadays, with the availability of cheap and powerful digital electronics for control and data processing, nearly all instruments can be remotely controlled. The microprocessor permits fast and reliable setting of the instrument and reading of the measured values. Extensive data treatment for error correction, complex calibration routines and self tests is a great improvement. However, the user of such a sophisticated system may not always be aware of what is really going on in the analogue section before all data are digitized. The basis of these analogue sections is discussed now.

In general, there are two types of spectrum analyser:

- scalar spectrum analysers (SAs) and
- vector spectrum analysers (VSAs).

The SA provides only information of the amplitude of an ingoing signal, while the VSA provides the phase as well.

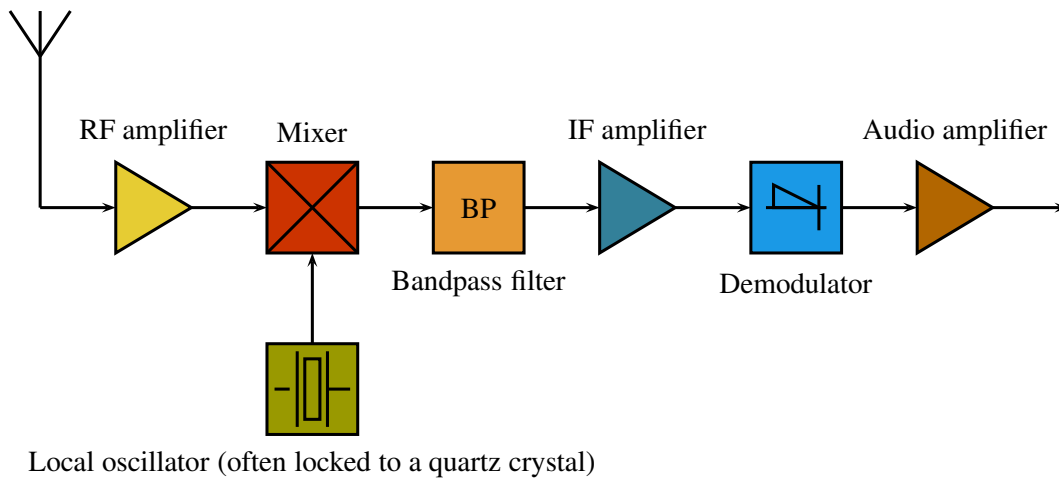


Fig. 13: Schematic drawing of a superheterodyne radio receiver

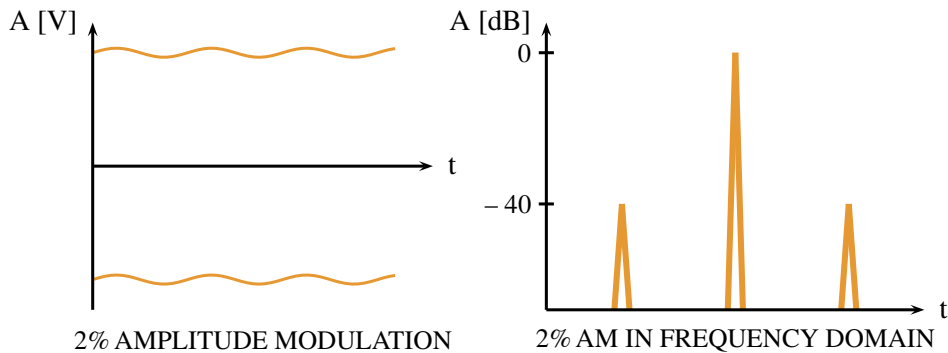


Fig. 14: Example of amplitude modulation in time and frequency domains

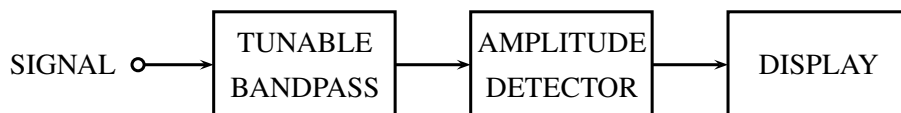


Fig. 15: A tunable bandpass as a simple spectrum analyser (SA)

3.1 Scalar spectrum analysers

A common oscilloscope displays a signal in the amplitude–time plane (time domain). The SA follows another approach and displays it in the frequency domain.

One of the major advantages of the frequency-domain display is the sensitivity to periodic perturbations. For example, 5% distortion is already difficult to see in the time domain but in the frequency domain the sensitivity to such ‘sidelines’ (Fig. 14) is very high (–120 dB below the main line). A very faint amplitude modulation (AM) of 10^{-12} (power) on some sinusoidal signals would be completely invisible on the time trace, but can be displayed as two sidelines 120 dB below the carrier in the frequency domain [5].

We will now consider only serial processing or swept tuned analysers (Fig. 15).

The easiest way to design a swept tuned spectrum analyser is by using a tunable bandpass. This

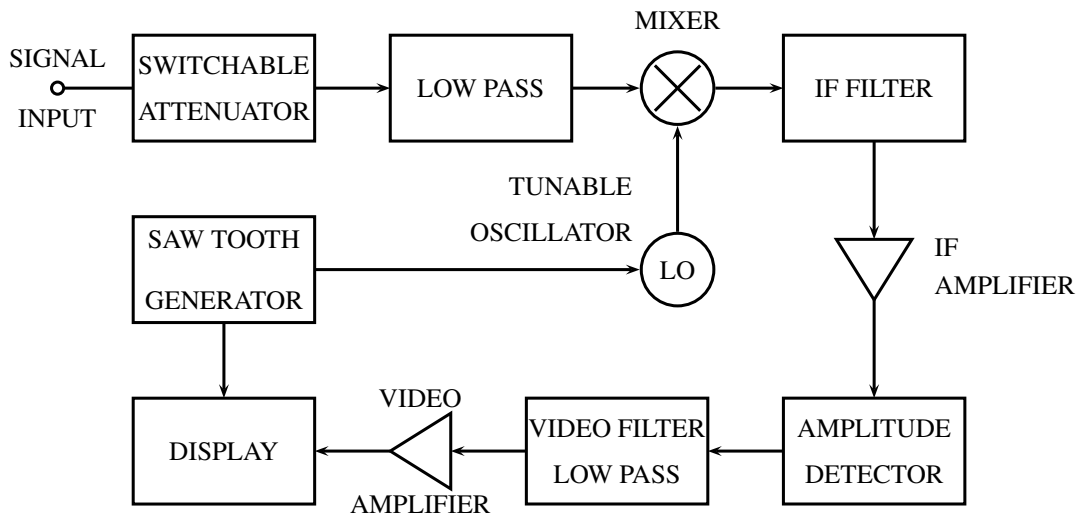


Fig. 16: Block diagram of a spectrum analyser

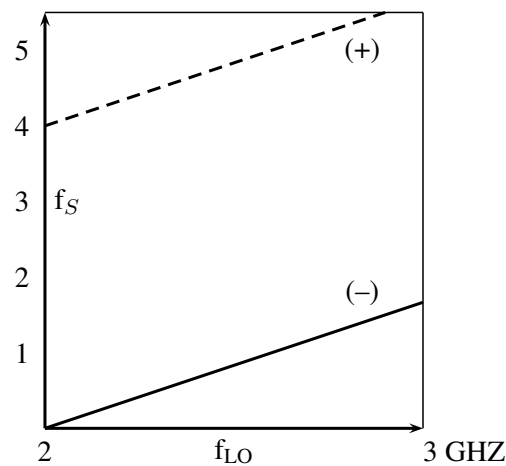


Fig. 17: Frequency chart of the SA of Fig. 16, IF = 2 GHz

may be an LC circuit or a YIG filter (YIG = yttrium iron garnet) beyond 1 GHz. The LC filter exhibits poor tuning, stability and resolution. YIG filters are used in the microwave range (as preselectors) and for YIG oscillators. Their tuning range is about one decade, with Q values exceeding 1000.

For much better performance, the superheterodyne principle can be applied (Fig. 13).

As already mentioned, the non-linear element (four-diode mixer or double-balanced mixer) delivers mixing products as

$$f_s = f_{LO} \pm f_{IF}. \quad (8)$$

Assuming a signal range from 0 to 1 GHz for the spectrum analyser depicted in Fig. 16 and f_{LO} between 2 and 3 GHz, we get the frequency chart shown in Fig. 17.

Obviously, for a wide input frequency range without image response we need a sufficiently high IF. A similar situation occurs for AM- and FM-broadcast receivers (AM-IF = 455 kHz, FM-IF = 10.7 MHz). But, for a high IF (e.g. 2 GHz) a stable narrowband IF filter is difficult to construct, which is why most SAs and high-quality receivers use more than one IF. Certain SAs have four different LOs, some fixed, some tunable. For a large tuning range the first, and for a fine tune (e.g. 20 kHz), the third LO is

tuned.

Multiple mixing is necessary when going to a lower IF (required when using high- Q quartz filters) for good image response suppression of the mixers.

It can be shown that the frequency of the n th LO must be higher than the (say) 80 dB bandwidth (BW) of the $(n - 1)$ th IF-band filter. A disadvantage of multiple mixing is the possible generation of intermodulation lines if amplitude levels in the conversion chain are not carefully controlled.

The requirements of a modern SA with respect to frequency are

- high resolution,
- high stability (drift and phase noise),
- wide tuning range,
- no ambiguities

and, with respect to amplitude response, are

- large dynamic range (100 dB),
- calibrated, stable amplitude response,
- low internal distortions.

It should be mentioned that the size of the smallest IF-bandpass filter width Δf has an important influence on the maximum sweep rate (or step width and rate when using a synthesizer):

$$\frac{df}{dt} < (\Delta f)^2. \quad (9)$$

In other words, the signal frequency has to remain at $\Delta T = 1/\Delta f$ within the bandwidth Δf .

On many instruments the proper relation between Δf and the sweep rate is automatically set to the optimum value for the highest possible sweep speed, but it can always be altered manually (setting of the resolution bandwidth).

Caution is advised when applying, but not necessarily displaying, two or more strong (> 10 dBm) signals to the input. Intermodulation third-order products may appear (from the first mixer or amplifier) and could lead to misinterpretation of the signals to be analysed.

Spectrum analysers usually have a rather poor noise figure of 20–40 dB, as they often do not use pre-amplifiers in front of the first mixer (dynamic range, linearity). But, with a good pre-amplifier, the noise figure can be reduced to almost that of the pre-amplifier. This configuration permits amplifier noise-figure measurements to be made with reasonable precision of about 0.5 dB. The input of the amplifier to be tested is connected to hot and cold terminations and the two corresponding traces on the SA display are evaluated [6–10].

4 VECTOR SPECTRUM AND FAST FOURIER TRANSFORM ANALYSER

The modern vector spectrum analyser (VSA) is essentially a combination of a two-channel digital oscilloscope and a spectrum analyser fast Fourier transform (FFT) display. The incoming signal gets down-mixed, bandpass (BP) filtered and passes an analog-to-digital converter (ADC) (generalized Nyquist for BP signals; $f_{\text{sample}} = 2 \cdot \text{BW}$). A schematic drawing of a modern VSA can be seen in Fig. 18.

The digitized time trace then is split into I (in-phase) and Q (quadrature, 90 degree offset) components with respect to the phase of some reference oscillator. Without this reference, the term ‘vector’ is meaningless for a spectral component.

One of the great advantages is that a VSA can easily separate AM and FM components.

An example of vector spectrum analyser display and performance is given in Figs. 19 and 20. Both figures were obtained during measurements of the electron cloud in the CERN SPS (Super Proton Synchrotron).

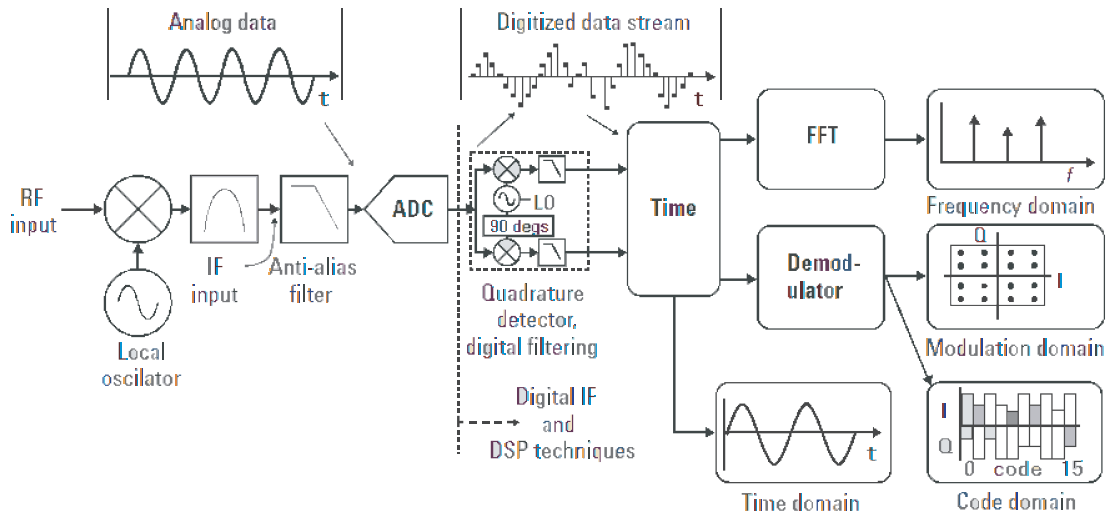


Fig. 18: Block diagram of a vector spectrum analyser

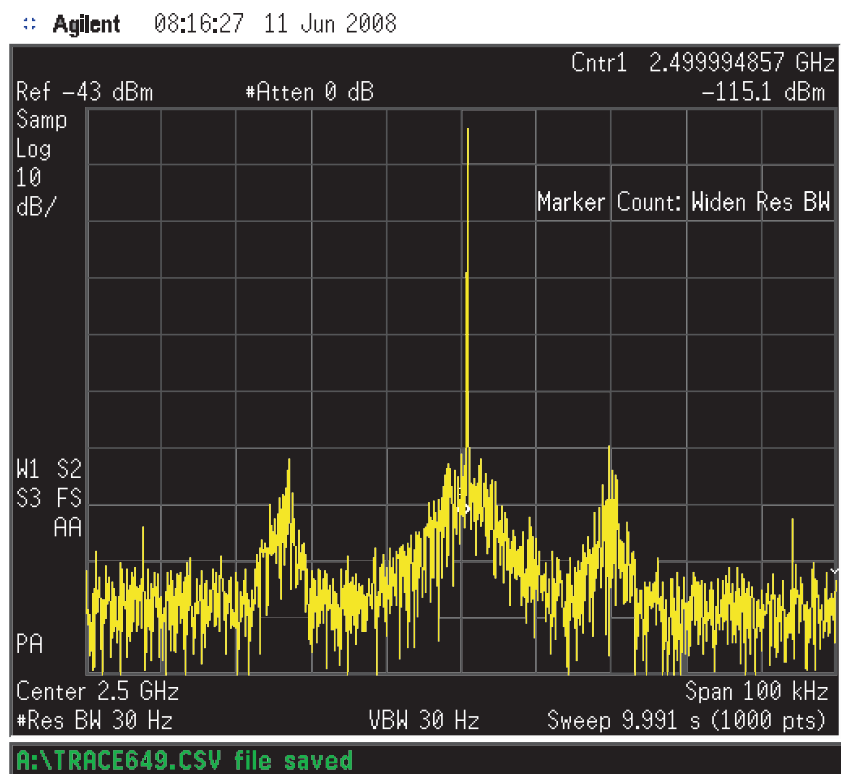


Fig. 19: Single-sweep FFT display similar to a very slow scan on a swept spectrum analyser

5 NOISE BASICS

The concept of ‘noise’ was applied originally to the type of audible sound caused by statistical variations of the air pressure with a wide flat spectrum (white noise). It is now also applied to electrical signals, the noise ‘floor’ determining the lower limit of signal transmission. Typical noise sources are: Brownian movement of charges (thermal noise), variations of the number of charges involved in the conduction (flicker noise) and quantum effects (Schottky noise, shot noise). Thermal noise is only emitted by structures with electromagnetic losses, which, by reciprocity, also absorb power. Pure reactances do not emit noise (emissivity = 0).

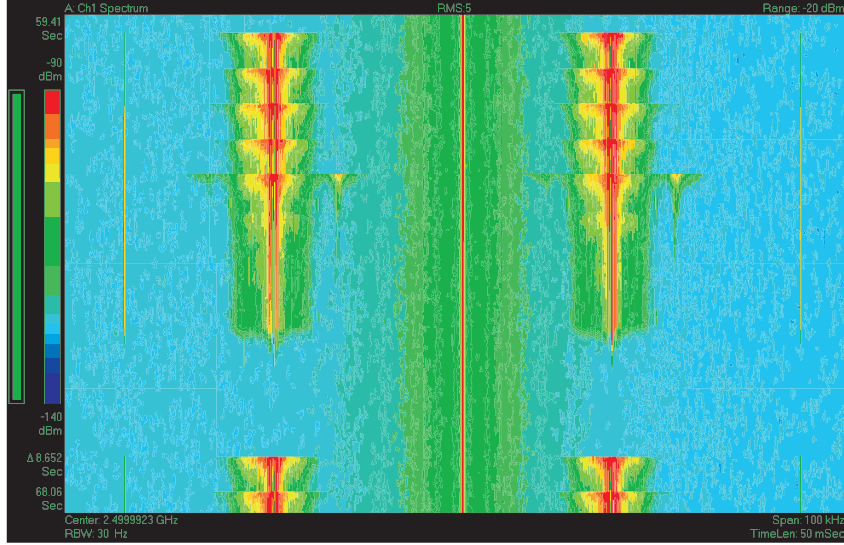


Fig. 20: Spectrogram display containing about 200 traces as shown on the left-hand side in colour coding. Time runs from top to bottom.

