S-26.3120 Lecture *Microwave Measurement Techniques*

Introduction

This handout on microwave measurement techniques supplements the lectures of the RAD-Department course S-26.3120 *Radio Engineering, Laboratory Course*. It is based on the book: A. Lehto, A. Räisänen *Mikroaaltomittaustekniikka.* It deals with how to perform sufficiently reliable and accurate microwave measurements. We get closer to this goal when we understand well the measured variables, the design and the operation of the measurement devices, as well as the measurement error sources. In this course the learning goal is therefore to be able to describe measurement devices and their functions, such as microwave signal generation and detection, to explain measurements methods for impedance, frequency and spectrum, power and attenuation, non-linearity and noise, as well as for antenna characteristics such as gain, efficiency, radiation patterns, and finally - through the four laboratories - to acquire routine in performing two-port S-parameter measurements with a VNA, as well as spectrum and noise measurements with a spectrum analyzer, incl. measurement calibration routines, measurement analysis incl. measurement uncertainty estimation. The course also offers the possibility to familiarize with microwave-based measurements of material properties, EMC measurements, time-domain reflectometry, and finally to apply the design of a microwave amplifier in the last of the four laboratories.

This course's prerequisite is the course S-26.2100/2110 *Fundamentals of Radio Engineering*, which means it is expected that the student masters the basic concepts of: transmission lines and waveguides (coaxial cables, microstrip lines), impedance matching (scattering parameters, Smith diagram, gain of a two-port), passive and active microwave devices (attenuators, filters, oscillators, amplifiers, mixers), and of antennas (far field, directivity, efficiency) and radio systems (transmitters, receivers, PLL). For the fourth laboratory (in spring), it is necessary to acquire knowledge on microstrip-circuit design from the course S-26.3100 *RF and Microwave Engineering* held in parallel to the labs in the spring term.

Microwaves are radio waves, which cover the electromagnetic spectrum of the frequency range 300 MHz – 30 GHz. Below this range we talk of RF waves. Sometimes the border is considered to be at 1 GHz. Millimeter waves cover the frequency rage 30 – 300 GHz and sub-millimeter waves the range 300 – 3000 GHz. Still shorter waves are infrared waves. From the technical point of view, however, these boundaries are often quite artificial. The measuring equipment and techniques we deal with in this handout in many cases work also at RF- and millimeter-wave ranges, or even at sub-millimeter-wave range. In microwave and RF applications, such as mobile communications, wireless local-area networks, radar technology, radio navigation, and a variety of sensors, the measurement of the circuits needed in the devices are an integral part of both the design process and production. Measurements are needed also when installing and servicing the equipment. Naturally, in scientific applications and basic research accurate measurements are very important.

Microwave and RF measurements differ in many respects from low-frequency measurements. Voltages and currents are not constant along the transmission lines, i.e. measured quantities are rather properties of the electromagnetic fields. Since the dimensions of the circuits are normally in the order of a wavelength, impedance and scattering parameters depend on location. Therefore, the measurement reference locations/planes must be clearly defined. This is why microwave measurements are often difficult and time-consuming, and sometimes the final desired result is obtained only after complicated calculations.

In the 1960s (sometimes still today), measurements were made manually with fixed-frequency signal sources. For example, the impedance was measured through the standing-wave pattern inside a slotted line (see Fig. 1), and the comparison with the short circuit had to be done for each frequency separately. Nowadays we use wideband sweep frequency generators, which synthesize the signal with help of a low-frequency quartz oscillator. E.g. impedance can be measured with a vector network analyzer (VNA), which corrects the measurement result with help of an error model that is based on a calibration measurement.

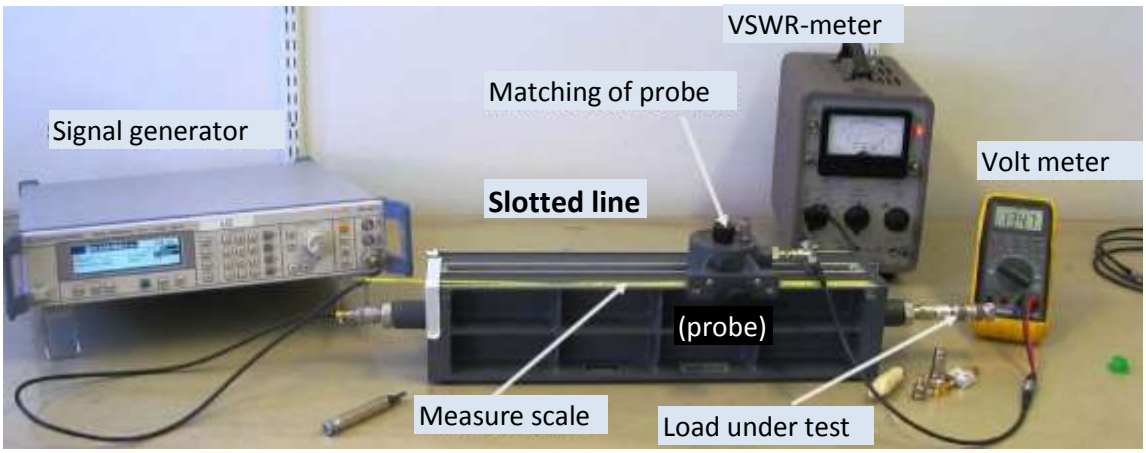

Figure 1. Single-frequency impedance measurement setup with a slotted line.