Different categories of noise can be defined:

- white, which has a flat spectrum,
- pink, being low-pass filtered and
- blue, being high-pass filtered.

In addition to the spectral distribution, the amplitude density distribution is also required in order to characterize a stochastic signal. For signals coming from very many independent sources, the amplitude density has a Gaussian distribution. The noise power density delivered to a load by a black body is given by Planck's formula:

$$\frac{N_L}{\Delta f} = hf \left(e^{hf/kT} - 1 \right)^{-1}, \quad (10)$$

where N_L is the noise power delivered to the load, $h = 6.625 \cdot 10^{-34}$ J s the Planck constant and $k = 1.38056 \cdot 10^{-23}$ J/K Boltzmann's constant.

Equation (10) indicates constant noise power density up to about 120 GHz (at 290 K) with 1% error. Beyond, the power density decays and there is no 'ultraviolet catastrophe', i.e. the total noise power is finite.

The radiated power density of a black body is given as

$$W_r(f, T) = \frac{hf^3}{c^2 [e^{hf/kT} - 1]}. \quad (11)$$

For $hf \ll kT$ the Rayleigh-Jeans approximation of Eq. (10) holds:

$$N_L = kT\Delta f, \quad (12)$$

where in this case N_L is the power delivered to a matched load. The no-load noise voltage $u(t)$ of a resistor R is given as

$$\overline{u^2(t)} = 4kTR\Delta f \quad (13)$$

and the short-circuit current $i(t)$ by

$$\overline{i^2(t)} = 4\frac{kT\Delta f}{R} = 4kTG\Delta f, \quad (14)$$

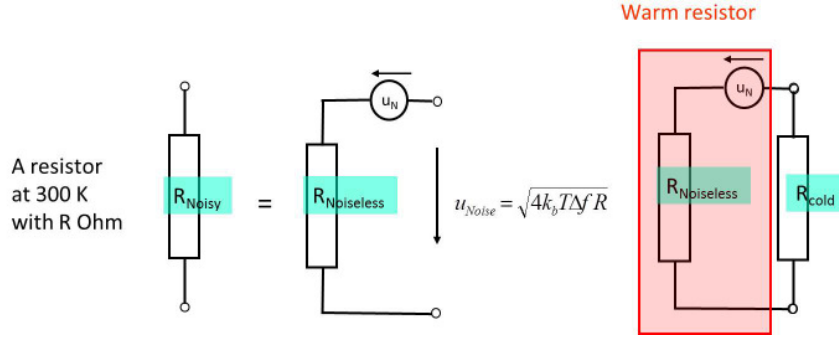


Fig. 21: Equivalent circuit for a noisy resistor terminated by a noiseless load

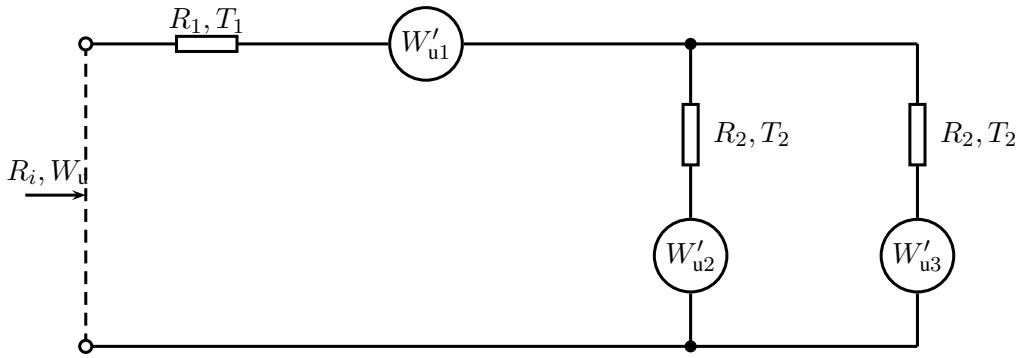


Fig. 22: Noisy one-port with resistors at different temperatures [6, 11]

where $u(t)$ and $i(t)$ are stochastic signals and G is $1/R$. The linear average $\overline{u(t)}$, $\overline{i(t)}$ vanishes. Of special importance is the quadratic average $\overline{u^2(t)}$, $\overline{i^2(t)}$.

The available power (which is independent of R) is given by (Fig. 21)

$$\frac{\overline{u^2(t)}}{4R} = kT \Delta f. \quad (15)$$

We define a spectral density function [6]

$$\begin{aligned} W_u(f) &= 4kTR, \\ W_i(f) &= 4kTG, \\ \overline{u^2(t)} &= \int_{f_1}^{f_2} W_u(f) df. \end{aligned} \quad (16)$$

A noisy resistor may be composed of many elements (resistive network). In general, it is made from many carbon grains which have homogeneous temperatures. But, if we consider a network of resistors with different temperatures and hence with an inhomogeneous temperature distribution (Fig. 22), the spectral density function changes to

$$W_u = \sum_j W_{uj} = 4kT_n R_i, \quad (17)$$

$$T_n = \sum_j \beta_j T_j, \quad (18)$$

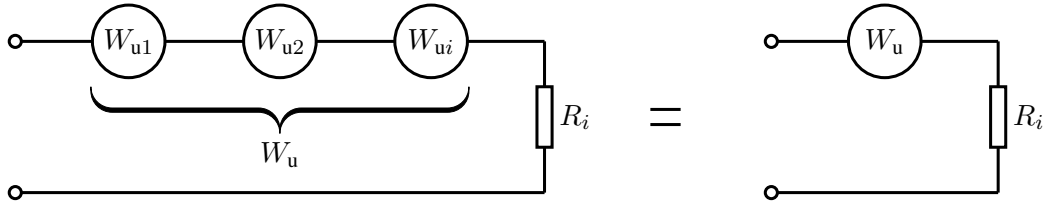


Fig. 23: Equivalent sources for the circuit of Fig. 22

where W_{uj} are the noise sources (Fig. 23), T_n is the total noise temperature, R_i the total input impedance and β_j are coefficients indicating the fractional part of the input power dissipated in the resistor R_j . It is assumed that the W_{uj} are uncorrelated for reasons of simplicity.

The relative contribution (β_j) of a lossy element to the total noise temperature is equal to the relative dissipated power multiplied by its temperature:

$$T_n = \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \dots \quad (19)$$

A nice example is the noise temperature of a satellite receiver, which is nothing else than a directional antenna. The noise temperature of free space amounts roughly to 3 K. The losses in the atmosphere, which is an air layer of 10 to 20 km length, cause a noise temperature at the antenna output of about 10 to 50 K. This is well below room temperature of 290 K.

So far, only pure resistors have been considered. Looking at complex impedances, it can be seen that losses from dissipation occur in $\text{Re}(Z)$ only. The available noise power is independent of the magnitude of $\text{Re}(Z)$ with $\text{Re}(Z) \neq 0$. For Figs. 22 and 23, Eq. (18) still applies, except that R_i is replaced by $\text{Re}(Z_i)$. However, it must be remembered that in complex impedance networks the spectral power density W_u becomes frequency dependent [11].

The rules mentioned above apply to passive structures. A forward-biased Schottky diode (external power supply) has a noise temperature of about $T_0/2 + 10\%$. A biased Schottky diode is not in thermodynamic equilibrium and only half of the carriers contribute to the noise [6]. But, it represents a real 50Ω resistor when properly forward biased. For transistors and in particular field-effect transistors (FETs), the physical mechanisms are somewhat more complicated. Noise temperatures of 50 K have been observed on a FET at 290 K physical temperature.

6 NOISE-FIGURE MEASUREMENT WITH THE SPECTRUM ANALYSER

Consider an ideal amplifier (noiseless) terminated at its input (and output) with a load at 290 K with an available power gain (G_a). We measure at the output [7, 12]:

$$P_a = kT_0 \Delta f G_a. \quad (20)$$

For $T_0 = 290$ K (or often 300 K), we obtain $kT_0 = -174$ dBm/Hz ($-$ dBm = decibel below 1 mW). At the input we have for some signal S_i a certain signal/noise ratio S_i/N_i and at the output S_o/N_o . For an ideal (= noiseless) amplifier S_i/N_i is equal to S_o/N_o , i.e. the signal and noise levels are both shifted by the same amount. This gives the definition of the noise figure F :

$$F = \frac{S_i/N_i}{S_o/N_o}. \quad (21)$$

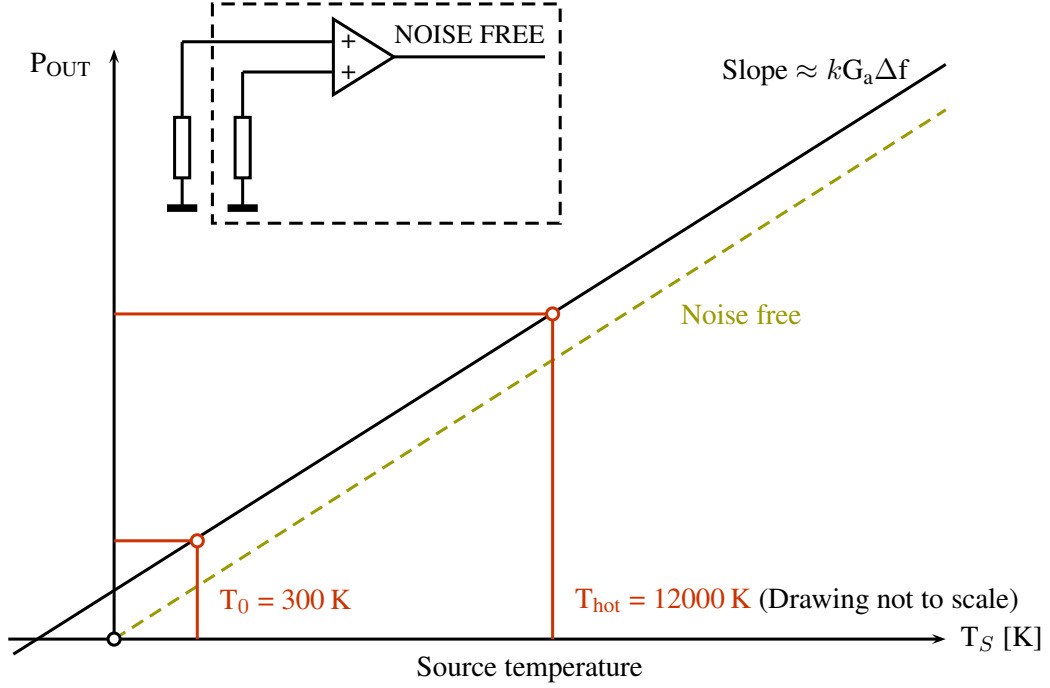


Fig. 24: Relation between source noise temperature T_s and output power P_{out} for an ideal (noise-free) and a real amplifier [7, 12].

The ideal amplifier has $F = 1$ or $F = 0$ dB and the noise temperature of this amplifier is 0 K. The real amplifier adds some noise, which leads to a decrease in S_o/N_o due to the noise added ($= N_{ad}$):

$$F = \frac{N_{ad} + kT_0 \Delta f G_a}{kT_0 \Delta f G_a}. \quad (22)$$

For a linear system the minimum noise figure amounts to $F_{min} = 1$ or 0 dB. However, for non-linear systems one may define noise figures $F < 1$. Now assume a source with variable noise temperature connected to the input and measure the linear relation between amplifier output power and input termination noise temperature ($T_s = T_{source}$).

In a similar way, a factor Y can be defined:

$$Y = \frac{T_e + T_H}{T_e + T_C},$$

$$T_e = \frac{T_H - Y T_C}{Y - 1},$$

$$F = \frac{[(T_H/290) - 1] - Y [(T_C/290) - 1]}{Y - 1}, \quad (23)$$

where T_e is the effective input noise temperature (see Fig. 24) and T_H and T_C are the noise temperatures of a hot or a cold input termination. To find the two points on the straight line of Fig. 24, one may switch between two input terminations at 373 K (100°C) and 77 K. For precise reading of RF power, calibrated piston attenuators in the IF path (IF superheterodyne receiver) are in use. This is the hot/cold method. The difference between the Y factor and the hot/cold method is that for the latter the input of the amplifier becomes physically connected to resistors at different temperatures (77 and 373 K). For the Y factor, the noise temperature of the input termination is varied by electronic means between 300 and 12000 K (physical temperature always around 300 K).

As a variant of the 3 dB method with a controllable noise source, the excess noise temperature definition ($T_{\text{ex}} = T_{\text{H}} - T_{\text{C}}$) is often applied. A switchable 3 dB attenuator at the output of the amplifier just cancels the increase in noise power from $T_{\text{H}} - T_{\text{C}}$. Thus, the influence of non-linearities of the power meter is eliminated. To measure the noise of a one-port one may also use a calibrated spectrum analyser. However, spectrum analysers have high noise figures (20–40 dB) and the use of a low-noise pre-amplifier is recommended. This ‘total power radiometer’ [8] is not very sensitive but often sufficient, e.g., for observation of the Schottky noise of a charged-particle beam. Note that the spectrum analyser may also be used for two-port noise-figure measurements. An improvement of this ‘total power radiometer’ is the ‘Dicke radiometer’ [8]. It uses a 1 kHz switch between the unknown one-port and a controllable reference source. The reference source is made equal to the unknown via a feedback loop, and one obtains a resolution of about 0.2 K. Unfortunately, switch spikes sometimes appear. Nowadays, switch-free correlation radiometers with the same performance are available [13].

The noise figure of a cascade of amplifiers is [6, 7, 11–13]

$$F_{\text{total}} = F_1 + \frac{F_2 - 1}{G_{a1}} + \frac{F_3 - 1}{G_{a1}G_{a2}} + \dots \quad (24)$$

As can be seen from Eq. (24), the first amplifier in a cascade has a very important effect on the total noise figure, provided G_{a1} is not too small and F_2 is not too large. In order to select the best amplifier from a number of different units to be cascaded, one can use the noise measure M :

$$M = \frac{F - 1}{1 - (1/G_a)} \quad (25)$$

The amplifier with the smallest M has to be the first in the cascade [12].

7 INTRODUCTION TO NETWORK ANALYSIS AND S-PARAMETERS

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 non-linear components with several input and output ports.

A network analyser is one of the most versatile and valuable pieces of measurement equipment used in a RF laboratory or particle accelerator control room. The best commercially available network analysers 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 the following sections, scalar and vector network analysers 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 non-linear (e.g. active elements such as amplifiers or mixers). For analysis of these devices certain approximations or extensions of the definitions are required [15].

Another interesting application is the determination of the beam transfer function (BTF), where the DUT is a circulating particle beam in an accelerator.

7.1 One-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 waves (b). The incident wave travels from a source to the DUT – the reflected one in the opposite direction. The fundamental reason for this

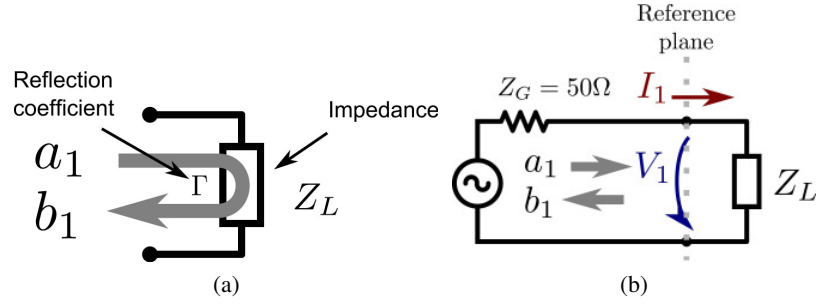


Fig. 25: Wave quantities of a one-port (with two poles) and impedance Z_L : (a) incident (a_1) and reflected (b_1) wave; (b) relation of a_1 and b_1 to V_1 and I_1 .

terminology is the fact that in RF engineering the linear geometrical dimensions are often larger than 10% of the corresponding free-space wavelength. This also requires the definition of a reference plane to which the measurement is referring. 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 at very low frequencies and voltage and current descriptions can be used at 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. 25a) is described as a 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

$$a_1 = \frac{V_1 + I_1 Z_0}{2\sqrt{Z_0}}, \quad b_1 = \frac{V_1 - I_1 Z_0}{2\sqrt{Z_0}}, \quad (26)$$

where V_1 and I_1 represent the voltage and current respectively at the port as depicted in Fig. 25b. Z_0 is an arbitrary reference impedance (often, but not necessarily always, the characteristic impedance $Z_0 = Z_G = 50\ \Omega$ of the system).

The wave quantities have the dimension of \sqrt{W} (see [14]). This normalization is important for the conservation of energy. The power which is travelling towards the DUT can be calculated by $P_{\text{inc}} = |a|^2$, the reflected power by $|b|^2$. It is important to note that this definition is mainly used in the USA – in European notation, the incident power is usually calculated by $P_{\text{inc}} = 0.5|a|^2$. 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 as

$$\Gamma = \frac{b_1}{a_1}. \quad (27)$$

By substitution with Eq. (26), we can find a relation between the complex impedance Z_L of a

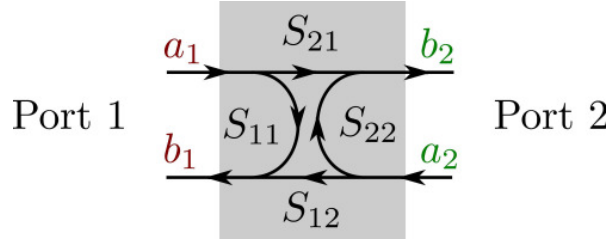


Fig. 26: All possible S-parameters of a two-port network

one-port and its complex reflection coefficient Γ :

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0}. \quad (28)$$

There are some particular cases which are worth noting:

7.2 Two-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 directions 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. 26. A two-port has four complex and frequency-dependent scattering parameters:

$$S_{11} = \frac{b_1}{a_1}, \quad S_{12} = \frac{b_1}{a_2}, \quad S_{21} = \frac{b_2}{a_1}, \quad S_{22} = \frac{b_2}{a_2}. \quad (29)$$

Here S_{11} and S_{22} are equal to the reflection coefficients Γ on the respective ports – but *only* with the condition that the corresponding other port is terminated with the characteristic impedance. S_{21} and S_{12} are the forward and reverse transmission coefficients, 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 analyser 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 impedances).

It is very important to *terminate all ports of the DUT with the respective characteristic impedances*. 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 words the generator impedance must also match the impedance of the DUT for the port under consideration. As a practical example one cannot measure this in a straightforward manner – unless a special calibration procedure is used – with a 50Ω network analyser 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.

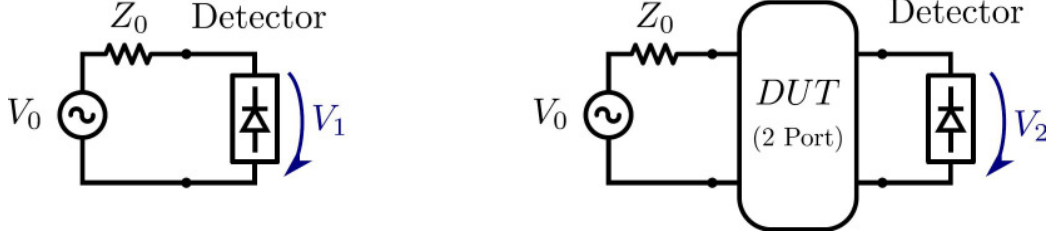


Fig. 27: A simple measurement set-up for the scalar transmission coefficient ($|S_{21}|$)

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-parameters measured shall be independent of the instrumentation used for the measurement.

Once all n^2 S-parameters for an 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

$$\begin{aligned} b_1 &= S_{11}a_1 + S_{12}a_2, \\ b_2 &= S_{21}a_1 + S_{22}a_2. \end{aligned} \quad (30)$$

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

$$\vec{b} = \mathbf{S}\vec{a} \quad (31)$$

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}. \quad (32)$$

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 can be found in [14].

8 SCALAR NETWORK ANALYSIS

In a scalar network analyser, 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 kinds of devices are generally less expensive than vector network analysers. A very simple measurement set-up as used more than 50 years ago is shown in Fig. 27.

The measurement is done twice, the first time (Fig. 27, left) without a DUT to measure the power of the incident signal (V_1). Then the DUT is inserted (Fig. 27, right), V_2 is measured and the magnitude of the transmission coefficient can be calculated by

$$|S_{21}| \propto \frac{V_2}{V_1}. \quad (33)$$

For obtaining a reading in decibels, a logarithmic amplifier is sometimes used following the detector. It has a logarithmic transfer function ($V_{\text{out}} = \log V_{\text{in}}$) and permits the showing of a large dynamic range on a dB scale. Also, the mathematical operation of a division, required for normalization in Eq. (33), then transforms into the easier procedure of a subtraction.

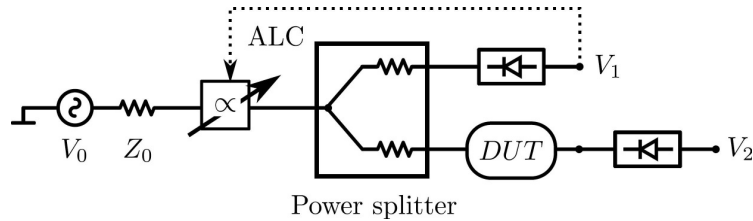


Fig. 28: Simplified circuit diagram of a typical automatic level control

The detector can be any kind of device converting the input RF signal into a DC voltage, which is ‘more or less’³ proportional to RF power. There are basically three possibilities to achieve this:

Rectifier A very fast Schottky diode and a low-pass filter are used to convert the input RF signal to a DC voltage. Operating the diode in its square-law region ($P_{in} < -10$ dBm) results in an output voltage proportional to the RF power; see Section 2.2. 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 non-linearity of the output signal versus 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 $Z_0 = 50 \Omega$.

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 metals or metal alloys is used. By a non-linear 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 platinum. 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 thermo-electrical coefficient of a junction between two different metals. A well-known example is the Sb-Bi 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 $250 \mu\text{V/K}$ can be obtained. For further details, see [16].

Mixer Multiplying two sinusoidal signals with different frequencies results in signals at the sum and difference frequencies at the multiplier’s output; see Section 2.3. This can technically be used to convert a band of high-frequency signals to a much lower IF. Now all the measurements can be carried out at this IF.

8.1 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 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 inputs and outputs matched to $Z_0 = 50 \Omega$ (Fig. 28). 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.

³With the term ‘more or less’, it should be stressed that many detectors have a non-linear relation between input power and output voltage.

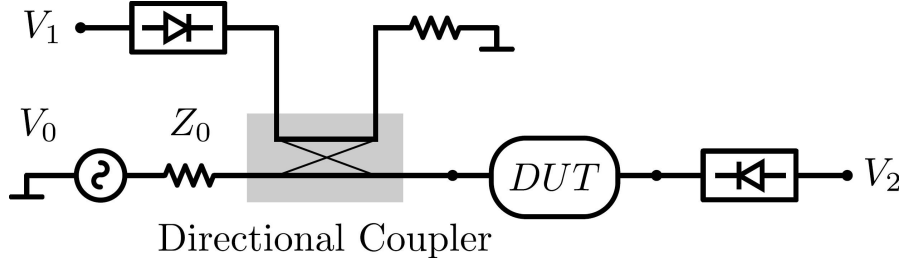


Fig. 29: Feedback loop of a typical automatic level control (ALC)

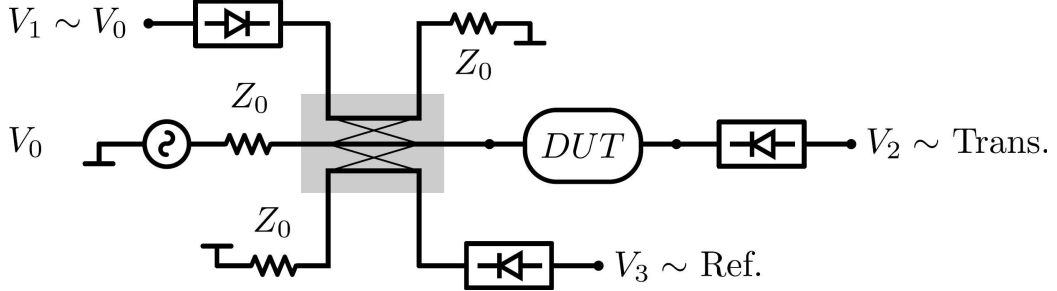


Fig. 30: Dual directional coupler in a network analyser

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 unlevelled generator, but in practice the levelling has a large number of advantages, in particular for measurements on weakly non-linear elements such as amplifiers.

8.2 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. 29. V_1 is an attenuated (by the coupling factor) replica of the forward-travelling wave, which is only used for levelling 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.

Modern network analysers (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 waves from the DUT. This is shown in Fig. 30. All those directional couplers in combination with switches and attenuators are commonly called a test set. 20 or 30 years ago, network analysers consisted of separate building blocks like S-parameter 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 input and output facilities.

From Fig. 30, the reflection and transmission coefficients are defined by

$$|S_{11}| \propto \frac{V_3}{V_1}, \quad |S_{21}| \propto \frac{V_2}{V_1}. \quad (34)$$

From the ratio of the reflected wave to the incident wave (S_{11}), quantities such as the standing wave ratio (SWR), reflection coefficient, impedance, admittance as well as return loss of the DUT can

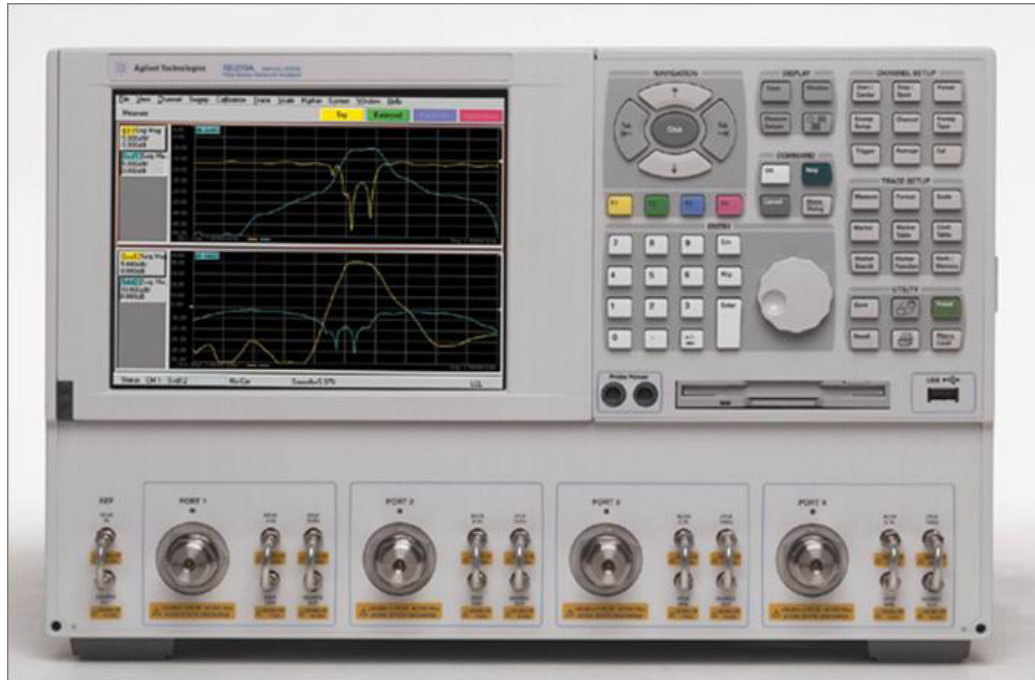


Fig. 31: A modern four-port VNA

be determined. From the ratio of the transmitted wave to the incident wave (S_{21}), the gain or insertion loss, the transmission coefficient, the insertion phase and group delay of the DUT can be found.

9 VECTOR MEASUREMENTS

A vector network analyser (VNA) is able to measure the magnitude *and phase* of a complex S-parameter. There are different possible hardware configurations for the implementation of such an instrument like six-port reflectometers, certain RF bridge methods or superheterodyne RF network analysers. Here only the last one will be described.

9.1 Modern vector network analysers

A modern network analyser contains a 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 analogue or digital down mixing, 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 [17, 18]. Note that similar to the spectrum analyser, the sweep time and resolution bandwidth cannot be adjusted independently. A modern four-port vector network analyser is shown in Fig. 31.

Although complete network analysis of any N -port can be done with a two-port device, a four-port 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.

9.2 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 transformation⁴ and vice versa. This is the basis of the synthetic pulse technique, available on many modern VNAs. It was commercially introduced by Hewlett-Packard in the 1980s for network analyser applications.

It renders the VNA even more versatile, allowing the display of the impulse (Gaussian) and also the step response of the DUT and the carrying out of 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.

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. 32. 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 versus time. The VNA permits the display of either the synthetic step (Fig. 32a) or the impulse response (Fig. 32b). 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 $t_d = 22$ ns 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:

$$l = \frac{c}{\sqrt{\epsilon}} \cdot \frac{1}{2} t_d. \quad (35)$$

The mechanical line length is given by Eq. (35). In this example we have $\epsilon = 2.3$, which returns a line length of $l = 2.2$ m. 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 ($1/\sqrt{\epsilon}$ 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. 32a returns the local reflection factor versus 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 Table 2).

The fact that in the pulse response display (Fig. 32b) 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 2 m long). The amplitude of this reflection from the open end indicates the attenuation over twice the electrical length of the cable at the centre frequency ($f_{\max} = 3$ GHz, $f_{\text{centre}} = 1.5$ GHz).

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 Table 3. For example: a long cable needs to be tested. Due to the long time window required in order to ensure that

⁴More 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.

Table 2: Key numbers for the reflection coefficient

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

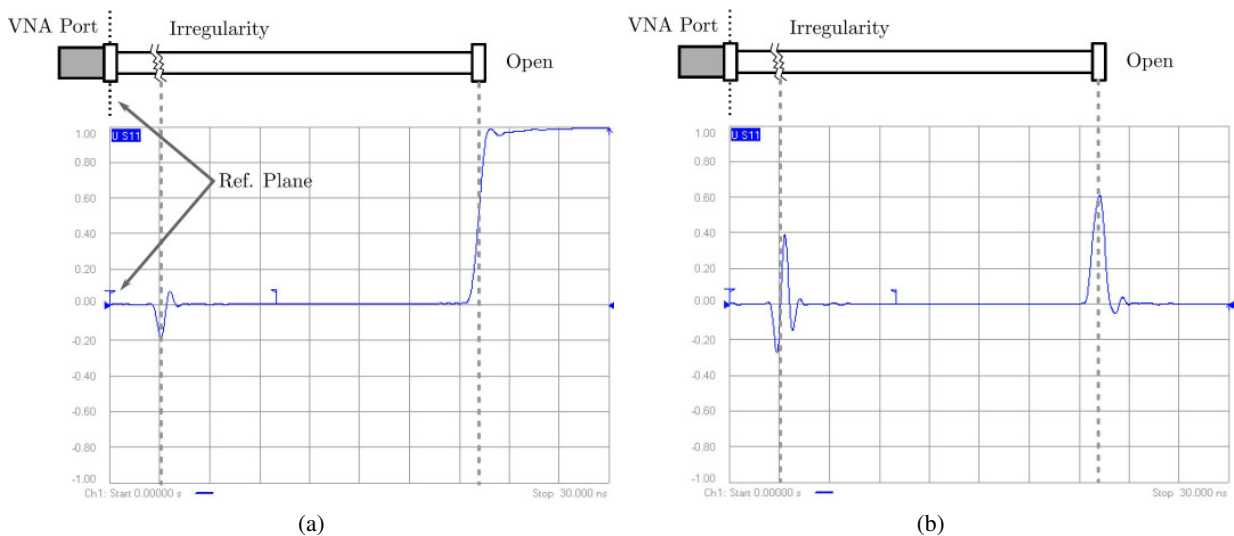


Fig. 32: Synthetic pulse measurement with a VNA: (a) step response; (b) impulse response. The measured frequency data is converted by an inverse discrete Fourier transformation (iDFT) to the time domain. Now the synthetic impulse response of the cable can be plotted over time. The reflections of the 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.

Table 3: Most important characteristics of the FFT

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

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 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

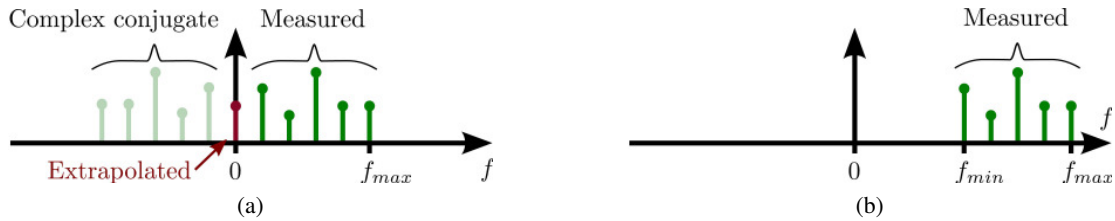


Fig. 33: Sampling of frequency points in the different operating modes: (a) low-pass mode; (b) bandpass mode.

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 ‘bandpass’ mode. Both of these modes are briefly discussed in the following sections.

9.3 ‘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. 33a). 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 cannot 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 frequencies by complex conjugation. Compared to the bandpass 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 time domain reflectometry (TDR) measurement routine usually goes as follows:

1. The DUT is connected, the port and type of measurement are selected (transmission or reflection).
2. The frequency range of interest and the number of data points are entered (this relates to the time domain by Table 3)
3. After pushing the soft key, ‘set frequency low-pass’⁵, 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 time domain reflectometry (TDR) and time domain transmission (TDT) is basically equivalent to what is shown on a real-time or sampling oscilloscope; see Section 9.8.

9.4 ‘Bandpass mode’

In the bandpass mode (Fig. 33b) 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 $f_{\min} = 1.2$ GHz to $f_{\max} = 1.5$ GHz). The start and stop frequencies 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 bandpass mode is the equivalent of a narrowband TDR (and also time-domain transmission TDT) using the synthetic pulse technique. It permits the display of an impulse

⁵This soft key may appear with slightly different naming, depending on the definitions of the manufacturer.

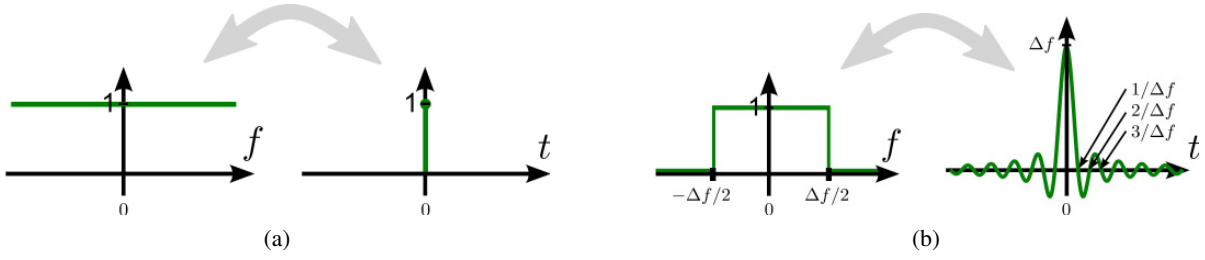


Fig. 34: (a) Infinite frequency span. (b) Limited frequency span. The limited frequency span Δf of the VNA leads to distortions in the time-domain synthetic pulse measurement. The ideal response is convoluted with a sinc function. Its characteristics depend on Δf .