Modern measuring device, such as the synthesized signal generator shown in Fig. 2, are at the same time computers, programmable and remote-controllable, that can be connected to a PC and a wide range of peripherals. A large digital display shows the device status. Different device functions are typically controlled with soft-key buttons next to the screen, whose function depend on the selected menu. Settings can be changed by entering values on the keypad or by turning the analog tuning wheel. Often, presets are available that correspond to different communication standards.

Figure 2. State-of-the-art signal generator

A lot of time and effort is saved by automated measurements, but problems may still be present. During the design phase of a circuit, a manual prototype measurement may give a better picture of the functions and in particular the errors of a new circuit. Automated quality-assurance measurements during the production must be designed carefully for each circuit, and calibrations must be repeated regularly enough. This requires microwave engineers to be very familiar with microwave measurement techniques.

This handout presents the most important and most frequently used microwave measurement techniques and equipment. In each measurement scenario we discuss the physical

background as well as the measurement methods. We can find manual and automated measurement methods, and the manual method often gives a better picture of the physical background and phenomena. We handle the matter in the following order:

1. measurement uncertainty in microwave measurements
2. the first measurement device discussed is the signal generators, because without a coherent signal practically no measurement can be made
3. passive components needed in measurements, and diode detectors
4. measuring equipment and methods for basic signal parameters: power and frequency
5. measuring equipment and methods for circuit parameters: different impedance measurement methods, including the vector network analyzer
6. spectrum analyzers offer many measurement possibilities, vector-signal analyzers are used for examining complex modulated signals
7. measurements of non-idealities of circuits: signal sources for noise and noise measurement methods, measurement methods for the non-linearity of amplifiers and spectral purity of oscillators
8. time-domain measurements
9. measurement methods and measurement sites for antenna characterization
10. a brief look at EMC testing
11. finally, measurement methods for the dielectric constant of materials

Some additional reference books for general microwave measurements are listed in references [1-4], which are suitable as a reference for most of the lecture material and handouts.

Planning a measurement in advance saves time and effort. The following points need to be thought of before starting a measurement:

- What quantities or parameters are to be measured?
- How does the device or circuit work, what measurement outcome is expected?
- What alternative measurement devices and methods are available?
- What does the measurement result mean and how to calculate the desired quantity?
- What is the uncertainty in the measurement result and is the accuracy good enough?

Microwave devices, and in particular measurement devices are expensive. All measurement equipment and other components needed in the measurement must be handled very carefully. Naturally, careful handling of equipment and components is also important in terms of measurement accuracy. The following rules should be followed:

- all connectors, adapters and cables must be treated carefully and in a proper manner.
- static discharges are to be avoided (especially during dry and cold weather conditions); a good way to be "grounded" is to touch for example a water pipe or radiator before touching the equipment, if available use a dedicated grounding wrist strap.
- before switching on a signal generator, check that the (maximum) power output is lower than the power rating of the attached load.
- before switching on the DC power supply of any active component, check the required voltage and that the output current limit of the power supply is set low enough.

[1] Bryant G. H.: *Principles of Microwave Measurements*. London 1988, Peter Peregrinus.
[2] Bailey A.E. (ed.): *Microwave Measurement incl. Suppl. Vol.* London 1988, P. Peregrinus.
[3] Laverghetta T. S.: *Handbook of Microwave Testing*. Dedham, Mass. 1981, Artech House.
[4] Laverghetta T. S.: *Modern Microwave Measurements and Techniques*. Norwood, Massachusetts 1988, Artech House.

Measurement Uncertainty

Measurement results are never exact. Therefore any presented measurement result is incomplete without a quantitative uncertainty [5]. Without this uncertainty any differences between results from different measurements cannot be evaluated: Are the measured results really different or is the difference a result of random variation? We now take a look at measurement uncertainty of microwave measurements.

Uncertainty and error

Uncertainty gives the limits within which the true value of the measured quantity can be expected. Uncertainty is usually expressed with a ±-value, whose unit is e.g. the same as for the quantity measured, a percentage, or $1{:}10^x$. Also the probability p needs to be given, with which the true value is expected to fall within these limits, i.e. the statistical confidence interval. Sometimes one might choose the largest possible uncertainty of p = 100%. But usually it is better to use a smaller probability, e.g. p = 95%, in order to avoid unrealistically large uncertainty limits. The determination of the uncertainty requires understanding of all the measurement steps and possible corrections. Experience is of great help when determining the measurement uncertainty.

Error is the difference between measured and true value of the measured quantity. The error can never be precisely determined. An approximation for the error can be obtained when comparing measurement results to international or national measurement standards or to measurement results obtained with normal or calibrated measuring instruments.