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. Their characterization in terms of capacitive, inductive or resistive properties is possible, but not straightforward [19]. More on the general properties and mathematical backgrounds of the low-pass and bandpass modes can be found in [18, 20].

9.5 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. 34.

An infinite spectrum with constant density (shown in Fig. 34a) leads to a Dirac-pulse function in the time domain. The Dirac pulse contains by definition all frequency components with equal power. In Fig. 34b, 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 denoted as a si function) in the time domain. This relation is shown in Eq. (36) and graphically in

$$\text{Frequency domain} \quad \Longleftrightarrow \quad \text{Time domain} \quad (36)$$

$$\text{rect}\left(\frac{f}{\Delta f}\right) \quad \Longleftrightarrow \quad \frac{\sin(\Delta f \pi t)}{\pi t} = \Delta f \cdot \text{sinc}(\Delta f \pi t). \quad (37)$$

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} (bandpass mode) but for f_{\max} only in the low-pass mode. This helps to reduce strong sidelobes (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. 35.

9.6 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. 36 at the range of about $t = 18$ ns to $t = 26$ ns), we are able to analyse the selected pulse only and are not bothered by multiple reflections and

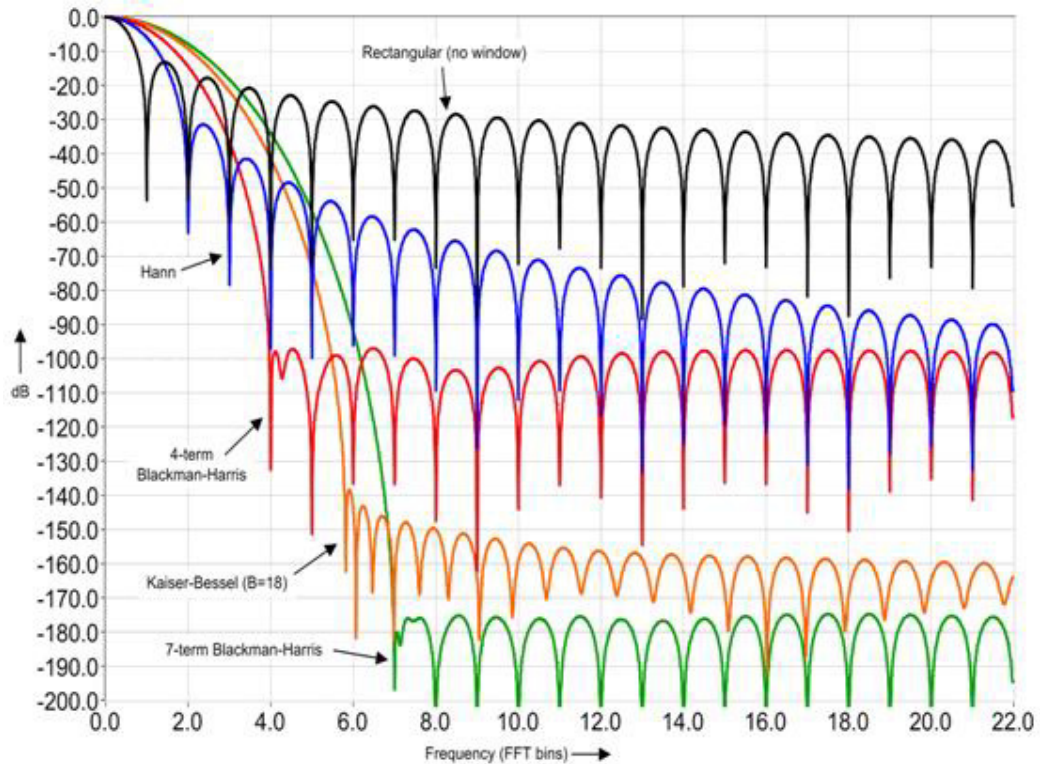


Fig. 35: Typical window functions to suppress strong sidelobes

other perturbations. Thus, by 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 pulses 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.

However, when using the gating function, it should always be kept in mind that this operation is *non-linear*. This implies that 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.

9.7 Examples of synthetic pulse time-domain measurements

A collection of measurement examples of simple DUTs are shown in Fig. 37. In all cases depicted there, the VNA is set up to measure the step response. The traces from top to bottom show:

1. Matched load ($Z = Z_C$). As Γ is equal to zero, the response is zero over the whole time.
2. Moderate (resistive) mismatch ($Z = 2Z_C$, e.g. $100\ \Omega$ in a $50\ \Omega$ 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 circuit 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 circuit and terminates with an exponential function in a short circuit, as the inductor starts conducting.

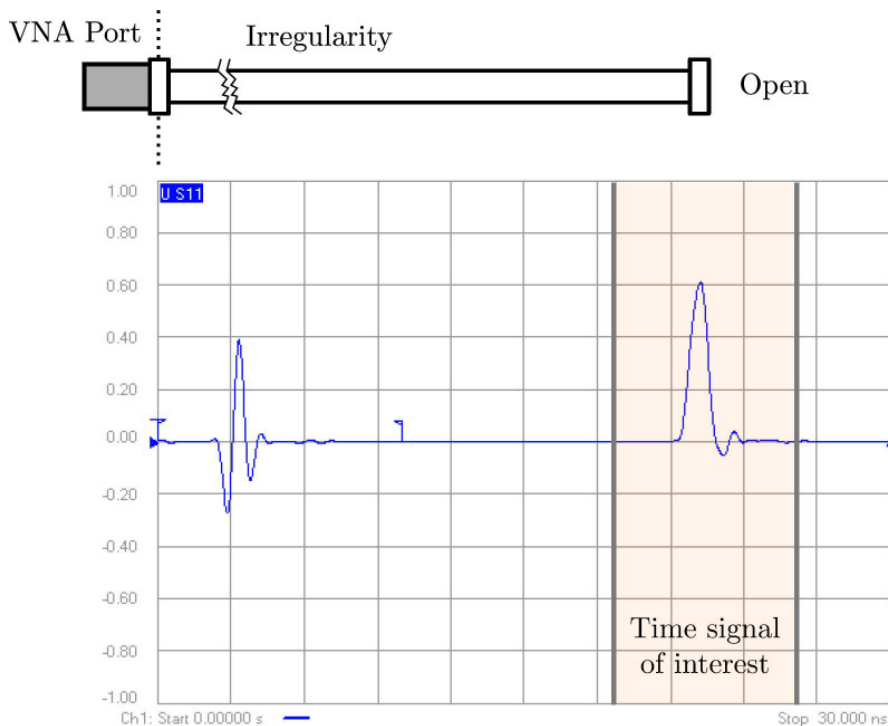


Fig. 36: Only the signal in a certain time window is of interest. After selection, the FFT of this window will be calculated. Here the real values of the synthetic impulse response are shown in a linear scale.

9.8 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 oscilloscope. 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 non-linear and time-varying DUTs, real pulse measurements are still indispensable, as in most of the air traffic radar systems, where we have linear but time-varying conditions.

The dynamic range of a typical sampling oscilloscope is limited to about 60 to 80 dB with a maximum input signal of 1 V and a noise floor around 0.1 to 1 mV (typical microwave oscilloscope). 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 (four diodes) or the sampling head (two or four 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 continuous-wave (CW) signal with bandwidth of a few Hz and thus can obtain with appropriate filtering a very good signal to noise ratio⁶.

With the sampling oscilloscope 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 20–50 GHz bandwidth). With this low average signal power (around a microwatt) 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).

⁶Remember that thermal noise is proportional to measurement bandwidth. Its density at room temperature is -174 dBm/Hz.

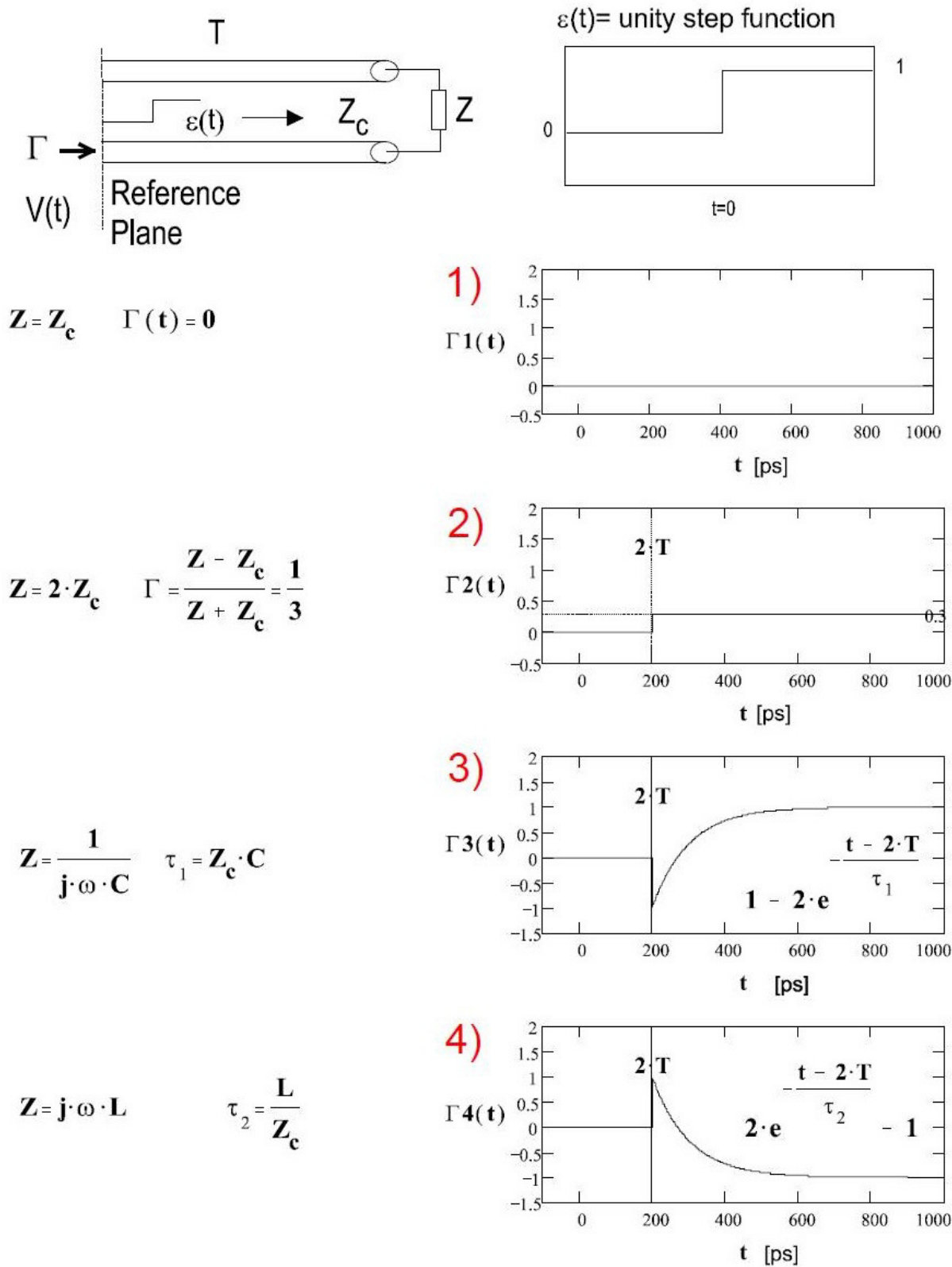


Fig. 37: Examples of an arbitrary impedance, measured in TDR

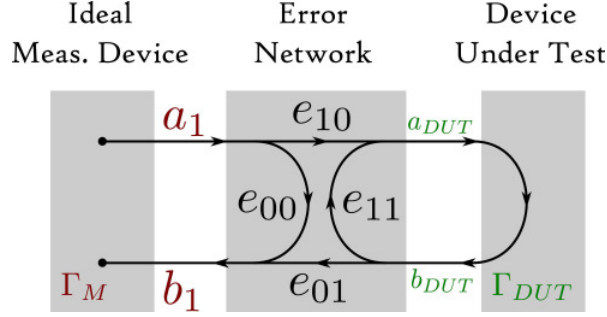


Fig. 38: Error model in a VNA. The parameters e_{xx} of the error network are determined by the calibration procedure and used to determine the real result (Γ_{DUT}) from the measured result (Γ_M).

A more detailed discussion about time-domain reflectometry with vector network analysers can be found in [20].

9.9 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 consists essentially in connecting, for a transmission measurement, the two ends of the test cables to each other and storing the amplitude and phase responses 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

$$S_{21\text{resp. cal.}} = \frac{S_{21\text{measured}}}{S_{21\text{reference}}}, \quad (38)$$

which is equivalent to a division in linear quantities. The phase is processed accordingly.

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 and match technique. This technique covers the three independent error sources mentioned above: finite directivity, generator mismatch and cable transfer function.

The VNA uses an internal error model, shown in Fig. 38. 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 Table 4.

We determine the unknowns of the error network by using three different, but known, calibration

Table 4: Interpretation of the error terms

Error term	Interpretation
e_{10}	Reflection tracking
e_{00}	Directivity
e_{11}	Test-port match

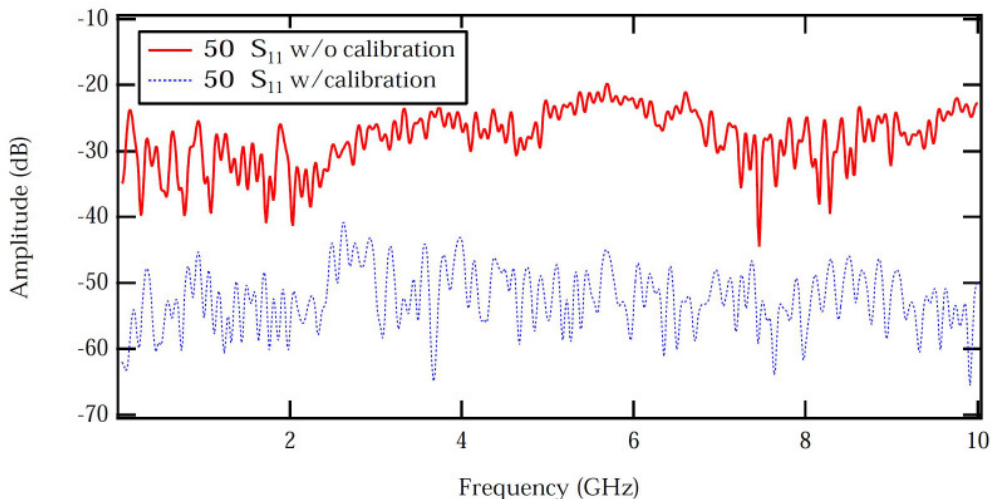


Fig. 39: S_{11} measurement of a 50Ω termination with and without calibration. The calibration provides 20 dB improvement over this frequency band.

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 a calibration kit), in electronic format.

They usually represent an open circuit, a short circuit and a match. In 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. 39 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 the range of several millidecibels.

So far, we have discussed the ‘response calibration’ and the complete one-port calibration. In order to carry out completely error-corrected transmission measurements, we need the ‘full two-port calibration’. In this case, the error model must be expanded to include errors at the receiving port, requiring calibration of each port using three known loads in reflection. Also, for transmission, we need two 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, an open and a short circuit are available (shown in Fig. 40a). As already mentioned for the

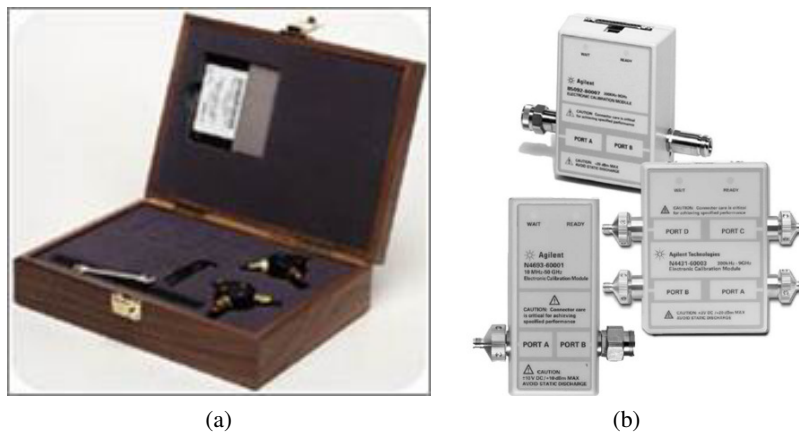


Fig. 40: Typical calibration kits for a VNA: (a) manual (open, short, match); (b) electronic

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 two-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 four-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. 40b), which switches the different standards automatically by communicating with the VNA. With this method, a full four-port calibration takes less than a minute. Again, as in the manual calibration method, the standards do not need to be perfect but well known, reproducible (switching) and stable. More details can be found in [17, 18].

9.10 1 dB compression point

A single tone source is connected to the input of an amplifier and its intensity gradually increased versus time. Monitoring the output of this amplifier, we notice a proportional dependence between input and output powers for small signal levels. This proportionality is referred to as the linear gain factor. For higher input signal levels, this will not hold any more 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 are typically the second and third harmonics and the distortion is referred to as 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. 41). When this deviation amounts to 1 dB, we have reached the 1 dB compression point.

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 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. 41.