Measurement errors are divided into random and systematic errors. Other possible errors such as the misinterpretation of a scale are not considered here. Random errors do not bias the results in a particular direction. It makes the result uncertain due to its statistical probability distributions. Systematic error always influences results into the same direction, and it can often be partially corrected. For example, when we once compare our results with a measurement standard, the obtained difference between the results can be used to correct all measurement results in future. The division into random and systematic errors is not always obvious, instead it may depend on the circumstances. For example, the calibration of a measurement device will vary because of random error. The error of the derived correction factor is then systematic, since it affects all the following measurements in the same way.

Random uncertainty

Even when repeating a measurement many times under the same conditions, the results vary due to several random factors. Typical causes of variations in electronic systems are noise and supply voltage fluctuations.

The probability distribution of random errors can be measured (given sufficient resolution of the measurement equipment) by repeating the measurement several times. Usually the probability distribution follows Gaussian distribution or normal distribution (Fig. 3). The normal distribution

$$y(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-x_0)^2/(2\sigma^2)},$$

where σ is the standard deviation and x_0 is the expected value. y(x) is the probability density function and its integral from $-\infty$ to $+\infty$ is 1.

The probabilities with which a measurement result is within the limits $x_0 \pm K\sigma$, are shown in Table 1. For example, the K value that corresponds to a 95% probability is K = 1.96. If the measurement error is made up of several random factors that are of the same magnitude, the distribution is still almost normal even if the separate contributions are not normal distributed.

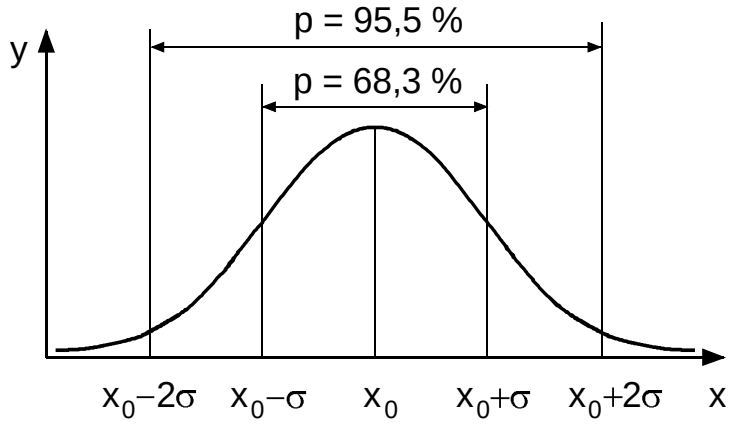

Figure 3. Normal (Gaussian) distribution.

Table 1. Probabilities p, with which the result of a measurement lies within the limits $x_0 \pm K\sigma$, presuming that the measurement error has normal distribution:

K	p / %
0,675	50,0
1	68,3
1,96	95,0
2	95,5
2,58	99,0
3	99,7

Let's assume that a measurement is repeated n times and the results obtained are x_1, x_2, ... and x_n. The arithmetic mean (or average) is $\bar{x} = (x_1 + x_2 + ... + x_n)/n$, which varies from one series of measurement to another. Now the random component of the uncertainty is difference between $\bar{x}$ and x_0. The standard deviation of this particular measurement series is

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} .$$

The estimate for the standard deviation of the probability function is

$$\sigma(est) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = S\sqrt{\frac{n}{n-1}} .$$

This estimate differs significantly from S only if n is small. When n is large, $\sigma(est)$ approaches the standard deviation σ.

The standard deviation describes how much individual measurements within the series of measurements can differ from the mean. If the measurement series is repeated several times, the variation of the mean is smaller than the variation of individual measurements. A

single series of measurements can be used to calculate the uncertainty of the mean value, which is expressed with help of the *K*-factor that corresponds to a certain probability

$$U_r = \frac{K\sigma}{\sqrt{n}}.$$

Since in practice $\sigma(est)$ is based on a limited number of measurements, this will result in additional uncertainty. This is taken into account by replacing the *K*-factor with an *n*-dependent *t*-factor [6], which is presented in Table 2. When *n* approaches infinity, *t* approaches *K*. The random uncertainty given for the desired probability is thus

$$U_r = \frac{t\sigma(est)}{\sqrt{n}}.$$

This shows that the random uncertainty decreases when the number of measurements increases. The attained improvement, however, gradually decreases, since one needs four times the number of measurements to reduce the uncertainty by half.

Table 2. Coefficient *t* as a function of the number of measurements *n*, given for the probabilities *p* = 0.5, 0.95 and 0.99:

n	*p* = 0,5	*p* = 0,95	*p* = 0,99
2	1,000	12,7	–
3	0,817	4,30	9,92
4	0,765	3,18	5,84
5	0,741	2,78	4,60
10	0,703	2,26	3,25
15	0,692	2,14	2,98
20	0,688	2,09	2,86
∞	0,675	1,96	2,58