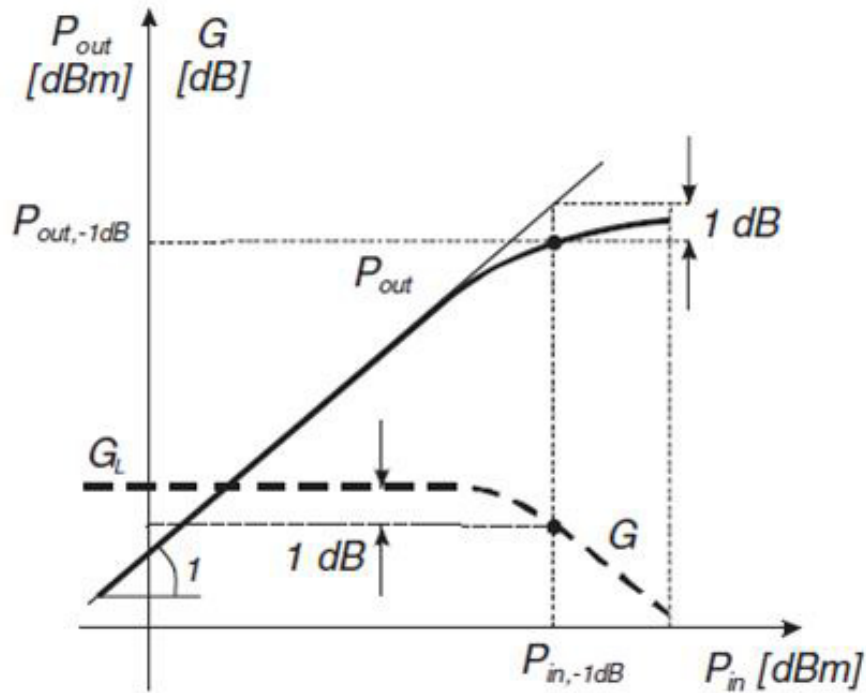


Fig. 41: Definition of the 1 dB compression point for an amplifier: input power, where the output power falls below 1 dB from its (linearly) predicted value.

10 INTRODUCTION TO THE SMITH CHART

The Smith chart is a very valuable and important tool that facilitates interpretation of S-parameter measurements. This section will give a brief overview of why and more importantly of how to use the chart. Its definition as well as an introduction of how to navigate inside the chart are illustrated. Useful examples show the broad possibilities for use of the chart in a variety of applications.

10.1 Voltage standing wave ratio

With the equipment at hand today, it has become rather easy to measure the reflection factor Γ even for complicated networks. In the ‘good old days’ though, this was done by measuring the electrical field strength⁷ at a coaxial measurement line with a slit at different positions in the axial direction (Fig. 42). A small electric field probe, protruding into the field region of the coaxial line near the outer conductor, was moved along the line. Its signal was picked up and displayed on a microvoltmeter after rectification via a microwave diode. While moving the probe, field maxima and minima as well as their position and spacing could be found. From this the reflection factor Γ and the voltage standing wave ratio (VSWR or SWR) could be determined using the following definitions:

- Γ is defined as the ratio of the electrical field strength E of the reflected wave over the forward-travelling wave:

$$\Gamma = \frac{E \text{ of reflected wave}}{E \text{ of forward-travelling wave}}. \quad (39)$$

- The VSWR is defined as the ratio of maximum to minimum measured voltages:

$$\text{VSWR} = \frac{U_{\max}}{U_{\min}} = \frac{1 + |\Gamma|}{1 - |\Gamma|}. \quad (40)$$

⁷The electrical field strength was used, since it can be measured considerably more easily than the magnetic field strength.

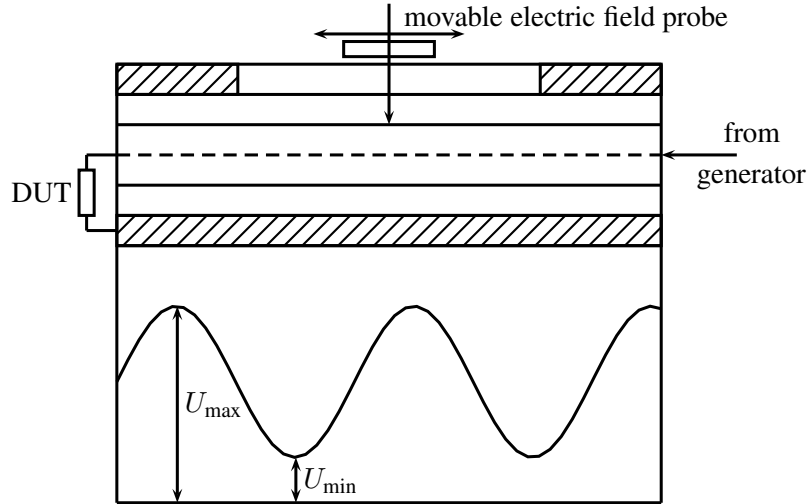


Fig. 42: Schematic view of a measurement set-up used to determine the reflection coefficient as well as the voltage standing wave ratio of a device under test (DUT) [21].

Although today these measurements are far easier to conduct, the definitions of the aforementioned quantities are still valid. Also, their importance has not diminished in the field of microwave engineering and so the reflection coefficient as well as the VSWR are still a vital part of the everyday life of a microwave engineer be it for simulations or measurements.

11 DEFINITION OF THE SMITH CHART

The Smith chart [22] provides a graphical representation of Γ that permits the determination of quantities such as the VSWR or the terminating impedance of a device under test (DUT). It uses a bilinear Moebius transformation, projecting the complex impedance plane onto the complex Γ plane:

$$\Gamma = \frac{Z - Z_0}{Z + Z_0} \quad \text{with} \quad Z = R + jX. \quad (41)$$

As can be seen in Fig. 43, the half-plane with positive real part of impedance Z is mapped onto the interior of the unit circle of the Γ plane.

11.1 Properties of the transformation

In general, this transformation has two main properties:

- generalized circles are transformed into generalized circles (note that a straight line is nothing else than a circle with infinite radius and is therefore mapped as a circle in the Smith chart);
- angles are preserved locally.

Figure 44 illustrates how certain basic shapes transform from the impedance to the Γ planes.

11.2 Normalization

The Smith chart is usually normalized to a terminating impedance Z_0 (= real):

$$z = \frac{Z}{Z_0}. \quad (42)$$

This leads to a simplification of the transform:

$$\Gamma = \frac{z - 1}{z + 1} \quad \Leftrightarrow \quad z = \frac{1 + \Gamma}{1 - \Gamma}. \quad (43)$$

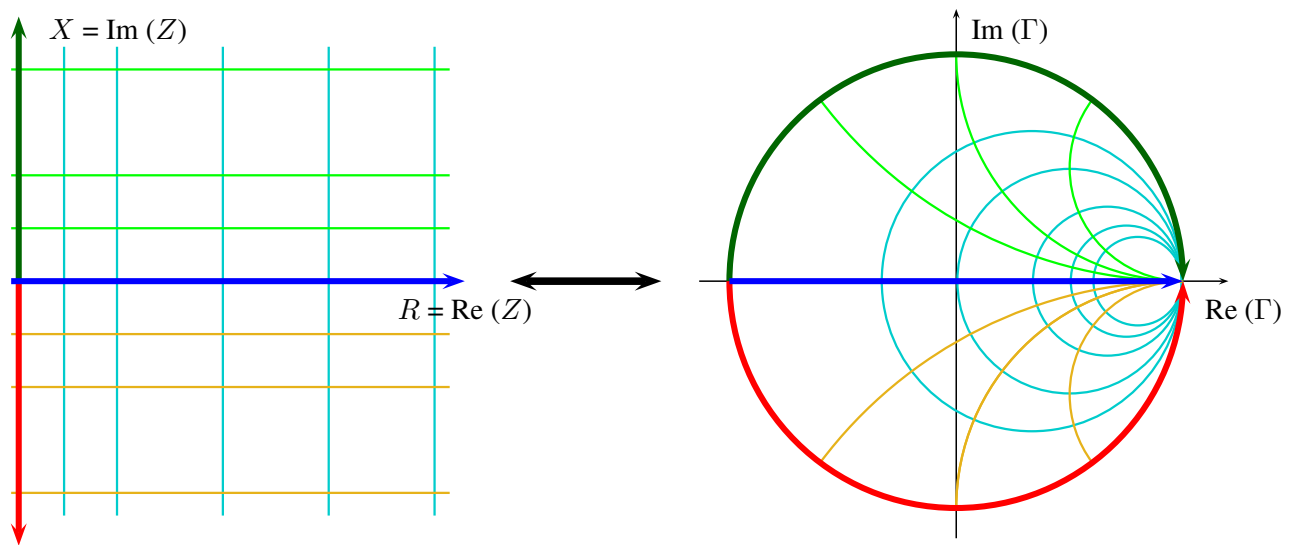


Fig. 43: Illustration of the Moebius transform from the complex impedance plane to the Γ plane commonly known as Smith chart.

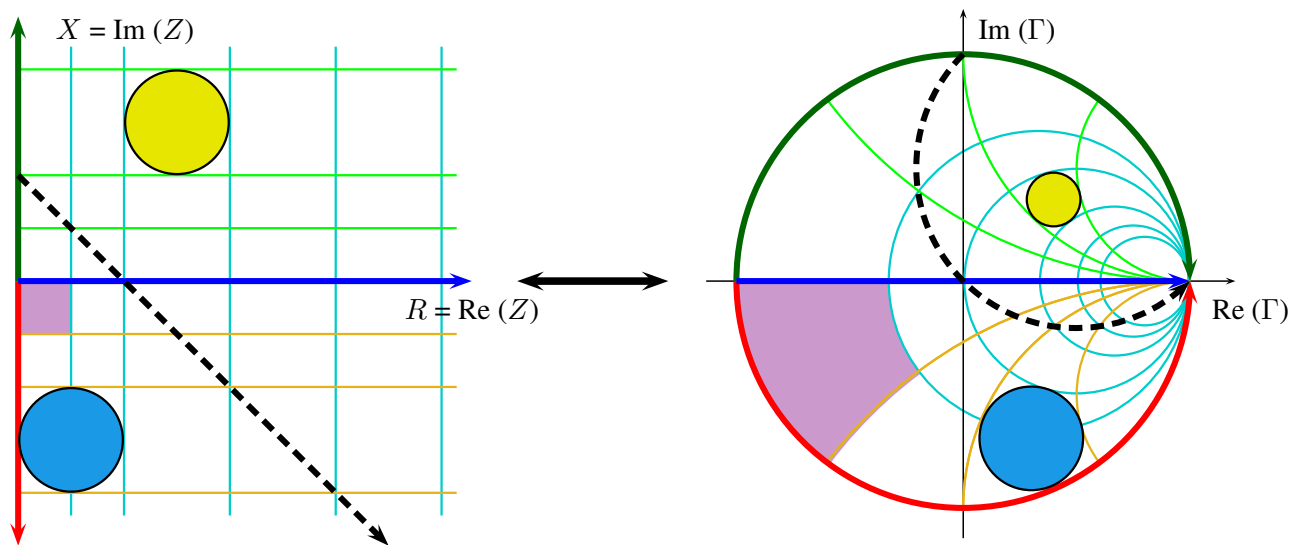


Fig. 44: Illustration of the transformation of basic shapes from the Z to the Γ planes

Although $Z = 50 \Omega$ is the most common reference impedance (characteristic impedance of coaxial cables) and many applications use this normalization, any other real and positive value is possible. *Therefore, it is crucial to check the normalization before using any chart.*

Commonly used charts that map the impedance plane onto the Γ plane always look confusing at first, as many circles are depicted (Fig. 45).

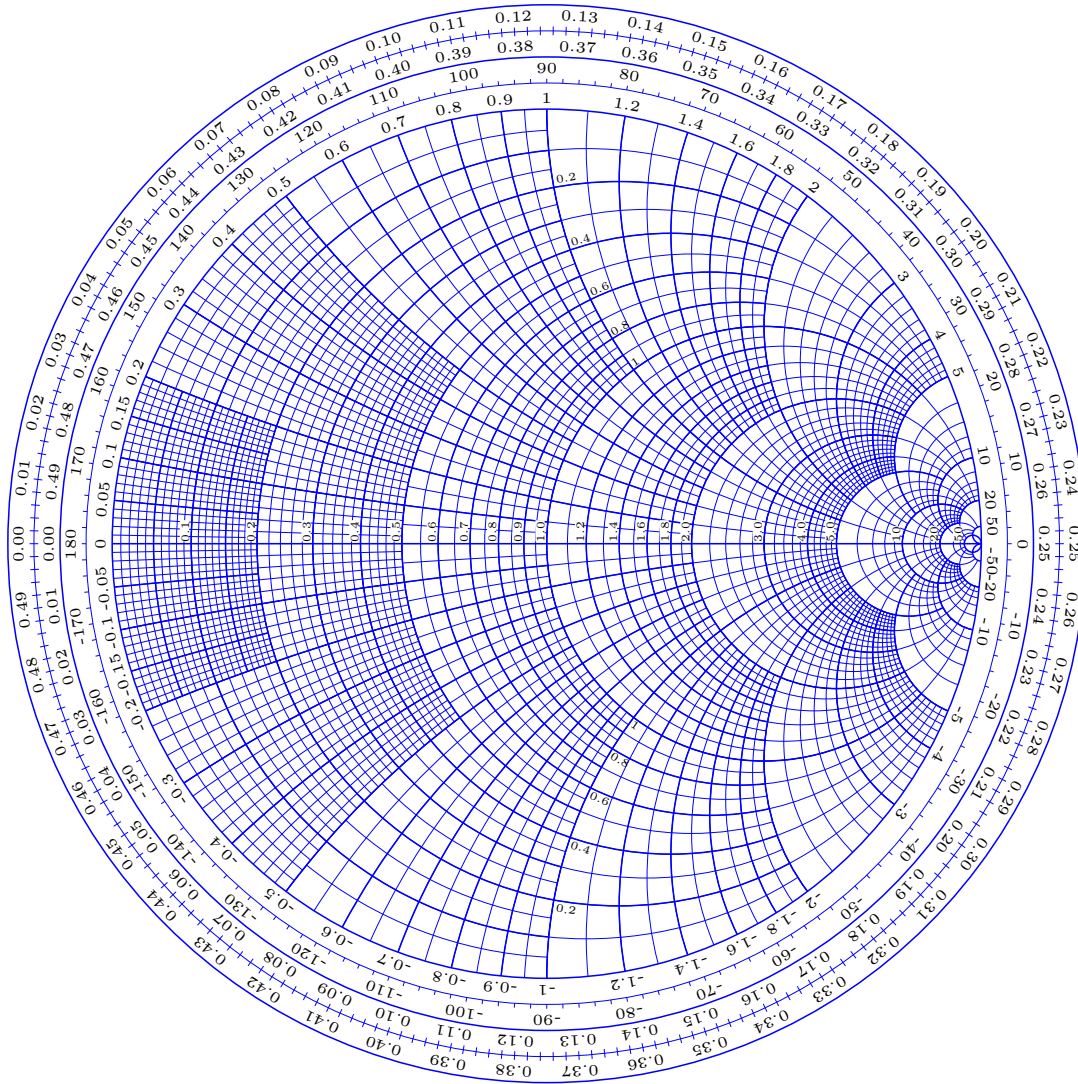


Fig. 45: Example of a commonly used Smith chart

11.3 Admittance plane

The Moebius transform that generates the Smith chart also provides a mapping of the complex admittance plane ($Y = \frac{1}{Z}$ or normalized $y = \frac{1}{z}$) into the same chart:

$$\Gamma = -\frac{y-1}{y+1} = -\frac{Y-Y_0}{Y+Y_0} = -\frac{1/Z-1/Z_0}{1/Z+1/Z_0} = \frac{Z-Z_0}{Z+Z_0} = \frac{z-1}{z+1}. \quad (44)$$

Using this transformation, the result is the same chart, but mirrored at the centre of the Smith chart (Fig. 46). Often both mappings, the admittance and the impedance planes, are combined into one chart, which looks even more confusing. For reasons of simplicity all illustrations in this article will use only the mapping from the impedance to the Γ planes.

12 NAVIGATION IN THE SMITH CHART

The representation of circuit elements in the Smith chart is discussed in this section starting with the important points inside the chart. Then several examples of circuit elements will be given and their representation in the chart will be illustrated.