Systematic uncertainty

In order to properly determine the uncertainty we need to identify all sources of systematic errors. Typical sources of systematic uncertainty are the measurement device calibration, and errors caused when connecting measurement devices and circuits under tests. Calibration uncertainty is often a significant source of error. Calibration is valid at a given moment and under certain circumstances. Slow drift in equipment characteristics and differences in the environmental conditions during calibration and measurement (temperature, humidity, etc.) may increase the uncertainty. On calibration certificates of measurement equipment correction factors are often given for specified e.g. frequencies or power levels. If the measurement frequency or power level happens to fall between these given values, the necessary interpolation increases uncertainty. Also connecting a measuring device to other devices can affect the result. If the set-up differs between calibration and measurement, more uncertainty is created, therefore especially in microwave measurements the impedance mismatch between circuits or equipment is a common source of error.
The systematic uncertainty of the probability distribution is not usually measured, but it is based on theory, earlier experimental work or simply estimates. If we know only the limits of the systematic uncertainty, it is safe to assume rectangular distribution between $x_0 \pm a$ (see Fig. 4 a). The standard deviation of the rectangular distribution is

$$\sigma_{sq} = \frac{a}{\sqrt{3}}.$$

In microwave measurements the power transfer between circuits that are not perfectly matched, such as from a generator to a load (see example below), is an example of a theoretically predictable probability distribution. We often know the absolute value of the reflection coefficients of generator and load (in presence of the connecting cable), but the actual power coupled to the load depends also on the phase of the reflection coefficients and the distance between the reflections, both of which are usually not known. The probability distribution is now U-shaped between $x_0 \pm a$ as shown in Fig. 4(b), and the standard deviation of this distribution is

$$\sigma_u = \frac{a}{\sqrt{2}}.$$

The U-shaped distribution can be understood by looking at the arrows that describe the propagating waves. The first arrow indicates the direct wave approaching the load and the second arrow indicates the wave(s) reflected by the mismatch. The length of the resulting arrow is proportional to the voltage across the load. If the phase difference between the first two arrows is uniformly distributed between 0-360°, the length of the resulting arrow is more likely to be quite large or quite small, rather than near the average.

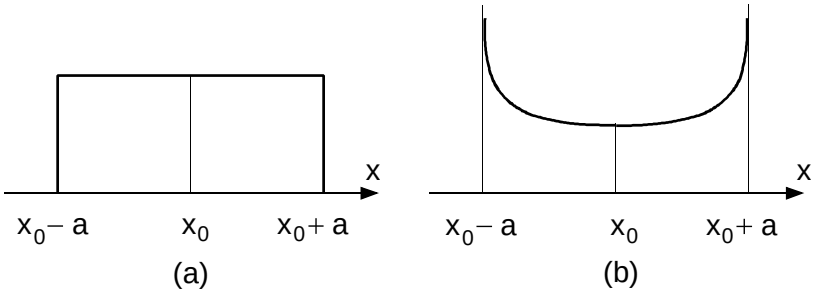

Figure 4. Probability distributions: (a) rectangular distribution, (b) U-distribution of power transfer between generator and load.

If the systematic uncertainty consists of non-correlated factors, the combined standard deviation is

$$\sigma_s = \sqrt{\sum_i (\sigma_i^{\,2})}$$

and the combined uncertainty is

$$U_s = K\sigma_s.$$

The probability distributions of different factors can have different shapes. If various factors are of the same magnitude, the total distribution is approximately normal-shaped. Figure 5 is an example of a combination of three rectangular probability distributions, where $x_0 = 0$ and $a_1 = 1$, $a_2 = 1.5$ and $a_3 = 3$. For such a combination of rectangular probability distributions and if $K > 1.8$, the probability that the measured value in within the limits $\pm K\sigma$ is always greater than the probability given by the normal distribution.

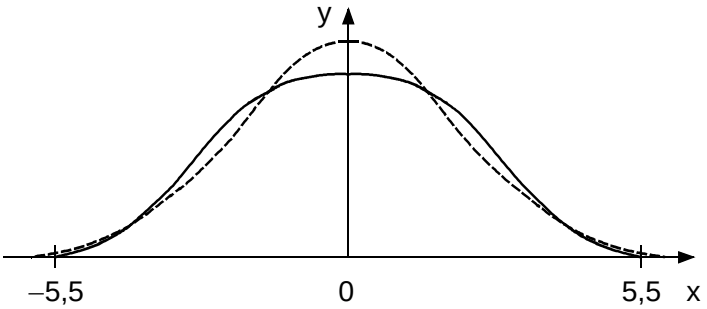

Figure 5. Combination of three rectangular probability distributions (solid line) versus a normal distribution (dashed line) with the same standard deviation.

When selecting a K value and calculating the corresponding U_s, one can avoid too pessimistic uncertainties by verifying that U_s does not exceed $\Sigma_i a_i$, which can happen when there is a dominant factor in the sum. This dominant factor a_d should be separated from the others, i.e. $U_s' = a_d + U_s$.

Total Uncertainty

When random (U_r) and systematic uncertainties (U_s) have been determined (for the same probability p), the total uncertainty

$$U = \sqrt{U_r^2 + U_s^2} \ .$$

In case of a dominant factor in the systematic uncertainty, the total uncertainty

$$U = a_d + \sqrt{U_r^2 + U_s^2} \ .$$

A measurement result is now expressed in the form $\bar{x} \pm U$ along with the information with which probability p the limits hold. Typically p = 95% or p = 99%. Usually one significant digit (or decimal) is sufficient for the uncertainty (more than two decimals should never be given).