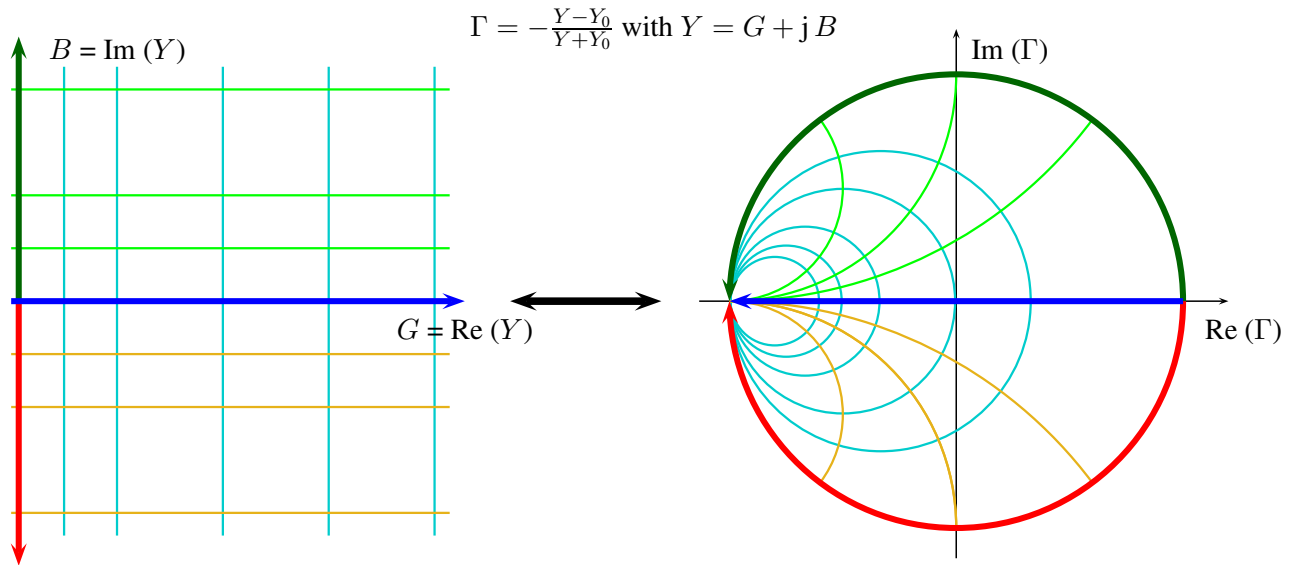


Fig. 46: Mapping of the admittance plane into the Γ plane

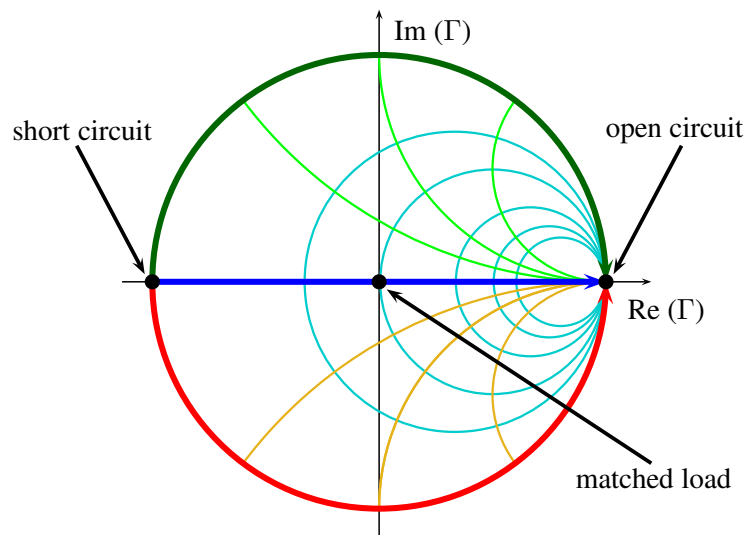


Fig. 47: Important points in the Smith chart

12.1 Important points

There are three important points in the chart:

1. Open circuit with $\Gamma = 1, z \rightarrow \infty$.
2. Short circuit with $\Gamma = -1, z = 0$.
3. Matched load with $\Gamma = 0, z = 1$.

They all are located on the real axis at the beginning, the end and the centre of the circle (Fig. 47). The upper half of the chart is inductive, since it corresponds to the positive imaginary part of the impedance. The lower half is capacitive as it is corresponding to the negative imaginary part of the impedance.

Concentric circles around the diagram centre represent constant reflection factors (Fig. 48). Their radius is directly proportional to the magnitude of Γ ; therefore, a radius of 0.5 corresponds to reflection

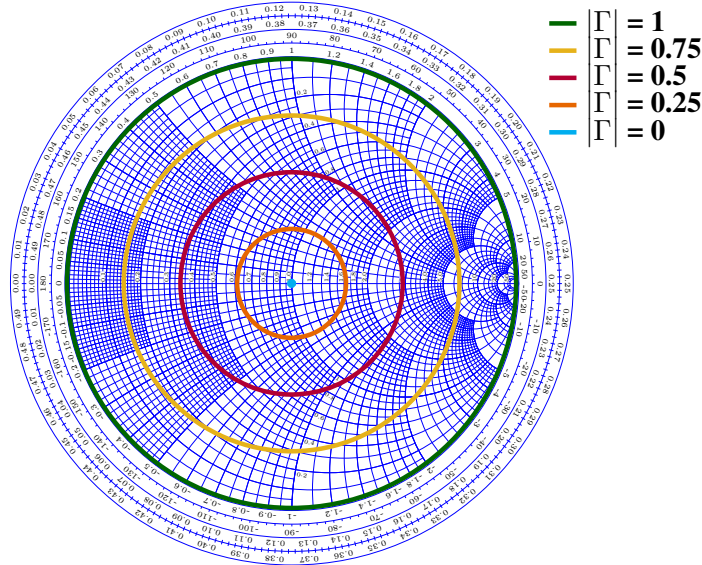


Fig. 48: Illustration of circles representing a constant reflection factor

of 3 dB (half of the signal is reflected) whereas the outermost circle (radius = 1) represents full reflection. Therefore, matching problems are easily visualized in the Smith chart, since a mismatch will lead to a reflection coefficient larger than 0 (see Eq. (45)).

$$\text{Power into the load} = \text{forward power} - \text{reflected power: } P = \frac{1}{2} (|a|^2 - |b|^2) = \frac{|a|^2}{2} (1 - |\Gamma|^2). \quad (45)$$

In Eq. (45), the European notation⁸ is used, where power = $\frac{|a|^2}{2}$. Furthermore, $(1 - |\Gamma|^2)$ corresponds to the mismatch loss.

Although only the mapping of the impedance plane to the Γ plane is used, one can easily use it to determine the admittance, since

$$\Gamma\left(\frac{1}{z}\right) = \frac{\frac{1}{z} - 1}{\frac{1}{z} + 1} = \frac{1 - z}{1 + z} = \left(\frac{z - 1}{z + 1}\right) \text{ or } \Gamma\left(\frac{1}{z}\right) = -\Gamma(z). \quad (46)$$

In the chart this can be visualized by rotating the vector of a certain impedance by 180° (Fig. 49).

12.2 Adding impedances in series and parallel (shunt)

A lumped element with variable impedance connected in series is an example of a simple circuit. The corresponding signature of such a circuit for a variable inductance and a variable capacitor is a circle. Depending on the type of impedance, this circle is passed through clockwise (inductance) or anticlockwise (Fig. 50). If a lumped element is added in parallel, the situation is the same as for an element connected in series mirrored by 180° (Fig. 51). This corresponds to taking the same points in the admittance mapping. Summarizing both cases, one ends up with a simple rule for navigation in the Smith chart:

For elements connected in series use the circles in the impedance plane. Go clockwise for an added inductance and anticlockwise for an added capacitor. For elements in parallel use the circles in the admittance plane. Go clockwise for an added capacitor and anticlockwise for an added inductance.

This rule can be illustrated as shown in Fig. 52.

⁸The commonly used notation in the USA is power = $|a|^2$. These conventions have no impact on S-parameters but they are relevant for absolute power calculation. Since this is rarely used in the Smith chart, the definition used is not critical for this

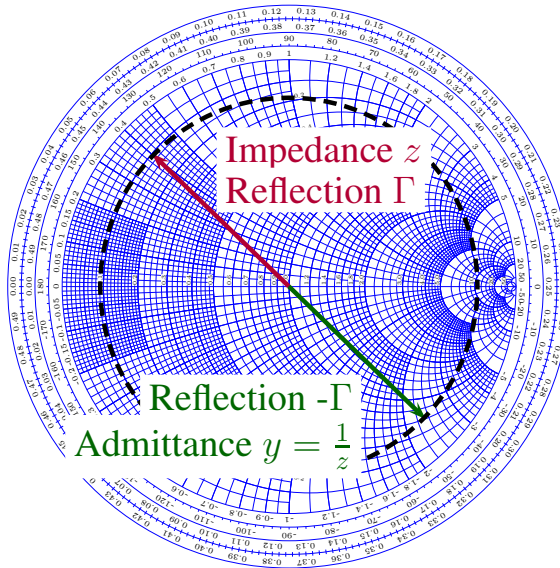


Fig. 49: Conversion of an impedance to the corresponding emittance in the Smith chart

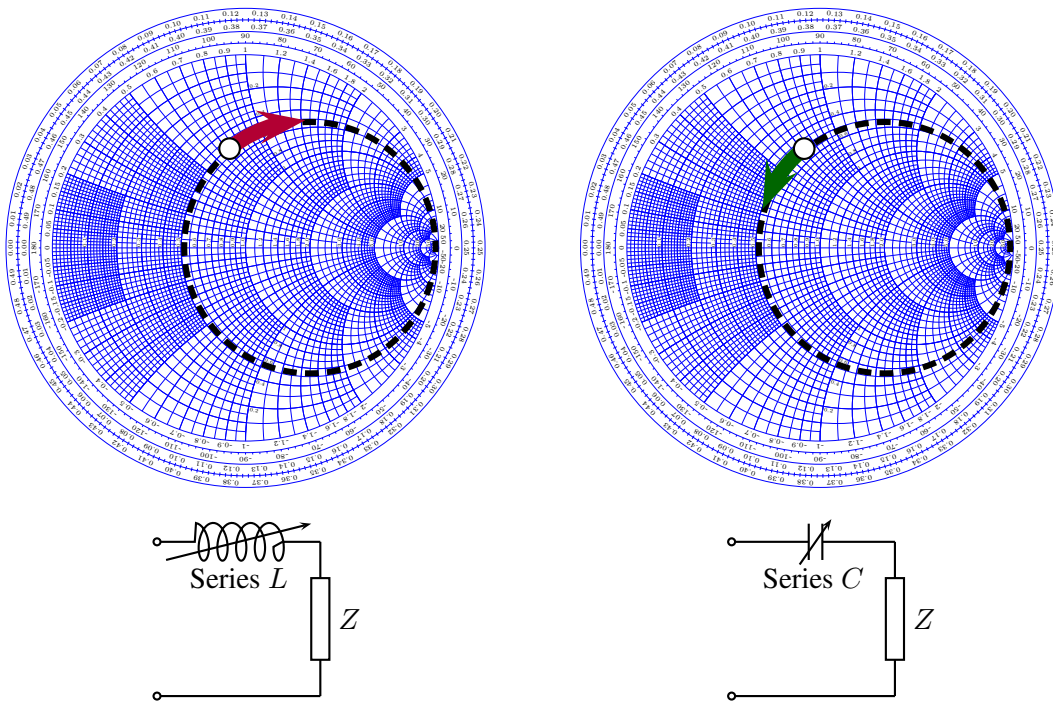


Fig. 50: Traces of circuits with variable impedances connected in series

12.3 Impedance transformation by transmission line

The S-matrix of an ideal, lossless transmission line of length l is given by

$$S = \begin{bmatrix} 0 & e^{-j\beta l} \\ e^{-j\beta l} & 0 \end{bmatrix}, \quad (47)$$

where $\beta = \frac{2\pi}{\lambda}$ is the propagation coefficient with the wavelength λ ($\lambda = \lambda_0$ for $\epsilon_r = 1$).

article.

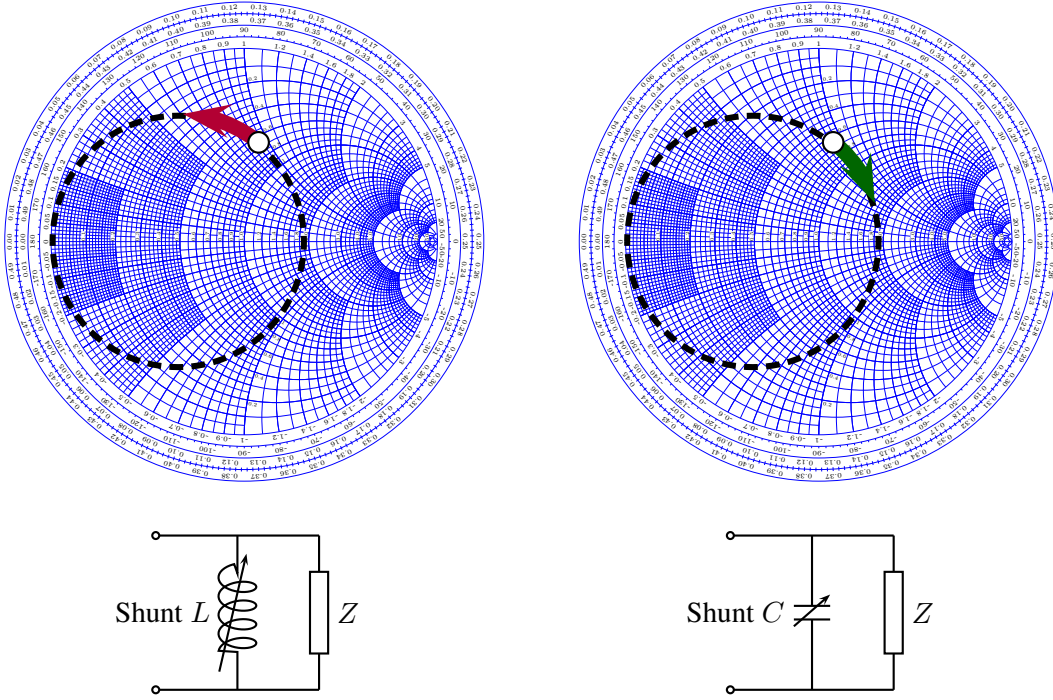


Fig. 51: Traces of circuits with variable impedances connected in parallel

When adding a piece of coaxial line, we turn clockwise on the corresponding circle leading to a transformation of the reflection factor Γ_{load} (without line) to the new reflection factor $\Gamma_{\text{in}} = \Gamma_{\text{load}}e^{-j2\beta l}$. Graphically speaking, this means that the vector corresponding to Γ_{in} is rotated clockwise by an angle of $2\beta l$ (Fig. 53).

The peculiarity of a transmission line is that it behaves as an inductance, a capacitor or a resistor depending on its length. The impedance of such a line (if lossless) is given by

$$Z_{\text{in}} = jZ_0 \tan(\beta l). \quad (48)$$

The function in Eq. (48) has a pole at a transmission line length of $\lambda/4$ (Fig. 54). Therefore, adding a transmission line with this length results in a change of Γ by a factor -1 :

$$\Gamma_{\text{in}} = \Gamma_{\text{load}}e^{-j2\beta l} = \Gamma_{\text{load}}e^{-j2(\frac{2\pi}{\lambda})l} \stackrel{l=\frac{\lambda}{4}}{=} \Gamma_{\text{load}}e^{-j\pi} = -\Gamma_{\text{load}}. \quad (49)$$

Again, this is equivalent to changing the original impedance z to its admittance $1/z$ or the clockwise movement of the impedance vector by 180° . Especially when starting with a short circuit (at -1 in the Smith chart), adding a transmission line of length $\lambda/4$ transforms it into an open circuit (at $+1$ in the Smith chart).

12.4 Examples of different two-ports

In general, the reflection coefficient when looking through a two-port Γ_{in} is given via the S-matrix of the two-port and the reflection coefficient of the load Γ_{load} :

$$\Gamma_{\text{in}} = S_{11} + \frac{S_{12}S_{21}\Gamma_{\text{load}}}{1 - S_{22}\Gamma_{\text{load}}}. \quad (50)$$

In general, the outer circle of the Smith chart as well as its real axis are mapped to other circles and lines.

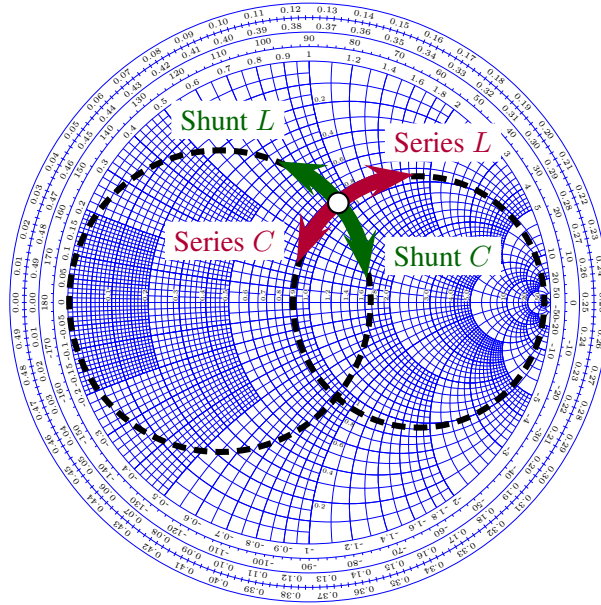


Fig. 52: Illustration of navigation in the Smith chart when adding lumped elements

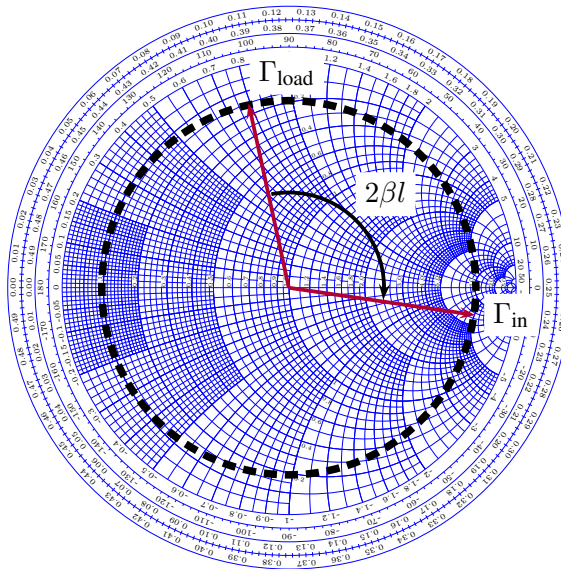


Fig. 53: Illustration of adding a transmission line of length l to an impedance

In the following three examples, different two-ports are given along with their S-matrix, and their representation in the Smith chart is discussed. For illustration, a simplified Smith chart consisting of the outermost circle and the real axis only is used for reasons of simplicity.