Example: An uncalibrated power meter is compared to a calibrated one. The power of the generator is set so that the calibrated power meter shows exactly 10 mW. Then the uncalibrated power meter is connected to the generator. (Fig. 6).

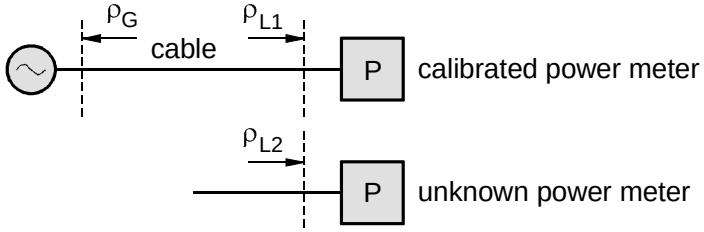

Figure 6. Comparison measurement with an uncalibrated and a calibrated power meter.

The systematic error sources (rectangular distribution) and their limits a are:

1) calibration uncertainty: 1 %,
2) drift between calibrations: 0,5 %,
3) interpolation error (no calibration point at measurement frequency) 0,2 %.

   The voltage standing wave ratios (*VSWR*) and reflection coefficients ($|\rho|$) of the devices are:

   – generator $VSWR_G$ = 1,10, $|\rho_G|$ = 0,048,

– calibrated power meter $VSWR_{L1}$ = 1,15, $|\rho_{L1}|$ = 0,070,
– uncalibrated power meter $VSWR_{L2}$ = 1,35, $|\rho_{L2}|$ = 0,149.

When reflection coefficients are small, the maximum error due to mismatch is $2|\rho_G||\rho_L|100$ %.

Then the error limits a of the systematic error due to mismatch (U-shaped distribution) are:

4) generator and calibrated power meter mismatch: 0,66 %,
5) generator and uncalibrated power meter mismatch 1,42 %.

For these five systematic error sources combined the standard deviation is σ_s = 1,29 %. We choose p = 95 % probability, i.e. K = 1,96, so that the systematic uncertainty yields U_s = 2,52 %.

In order to determine the random uncertainty we repeat measuring the power four times with the uncalibrated power meter, and we get the readings 10,04, 10,07, 10,03 and 10,06 mW. The average is 10,05 mW and the standard deviation is 0,018 mW, or 0,18 %. Since n = 4 and p = 95 %, according to Table 2 we get t = 3,18. So finally we obtain the random uncertainty U_r = 0,29 %.

The total uncertainty is now 2,54 %. The power 10 mW corresponds to the (uncalibrated) power meter reading of 10,05 mW ± 0,26 mW with 95 % probability.

Uncertainty of derived parameters

Often the investigated parameter is not measured directly, but derived from several measured parameters. Let's assume parameter G is a function of several, pairwise independent parameters X_i:

$$G = f(X_1, X_2, \ldots X_n).$$

Independent measurement parameters means, that no systematic error source affects two or more parameters simultaneously. When for each measurement parameter the average is $\overline{X}_i$ with the estimated uncertainties $U_{\overline{X}i}$ (for the same probability),

$$G = f(\overline{X}_1, \overline{X}_2, \ldots \overline{X}_n)$$

and the RSS-uncertainty (RSS = root-sum-of-squares) of G is

$$U = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f}{\partial X_i} U_{\overline{X}i} \right)^2} .$$

The partial derivatives are calculated by using $\overline{X}_i$ for X_i. With help of this equation we can estimate how much different error sources affect the total uncertainty.

The worst-case uncertainty for G (i.e. the total uncertainty that corresponds to p = 100 %) we get by using for each parameter X_i the maximum uncertainty $U_{\overline{X}i,max}$ and assuming that the errors do affect G in the same "direction":

$$U_{max} = \sum_{i=1}^{n} \left| \frac{\partial f}{\partial X_i} \right| U_{\overline{X}i,max} .$$

If we have several parameters X_i, it is, however, extremely unlikely that the error really is of the magnitude of the maximum uncertainty.

Example: We measure the noise temperature T_e of a device with the *Y*-factor method (see p. 91ff.). We get T_e from the equation

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

where T_H and T_C are the noise temperatures of the loads that are connected to the input of the device, Y is the power ratio P_H/P_C of the corresponding output powers. The RSS-uncertainty of T_e

$$U = \sqrt{\left[\frac{T_H - T_C}{(Y-1)^2}\right]^2 U_Y^2 + \left(\frac{1}{Y-1}\right)^2 U_{TH}^2 + \left(\frac{Y}{Y-1}\right)^2 U_{TC}^2}\ .$$

In the (unlikely) worst case we get a maximum uncertainty of

$$U_{max} = \frac{T_H - T_C}{(Y-1)^2} U_{Y,max} + \frac{1}{Y-1} U_{TH,max} + \frac{Y}{Y-1} U_{TC,max}.$$

If the uncertainties of the noise temperatures and the *Y*-factor are $T_H = 295 \pm 0{,}2$ K, $T_C = 78 \pm 1$ K and $Y = 2{,}00 \pm 0{,}02$, we obtain for $T_e = 139 \pm 4{,}8$ K (RSS). We can find that the uncertainty in the *Y*-factor is clearly the dominant one. If the three given uncertainties are the maximum uncertainties, we would get a worst-case uncertainty for T_e of ±6,5 K.