12.4.1 Transmission line $\lambda/16$

The S-matrix of a $\lambda/16$ transmission line is

$$S = \begin{bmatrix} 0 & e^{-j\frac{\pi}{8}} \\ e^{-j\frac{\pi}{8}} & 0 \end{bmatrix} \quad (51)$$

with the resulting reflection coefficient

$$\Gamma_{\text{in}} = \Gamma_{\text{load}} e^{-j\frac{\pi}{4}}. \quad (52)$$

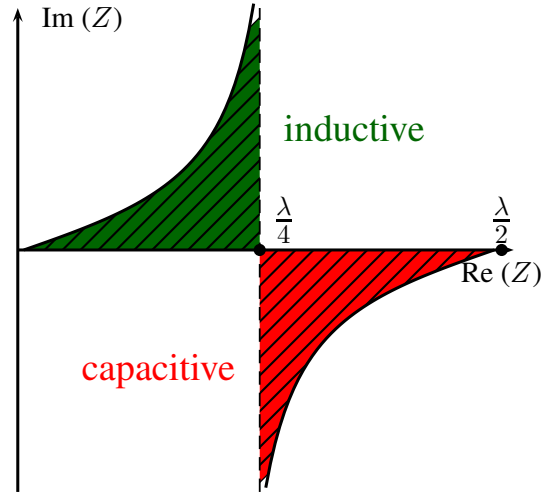


Fig. 54: Impedance of a transmission line as a function of its length l

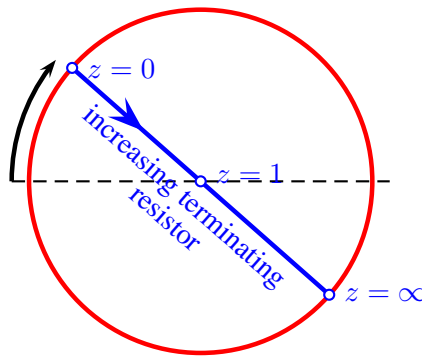


Fig. 55: Rotation of the reference plane of the Smith chart when adding a transmission line

This corresponds to a rotation of the real axis of the Smith chart by an angle of 45° (Fig. 55) and hence a change of the reference plane of the chart (Fig. 55). Consider, for example, a transmission line terminated by a short and hence $\Gamma_{\text{load}} = -1$. The resulting reflection coefficient is then equal to $\Gamma_{\text{in}} = e^{-j\frac{\pi}{4}}$.

12.4.2 Attenuator 3 dB

The S-matrix of an attenuator is given by

$$\mathbf{S} = \begin{bmatrix} 0 & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & 0 \end{bmatrix}. \quad (53)$$

The resulting reflection coefficient is

$$\Gamma_{\text{in}} = \frac{\Gamma_{\text{load}}}{2}. \quad (54)$$

In the Smith chart, the connection of such an attenuator causes the outermost circle to shrink to a radius of 0.5^9 (Fig. 56).

12.4.3 Variable load resistor

Adding a variable load resistor ($0 < z < \infty$) is the simplest case that can be depicted in the Smith chart. It means moving through the chart along its real axis (Fig. 57).

⁹An attenuation of 3 dB corresponds to a reduction by a factor 2 in power.

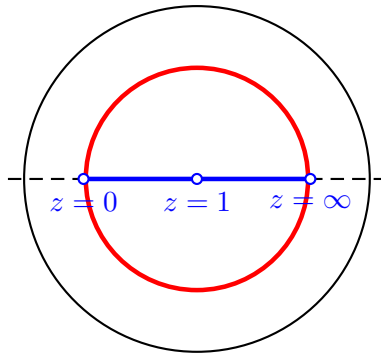


Fig. 56: Illustration of the appearance of an attenuator in the Smith chart

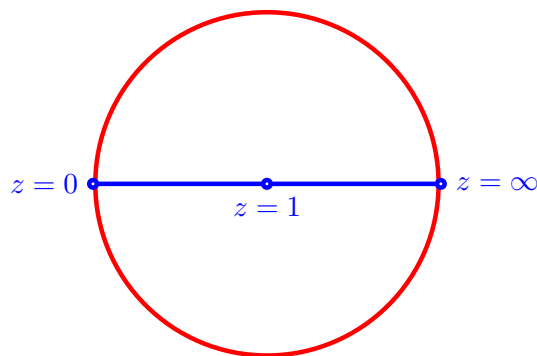


Fig. 57: A variable load resistor in the simplified Smith chart. Since the impedance has a real part only, the signal remains on the real axis of the Γ plane.

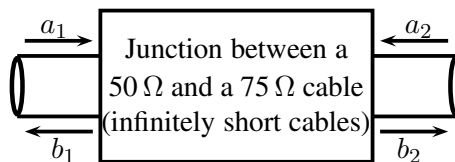


Fig. 58: Illustration of the junction between a coaxial cable with 50Ω characteristic impedance and another with 75Ω characteristic impedance, respectively. Infinitely short cables are assumed – only the junction is considered.

13 EXAMPLES FOR APPLICATIONS OF THE SMITH CHART

In this section two practical examples of common problems are given, where the use of the Smith chart greatly facilitates their solution.

13.1 A step in characteristic impedance

Consider a junction between two infinitely short cables, one with a characteristic impedance of 50Ω and the other with 75Ω (Fig. 58).

The waves running into each port are denoted with a_i ($i = 1, 2$), whereas the waves coming out of

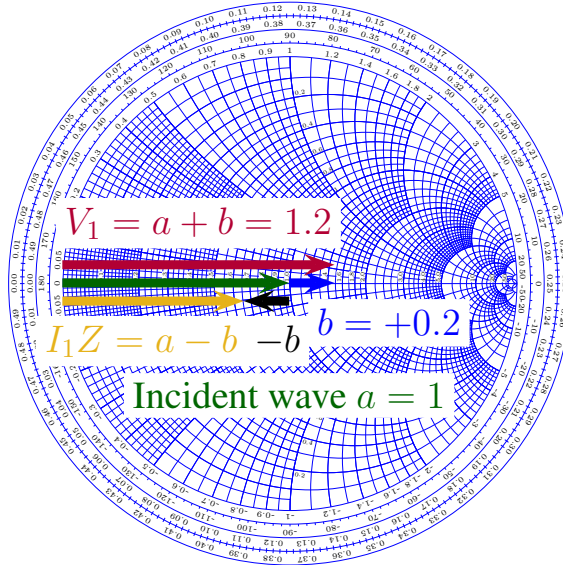


Fig. 59: Visualization of the two-port formed by the two cables of different characteristic impedances

every point are denoted with b_i . The reflection coefficient for port 1 is then calculated as

$$\Gamma_1 = \frac{Z - Z_1}{Z + Z_1} = \frac{75 - 50}{75 + 50} = 0.2. \quad (55)$$

Thus, the voltage of the reflected wave at port 1 is 20% of the incident wave ($a_2 = a_1 \cdot 0.2$) and the reflected power at port 1 is 4%¹⁰. From conservation of energy, the transmitted power has to be 96% and it follows that $b_2^2 = 0.96$.

A peculiarity here is that the transmitted energy is *higher* than the energy of the incident wave, since $E_{\text{incident}} = 1$, $E_{\text{reflected}} = 0.2$ and therefore $E_{\text{transmitted}} = E_{\text{incident}} + E_{\text{reflected}} = 1.2$. The transmission coefficient t is thus $t = 1 + \Gamma$. Also, note that this structure is not symmetric ($S_{11} \neq S_{22}$), but only reciprocal ($S_{21} = S_{12}$).

The visualization of this structure in the Smith chart is easy, since all impedances are real and thus all vectors are located on the real axis (Fig. 59).

13.2 Determination of the Q factors of a cavity

One of the most common cases where the Smith chart is used is the determination of the quality factor of a cavity. This section is dedicated to the illustration of this task.

A cavity can be described by a parallel RLC circuit (Fig. 60), where the resonance condition is given when

$$\omega L = \frac{1}{\omega C}. \quad (56)$$

This leads to the resonance frequency of

$$\omega_{\text{res}} = \frac{1}{\sqrt{LC}} \quad \text{or} \quad f_{\text{res}} = \frac{1}{2\pi} \frac{1}{\sqrt{LC}}. \quad (57)$$

The impedance Z of such an equivalent circuit is given by

$$Z(\omega) = \frac{1}{\frac{1}{R} + j\omega C + \frac{1}{j\omega L}}. \quad (58)$$

¹⁰Power is proportional to Γ^2 and thus $0.2^2 = 0.04$.

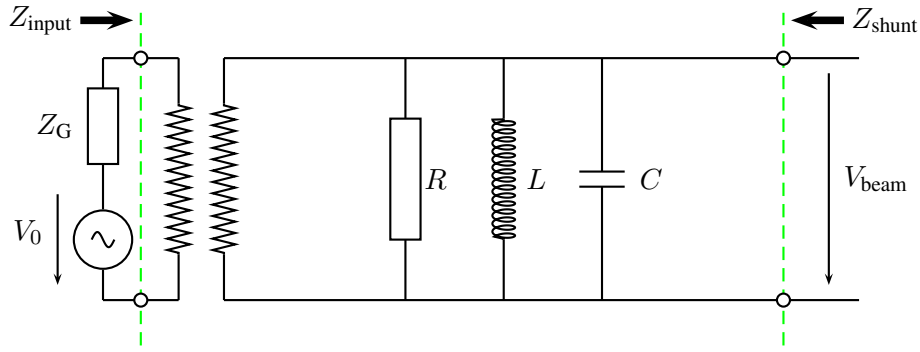


Fig. 60: The equivalent circuit that can be used to describe a cavity. The transformer is hidden in the coupling of the cavity ($Z \approx 1 \text{ M}\Omega$, seen by the beam) to the generator (usually $Z = 50 \text{ }\Omega$).

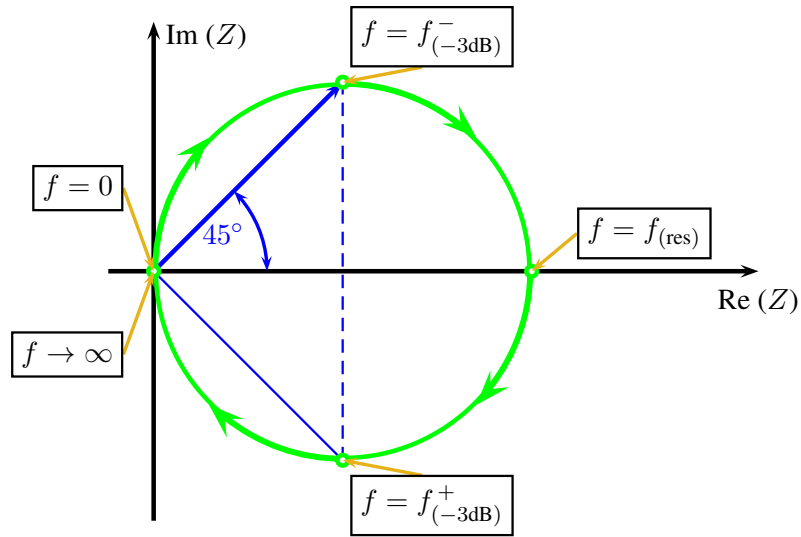


Fig. 61: Schematic drawing of the 3 dB bandwidth in the impedance plane

The 3 dB bandwidth Δf refers to the points where $\text{Re}(Z) = \text{Im}(Z)$, which correspond to two vectors with an argument of 45° (Fig. 61) and an impedance of $|Z_{(-3 \text{ dB})}| = 0.707R = R/\sqrt{2}$.

In general, the quality factor Q of a resonant circuit is defined as the ratio of the stored energy W over the energy dissipated in one cycle P :

$$Q = \frac{\omega W}{P}. \quad (59)$$

The Q factor for a resonance can be calculated via the 3 dB bandwidth and the resonance frequency:

$$Q = \frac{f_{\text{res}}}{\Delta f}. \quad (60)$$

For a cavity, three different quality factors are defined:

- Q_0 (unloaded Q): Q factor of the unperturbed system, i.e. the stand-alone cavity;
- Q_L (loaded Q): Q factor of the cavity when connected to generator and measurement circuits;
- Q_{ext} (external Q): Q factor that describes the degeneration of Q_0 due to the generator and diagnostic impedances.

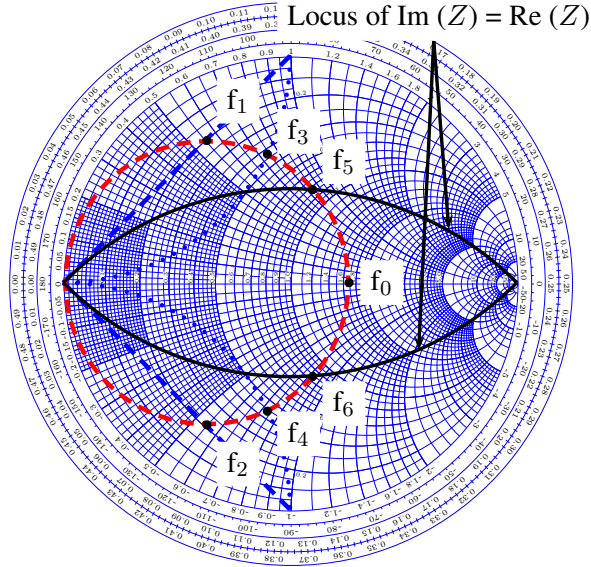


Fig. 62: Illustration of how to determine the different Q factors of a cavity in the Smith chart

All these Q factors are connected via a simple relation:

$$\frac{1}{Q_L} = \frac{1}{Q_0} + \frac{1}{Q_{\text{ext}}}. \quad (61)$$

The coupling coefficient β is then defined as

$$\beta = \frac{Q_0}{Q_{\text{ext}}}. \quad (62)$$

This coupling coefficient is not to be confused with the propagation coefficient of transmission lines, which is also denoted as β .

In the Smith chart, a resonant circuit shows up as a circle (Fig. 62, circle shown in the detuned short position). The larger the circle, the stronger the coupling. Three types of coupling are defined depending on the range of β (= size of the circle, assuming the circle is in the detuned short position):

- Undercritical coupling ($0 < \beta < 1$): the radius of the resonance circle is smaller than 0.25. Hence, the centre of the chart lies outside the circle.
- Critical coupling ($\beta = 1$): the radius of the resonance circle is exactly 0.25. Hence, the circle touches the centre of the chart.
- Overcritical coupling ($1 < \beta < \infty$): the radius of the resonance circle is larger than 0.25. Hence, the centre of the chart lies inside the circle.

In practice, the circle may be rotated around the origin due to the transmission lines between the resonant circuit and the measurement device.

From the different marked frequency points in Fig. 62 the 3 dB bandwidth and thus the quality factors Q_0 , Q_L and Q_{ext} can be determined as follows:

- The unloaded Q can be determined from f_5 and f_6 . The condition to find these points is $\text{Re}(Z) = \text{Im}(Z)$ with the resonance circle in the detuned short position.
- The loaded Q can be determined from f_1 and f_2 . The condition to find these points is $|\text{Im}(S_{11})| \rightarrow \text{max}$.

- The external Q can be calculated from f_3 and f_4 . The condition to determine these points is $Z = \pm j$.

To determine the points f_1 to f_6 with a network analyser, the following steps are applicable:

- f_1 and f_2 : set the marker format to $\text{Re}(S_{11}) + j \text{Im}(S_{11})$ and determine the two points where $\text{Im}(S_{11}) = \text{max}$.
- f_3 and f_4 : set the marker format to Z and find the two points where $Z = \pm j$.
- f_5 and f_6 : set the marker format to Z and locate the two points where $\text{Re}(Z) = \text{Im}(Z)$.

14 SUMMARY

In the present article several basic signal receiving techniques and devices used for this purpose were described. Advantages of several measurements using spectrum analysers and network analysers were presented. In the last part the definition of the Smith chart and its usage were discussed with several examples. All this information can give a good overview needed in the practical part of the CAS intermediate-level RF course. Within these practical exercises six test stages equipped with the most modern spectrum analysers and network analysers were prepared.

The following experiments were proposed:

14.1 Spectrum analyser test stand 1:

- Measurements of several types of modulation (AM, FM and PM) in the time and frequency domains.
- Superposition of AM and FM spectra (unequal carrier sidebands).
- Concept of a spectrum analyser: the superheterodyne method. Practice all the different settings (video bandwidth, resolution bandwidth etc). Advantage of FFT spectrum analysers.

14.2 Spectrum analyser test stand 2:

- Measurement of the TOI point of some amplifiers (intermodulation tests).
- Concept of noise-figure and noise-temperature measurements, testing a noise diode, the basics of thermal noise.
- Electromagnetic compatibility (EMC) measurements (e.g. analyse your cell-phone spectrum).
- Non-linear distortion in general concept and application of vector spectrum analysers, spectrogram mode.
- Measurement of the RF characteristic of a microwave detector diode (output voltage versus input power ... transition between regimes output voltage proportional to input power and output voltage proportional to input voltage).

14.3 Spectrum analyser test stand 3:

- Concept of noise-figure and noise-temperature measurements, testing a noise diode, the basics of thermal noise.
- Noise-figure measurements on amplifiers and also attenuators.
- The concept and meaning of excess noise ratio (ENR) numbers.
- Noise temperature of the fluorescent tubes in the room using a satellite receiver.

14.4 Network analyser test stand 1:

- Calibration of the vector network analyser.
- Navigation in the Smith chart.
- Application of the triple-stub tuner for matching.
- Measurements of the light velocity using a trombone (constant-impedance adjustable coaxial line) in the frequency domain.
- N -port ($N = 1-4$) S-parameter measurements for different reciprocal and non-reciprocal RF components.
- Self-made RF components: calculate, build and test your own attenuator (and then take it home).

14.5 Network analyser test stand 2:

- Measurements of the light velocity using a trombone (constant-impedance adjustable coaxial line) in the time domain.
- Two-port measurements for active RF components (amplifiers).
- A 1 dB compression point (power sweep).
- Beam transfer impedance measurements with the wire (button pick-up, stripline pick-up).