[5] Hinton L. J. T.: Uncertainty and confidence in measurements. Ch. 21 in *Microwave Measurement incl. Suppl. Vol.* (ed. Bailey A. E.). London 1988, Peter Peregrinus, s. 418–443.

[6] Harris I. A., Warner F. L.: New expressions and some more precise values for Student's *t*-factor. *Proceedings of IEE*, 125(1978)9, s. 902–904.

Signal Generators

In many measurements we rely on the exact frequency of the test signal, which is provided by a general-purpose signal generator. In some measurements, a simple oscillator is sufficient as a signal source. Or in noise measurements, the thermal noise generated by a matched termination acts as a test signal.

Actual signal generators characteristics vary greatly. With a simple signal generator the frequency can be tuned - manually or electronically - over a fairly small range. In somewhat more advanced signal generators we can automatically sweep the frequency and the power, and signal modulation is possible. Synthesized signal generators can produce highly accurate and stable signals over a wide frequency band. Synthesized vector signal generators are capable of additionally creating signals according to the various communications standards with their complex modulated signals.

Oscillators

The oscillator [7] is the core component of any signal generator. It is either a semiconductor or a tube oscillator. Oscillators can be frequency-wise freely floating or phase locked to an accurate reference signal. If the measurement frequency is difficult to generate directly, the oscillator can also be combined with a frequency multiplier.

Tube oscillators are used in millimeter and submillimeter frequency ranges, since at so short wavelengths it is difficult to generate high enough power with semiconductors. Tube oscillators are however difficult to use: they are large, require high voltages, and have a short lifetime. Klystrons are used for frequency up to 150 GHz. They can generate a very pure spectrum. Backward wave oscillators (BWO) are well suited for frequency sweeping. Their frequency can be tuned linearly over about one octave. BWOs can exceed 1000 GHz.

Semiconductor oscillators, i.e. transistor and diode oscillators, have largely replaced tube oscillators at RF and microwave range and also at millimeter waves. They are conveniently small and require only small supply voltages. In transistor oscillators bipolar transistors (BJT, HBT) and field-effect transistors (MESFET, HEMT) are used with a feedback circuit. The BJT operates up to about 20 GHz and the HBT up to 100 GHz. MESFETs operate as oscillators up to 100 GHz and HEMTs up to 200 GHz. With bipolar transistors the oscillator phase noise is typically 15 dB lower than with field-effect transistors.

A Gunn or IMPATT diode can act as a negative resistance and thus also as an oscillator. Gunn-oscillator operates at the fundamental mode up to about 150 GHz and at the harmonic mode up to 300 GHz. IMPATT oscillators are used up to 300 GHz. IMPATT oscillators are suitable for frequency sweeping, but its spectrum is quite noisy.

In semiconductor oscillators frequency *tuning* is often achieved with a YIG resonator (YIG = yttrium iron garnet). A YIG is a bulk ferrite whose resonance frequency depends linearly on the applied static magnetic field strength. The magnetic field is created by applying direct current to an electromagnetic coil. Frequency tuning ranges of 2 – 3 octaves are possible, and sweeping is very linear (deviations < 0.1%). YIG-resonators are usually spherical with a diameter of 0.2 – 2 mm. RF power is coupled to them with a coupling loop around the sphere. The unloaded quality factor is about 10 000. YIG-resonators are used up to 50 GHz. The tuning range can alternatively be moderately extended by adding a varactor diode to the circuit, whose capacitance depends on the applied bias voltage. Varactor tuning is used up to about 100 GHz.

Signal generator characteristics

An ideal signal generator would produce a clean sinusoidal or modulated signal of a frequency and power that are accurately tunable. In practice, the signal spectrum is not clean but contains noise and spurious emissions, the frequency is neither accurate nor stable, and power output is not accurate nor is the frequency response flat.

Listed below are the main characteristics of a signal generator, which will determine the usefulness of the device and also the price:

- The frequency range of the measurement equipment has to cover the frequency range of the device under test. In fact, often measurements should be extended beyond the actual operational bandwidth of the device. A simple generator can cover an octave or two. At very high frequencies the signal generator frequency range may be limited partly by the frequency band of the used waveguide. The frequency range of synthesized signal generators can extend from below 1 MHz to millimeter wave range.
- The frequency accuracy is given in frequency deviation Δf or in percent of the output frequency. The reading error rate for a mechanical scale may be as much as 1%. On a digital display the frequency may be given with a number of significant digits, as long as frequency counter is used. The frequency accuracy of an unsynthesized signal generator depends mainly on the active element and the temperature dependence of the coupled resonator, whereas for the synthesized signal generator all depends on the reference oscillator accuracy. If a measurement requires the use of two separate signal generators and the frequency difference of the two signals is to be measured accurately in the circuit under test, it is worthwhile using the reference oscillator of one of the two signal generators as the common reference.
- Frequency resolution refers to how accurately the frequency can be *read* on the screen. This is a property of the display; it does not describe the frequency accuracy. For a synthesized generator the resolution may be better than 1 Hz.
- Stability is one of the most important characteristics of a signal generator. It describes how well the frequency remains at the set value. Slow (taking more than one second) fluctuations or drift is given as a frequency change within a certain period or parts-per-million (ppm) as a function of temperature change and supply-voltage change. Frequency drift is largest during the warming phase after switching on the equipment. Fast (within less than a second) changes in the frequency spectrum are called residual frequency modulation, given as maximum frequency deviation, typically e.g. 5 – 10 kHz.
- Single-side-band phase noise describes the spectrum spread caused by phase deviations (can also be interpreted as frequency deviation). The phase noise is expressed in noise power relative to the carrier wave power, at several frequency offsets, normalized to 1 Hz bandwidth (dBc / Hz). The carrier in this context means the ideal signal. For a synthesized microwave signal generator, the phase noise can be less than 60 dBc / Hz, 10 Hz away from the carrier, and 100 dBc / Hz, 10 kHz from the carrier wave. Fig. 7 illustrates how phase noise interferes with the measurement of the receiver selectivity. The IF-filter frequency response of the tested receiver is shown, as well as a strong interference signal tuned to the neighboring channel. At high power levels, due to the phase noise of the generator the interference signal, power can also leak into the passband of the filter which directly affects measurement accuracy.

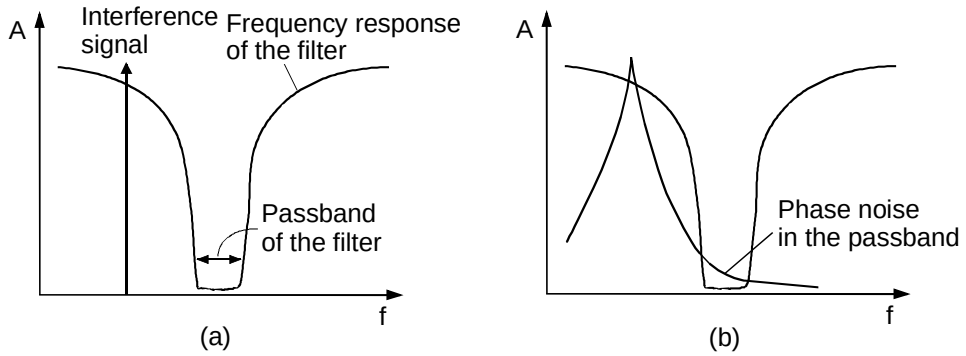

Figure 7. Effect of phase noise on channel-selectivity measurements: (a) ideal interference signal, (b) noisy interference signal.

- Spurious emissions are signals generated by the signal generator at the wrong frequency. They are divided into harmonic and non-harmonic. The harmonic spurious emissions are not harmful as they can be removed easily with a suitable low-pass filter. A good generator has harmonic spurious emissions levels of less than −50 dBc and non-harmonic spurious emission levels of −90 dBc. As seen for the phase noise, also spurious emissions can affect the measurement of channel selectivity.
- The range within which the output power level can be adjusted restricts the dynamic (power) range of the measurements. Often, the output power is not calibrated with sufficient accuracy and has to be separately measured with a power meter. The output power of high-end signal generators is calibrated from −120 dBm to +10 dBm and also power sweeping is possible. (dBm stands for power in decibels normalized to one milliwatt i.e. P (dBm) = 10log (P / 1 mW). The flatness of the frequency response is typically ± 1 dB.
- More advanced signal generators also offer digital and analogue modulation of the signal. Modulation parameters are defined by the user or are preset according to different communication system standards. Modulation can be controlled either internally or externally.

Simple signal generators

The block diagram of a simple signal generator is shown in Fig. 8. A freely oscillating (signal is not phase locked) oscillator is tuned electrically or mechanically. Frequency modulation (FM) is achieved through varying the control voltage. The buffer amplifier following the oscillator prevents the oscillator frequency from changing with load impedance. Often the output power of an oscillator power strongly varies with frequency. To allow measurements with constant power, an automatic level control (ALC) is added. In Fig. 8 the ALC circuit is based on a pin-diode attenuator. With the directional coupler a sample of the oscillator output is fed to the detector diode. If the detector voltage deviates from the reference value (ALC ref.) the attenuation of the pin-diode is adjust accordingly. Amplitude modulation (AM) is implemented within this ALC circuit. With the adjustable attenuator the output power of the generator is controlled (continuously or in discrete steps).