14.6 Network analyser test stand 3:

- Measurements of the characteristic cavity features (Smith-chart analysis).
- Cavity perturbation measurements (bead pull).
- Perturbation measurements using rectangular waveguides.
- Standing wave ratio (SWR) measurements using a waveguide measurement line and movable probe.

REFERENCES

- [1] G.D. Vendelin, A.M. Pavio and U.L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, second ed. (Wiley-Interscience, New Jersey, 2005), ISBN-10 0-471-41479-4.
- [2] F. Caspers, Proc. CERN Accelerator School, RF Engineering for Particle Accelerators, Oxford, UK, 1991, p.181.
- [3] M. Thumm, W. Wiesbeck and S. Kern, *Hochfrequenzmesstechnik* (Teubner, Stuttgart/Leipzig, 1998), ISBN 3-519-16360-8.
- [4] R.A. Witte, *Spectrum and Network Measurements* (Prentice-Hall, New Jersey, 1991), ISBN 0-13-826959-9.
- [5] W.O. Schleifer, *Hochfrequenz und Mikrowellenmesstechnik in der Praxis* (Hüthig, Heidelberg, 1981), ISBN 3-7785-0675-7.
- [6] B. Schiek and H.J. Sieveris, *Rauschen im Hochfrequenzschaltungen* (Hüthig, Heidelberg, 1984), ISBN 3-7785-2007-5.
- [7] P.C.L. Yip, *High Frequency Circuit Design and Measurement* (Chapman and Hall, London, 1990), ISBN 0-412-34160-3.
- [8] G. Evans and C.W. McLeisch, *RF-Radiometer Handbook* (Artech, Dedham, 1977), ISBN 0-89006-055-X.
- [9] F.R. Connor, *Noise* (Edward Arnold, London, 1973), ISBN 0-7131-3306-6.
- [10] F. Landstorfer and H. Graf, *Rauschprobleme der Nachrichtentechnik* (Oldenbourg, München, 1981), ISBN 3-486-24681-X.

- [11] O. Zinke and H. Brunswig, *Lehrbuch der Hochfrequenztechnik, Zweiter Band* (Springer, Berlin, 1974), ISBN 3-540-06245-9.
- [12] Agilent Technologies, Inc., *Fundamentals of RF and microwave noise figure measurements*, Agilent Application Note 57-1, 2010.
- [13] B. Schiek, *Messsysteme der Hochfrequenztechnik* (Hüthig, Heidelberg, 1984), ISBN 3-7785-1045-2.
- [14] F. Caspers, RF engineering basic concepts: S-parameters, CAS Proc., 2010, CERN Yellow Report CERN-2011-007, pp. 67-93.
- [15] J. Verspecht and D. Root, *Polyharmonic Distortion Modeling*, IEEE Microwave Magazine, Vol. 7, Issue 3, June 2006, pp. 44-57.
- [16] M. Thumm, W. Wiesbeck and S. Kern, *Hochfrequenzmesstechnik, Verfahren und Messsysteme* (Teubner, Stuttgart, 1998), ISBN 978-3519163602.
- [17] M. Hiebel, *Fundamentals of Vector Network Analysis* (Rohde & Schwarz, München, 2007), ISBN 3939837067.
- [18] Agilent Technologies, Inc., *Understanding the fundamental principles of vector network analysis*, Agilent Application Note AN 1287-1, 2000.
- [19] Anritsu Company, *Time domain measurements using vector network analyzers*, Anritsu Application Note No. 11410-00206, R , 2009.
- [20] Agilent Technologies, Inc., *Time domain analysis using a network analyzer*, Agilent Application Note 1287-12, 2012.
- [21] H. Meinke and F.-W. Gundlach, *Taschenbuch der Hochfrequenztechnik* (Springer, Berlin, 1992).
- [22] P. Smith, *Electronic Applications of the Smith Chart* (Noble Publishing, Atlanta, 2000), ISBN 1-884932-39-8.

15 APPENDIX: MEANING OF THE RULERS BELOW THE SMITH CHART

How to use the rulers that are often plotted underneath the Smith chart?

A commonly used set of rulers can be found below the Smith chart is shown in Fig. 63. There are four rulers, some with an upper and lower part, to quickly estimate and compare some important properties in terms of modulus values. For the following discussion lets split the upper three rulers at the line marked *CENTER* to a left and right part, each to be discussed separately. These rulers start at the *CENTER*, referring to the center of the Smith chart, and end at the left or right boundary, referring to the circular boundary of the Smith chart. The 4th ruler at the bottom is different, it starts at the left boundary *ORIGIN* and ends at the right boundary.

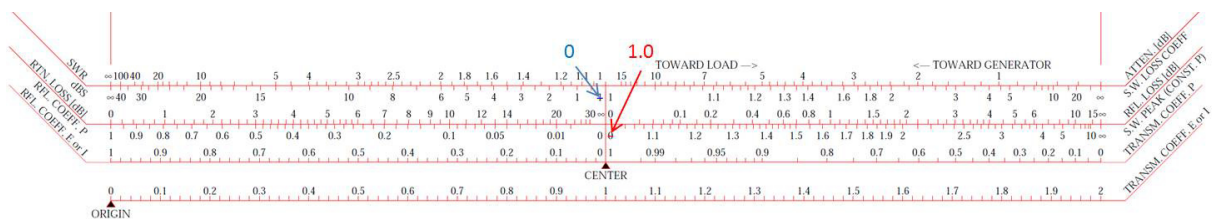


Fig. 63: Example for a set of rulers that can be found underneath the Smith chart (please note corrections in respect to the RF-course printouts)

First ruler, left/upper part in Fig. 64 is marked as *SWR* which mean actually *VSWR*, i.e. voltage standing wave ratio. It ranges between one – for the matched case (center of the Smith chart) and infinity – for total reflection (boundary of the Smith chart), respectively. The upper part is in linear scale, the lower part of this ruler is in dB, noted as *dBS* (dB referred to Standing Wave Ratio). Example: *SWR* = 10 corresponds to 20 *dBS*, *SWR* = 100 corresponds to 40 *dBS* [voltage ratios, not power ratios].

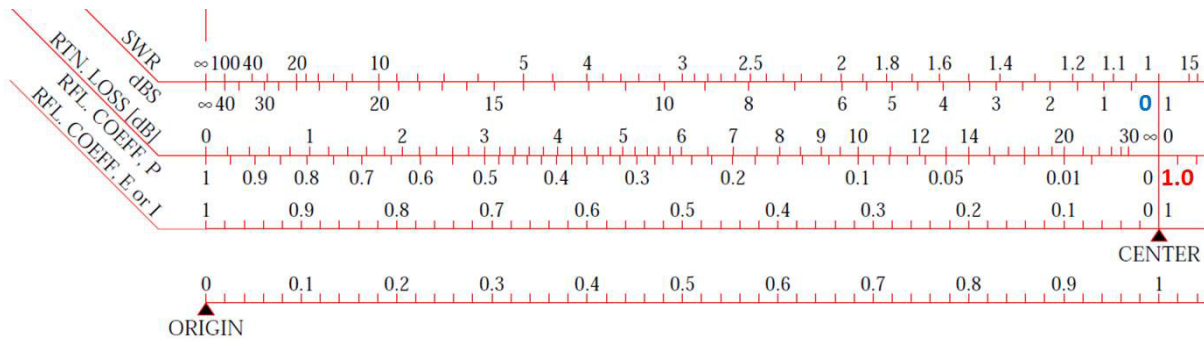


Fig. 64: Left part of the rulers usually plotted underneath the Smith Chart

Second ruler, left/upper part, marked as *RTN.LOSS* i.e. return loss in dB. This indicates the amount of reflected wave expressed in dB. Thus, in the center of SC nothing is reflected and the return loss is infinite. At the boundary we have full reflection, thus return loss is 0 dB. The lower part of the scale denoted as *RFL.COEFF.P* is a reflection coefficient in terms of POWER (proportional $|\Gamma|^2$). If there is no reflected power for the matched case locus is in the center of the Smith chart (SC). On the contrary, if normalized reflected power is equal to 1 locus is at the boundary.

Third ruler, left, marked as *RFL.COEFF.E or I* gives us the absolute value of the reflection coefficient in linear scale. Note that since we have the modulus we can refer it both to voltage or current as we have omitted the sign. Obviously in the center the reflection coefficient is zero, at the boundary it is one.

The fourth is a Voltage transmission coefficient. Note that the modulus of the voltage (and current) transmission coefficient has a range from zero, i.e. short circuit, to +2 (open = $1+\Gamma$ with $\Gamma = 1$). This ruler is only valid for $Z_{load} = \text{real}$, i.e. the case of a step in characteristic impedance of the coaxial line.

Third ruler, right (see Fig. 65) marked as *TRANSM.COEFF.P* refers to the transmitted power as a function of mismatch and displays essentially the relation $P_t = 1 - |\Gamma|^2$. Thus, in the center of the SC full match, all the power is transmitted. At the boundary we have total reflection and e.g. for a Γ value of 0.5 we see that 75 % of the incident power is transmitted.

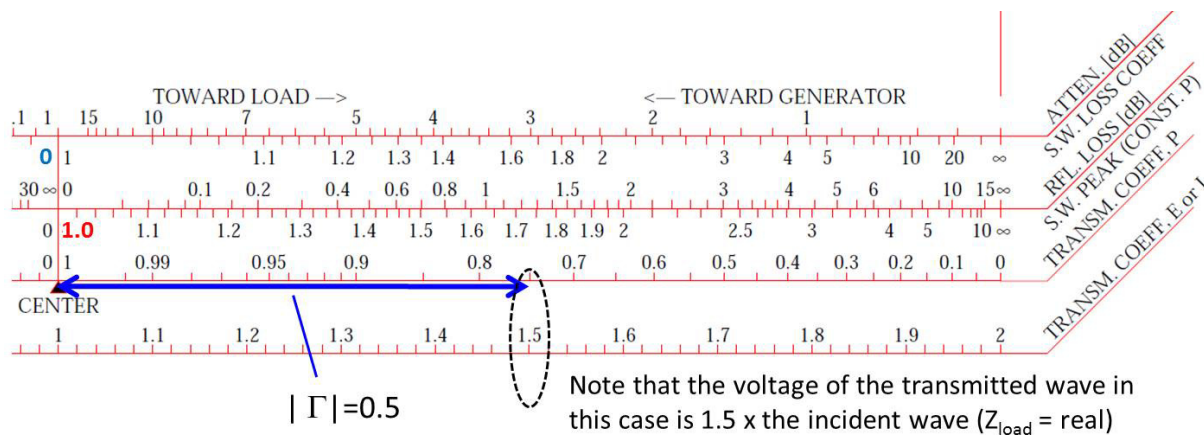


Fig. 65: Right part of the rulers usually plotted underneath the Smith Chart

Second ruler, right/upper part, denoted as *RFL.LOSS in dB* denotes reflection loss. This ruler refers to the loss in the transmitted wave, and should not be confounded with the return loss referring to the reflected wave. It displays the relation $P_t = 1 - |\Gamma|^2$ in dB. This ruler is nowadays rather not more in use.

Let us analyse an example from Fig. 66: $|\Gamma| = 1/\sqrt{2} = 0.707$, transmitted power = 50 % thus

loss = 50 % = 3 dB. Note that in the lowest ruler the voltage of the transmitted wave ($Z_{load} = \text{real}$) would be $V_t = 1.707 = 1 + 1/\sqrt{2}$ if referring to the voltage.

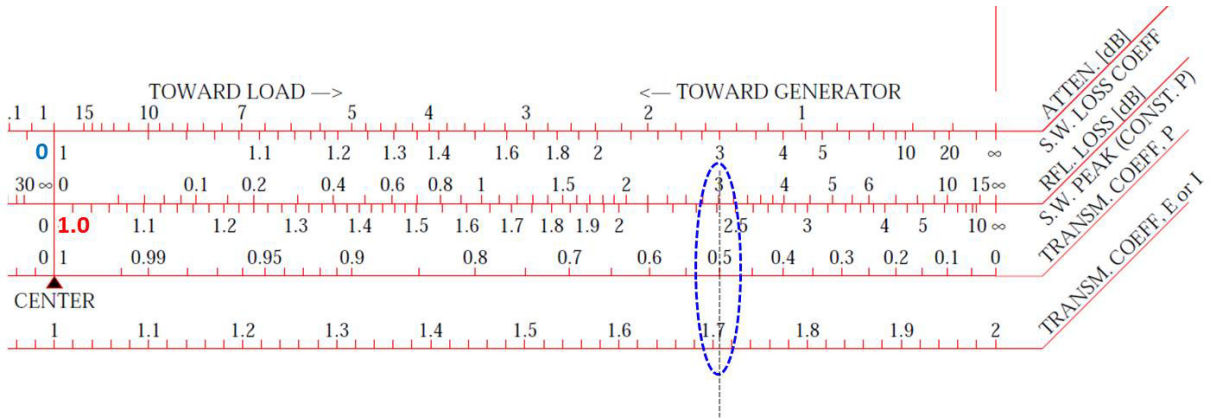


Fig. 66: Example for $|\Gamma| = 1/\sqrt{2} = 0.707$ and 50 % of transmitted power (i.e. 3 dB loss), see description in text

Finally, the First ruler, right/upper part, denoted as *ATTEN. in dB* assumes that one is measuring an attenuator or a lossy line which itself is terminated by an open or short circuit (full reflection). Thus the wave is traveling twice through the attenuator (forward and backward). The value of this attenuator can be between zero and some very high number corresponding to the matched case. The lower scale of first ruler displays the same situation just in terms of VSWR.

For the next example see Fig. 67: an 10 dB attenuator attenuates the reflected wave by 20 dB going forth and back and we get a reflection coefficient of $\Gamma = 0.1$. This correspond to the reflection of 10 % in voltage. Another example is 3 dB attenuator: for the forth and back transmission it gives 6 dB which correspond to half of the voltage. Table 5 is reprinted from an original paper of Phillip H. Smith [22] and summarizes reflection formulas discussed above.

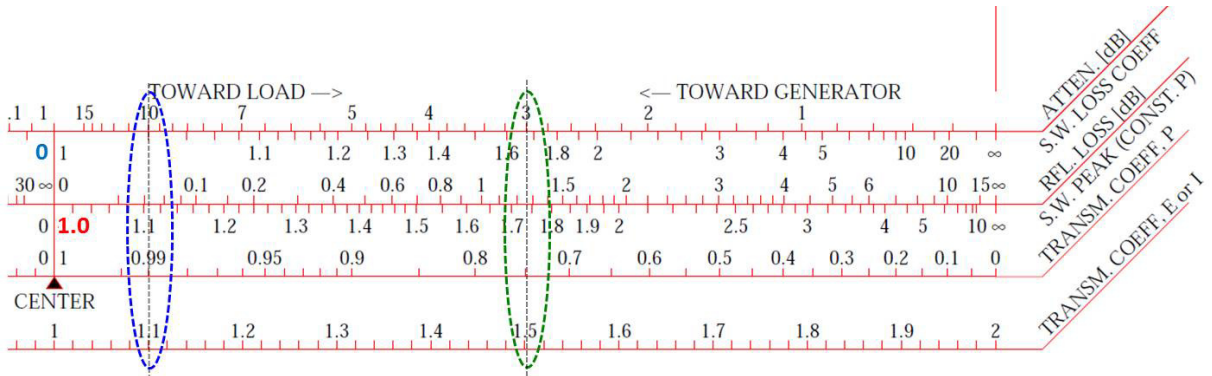


Fig. 67: Example for 10 dB and 3 dB attenuator, see description in text

Table 5: Reflection formulas

function	traveling waves	reflection coefficient	standing waves
VOLTAGE REFL. COEFF.	$\frac{r}{i}$	Γ	$\frac{S-1}{S+1}$
POWER REFL. COEF.	$(\frac{r}{i})^2$	Γ^2	$(\frac{S-1}{S+1})^2$
RETURN LOSS [dB]	$10 \cdot \log(\frac{i}{r})^2$	$-10 \cdot \log(\Gamma^2)$	$-10 \cdot \log(\frac{S-1}{S+1})^2$
REFLECTION LOSS [dB]	$10 \cdot \log(\frac{i^2}{i^2-r^2})$	$-10 \cdot \log(1 - \Gamma^2)$	$-10 \cdot \log[1 - (\frac{S-1}{S+1})^2]$
STDG. WAVE LOSS COEF.	$1 - \frac{[(i+r)/(i-r)]^2}{2[(i+r)/(i-r)]}$	$\frac{1-\Gamma+\Gamma^2-\Gamma^3}{1-\Gamma-\Gamma^2+\Gamma^3}$	$\frac{1+S^2}{2S}$
STDG. WAVE RATIO [dB]	$20 \cdot \log(\frac{i+r}{i-r})$	$20 \cdot \log(\frac{1+\Gamma}{1-\Gamma})$	$20 \cdot \log(S)$
MAX. OF STDG. WAVE	$(\frac{i+r}{i-r})^{1/2}$	$(\frac{1+\Gamma}{1-\Gamma})^{1/2}$	\sqrt{S}
MIN. OF STDG. WAVE	$(\frac{i-r}{i+r})^{1/2}$	$(\frac{1-\Gamma}{1+\Gamma})^{1/2}$	$\frac{1}{\sqrt{S}}$
STANDING WAVE RATIO	$\frac{i+r}{i-r}$	$\frac{1+\Gamma}{1-\Gamma}$	S
ATTENUATION [dB]	$-10 \cdot \log(\frac{r}{i})$	$-10 \cdot \log(\Gamma)$	$-10 \cdot \log(\frac{S-1}{S+1})$

whereas: i = incident wave amplitude, r = reflected wave amplitude, Γ = reflection coefficient,
 $S \equiv$ SWR = voltage standing wave ratio.