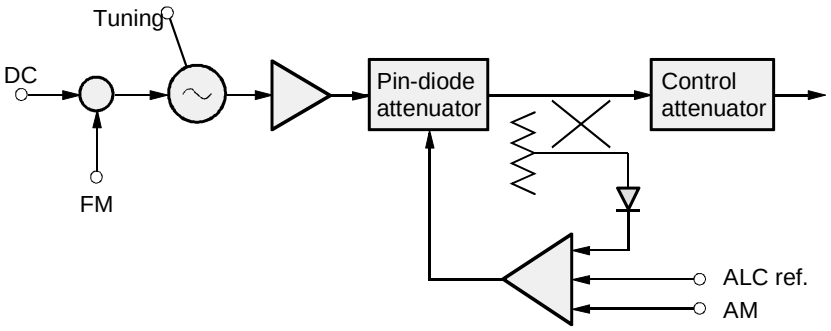

Figure 8. Simple signal generator

Sweep generators

Sweep generators are used when a continuous frequency band has to be covered in a measurement. We do not consider synthesized sweep generators in this section.

Sweep generators usually consist of a base unit and an interchangeable high frequency unit, as shown in Fig. 9. The base unit provides the control voltage to the high-frequency unit: a basic linear voltage ramp, a frequency modulation or tuning (FM) signal, a pulse- and amplitude modulation (AM) signal, and a reset signal between sweeps. You can choose between several sweep types, such as full bandwidth sweep, sweep between start and end frequencies, and a symmetrical Δf-sweep around a center frequency. The sweep duration can be selected, e.g. between 0.01 – 100 s. Also manual sweeping is possible.

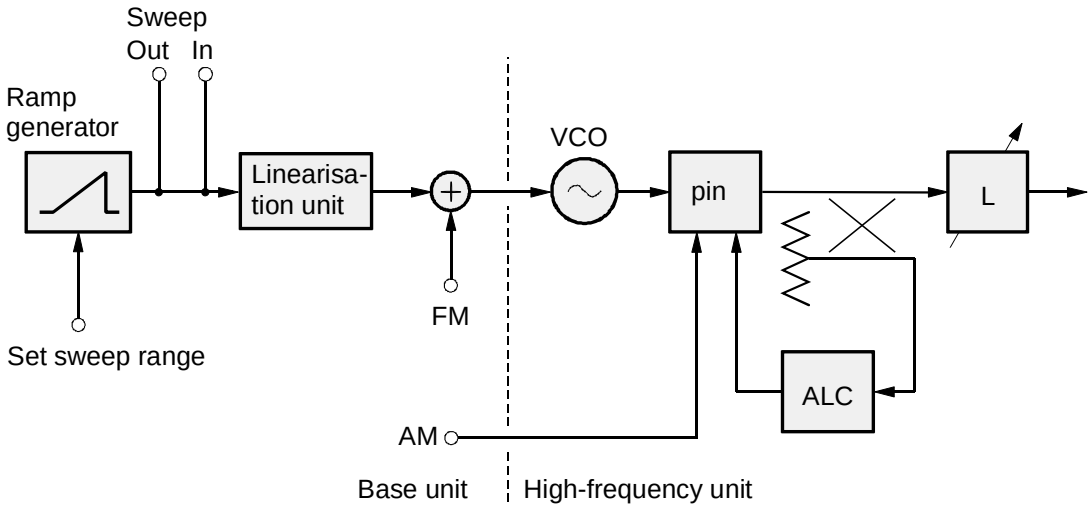

Figure 9. Block diagram of a sweep generator

The general base unit can be connected to a number of different high-frequency units for different frequency ranges. The high-frequency unit consists of a voltage-controlled oscillator (VCO), a level control circuit and the output attenuator. The frequency tuning of the oscillator is usually based on varactor diodes or YIG-resonators. Since the capacitance of a varactor diode exhibits a very non-linear dependence on the bias voltage, a linearization circuit is added in Fig. 9. This ensures a wide and linear frequency sweep.

Synthesized signal generators

The signal generated by a synthesized signal generator, i.e. a frequency synthesizer is based on a very stable, low-frequency reference oscillator. The synthesized signal is similarly stable, low-noise, and accurate in frequency.
The reference oscillator is often an oven-stabilized 10 MHz quartz oscillator, the drift is typically less than 5×10^{-10}/day. Oven stabilization means that the crystal is kept at a constant temperature independent of changes in the ambient temperature. Also a stable external oscillator can act as a reference oscillator. Frequency synthesis can be realized in two ways, as direct and as indirect synthesis [8-10].

Direct frequency synthesis

In direct frequency synthesis the desired output signal is obtained by frequency multiplication/division and mixing of the same reference signal (coherent frequency synthesis) or several reference signals (incoherent frequency synthesis). A step-recovery diode is often used as a multiplier. A step-recovery diode generates comb spectrum, i.e. a large number of harmonic frequencies. Fig. 10 (a) illustrates, how by multiplying (N_1), dividing (N_2), and mixing we obtain frequencies $f_i = N_1 f_{ref} + i f_{ref} / N_2$ (i = 0, 1, 2, 3, 4), afterwards we filter out the sum-frequency component from the output signal of the mixer. By combining several similar units with switches and mixers according to Fig. 10 (b) we obtain signals at frequencies $f_o = f_i + f_j + f_k + ...$
With direct frequency synthesis very fast frequency changes can be realized (less than 1 ps), mainly limited by the switching times of the switches. The disadvantage of direct frequency synthesis is the increasing amount of components and unwanted frequencies for an increasing number of available output frequencies.
In direct digital synthesizers (DDS) the output wave form is generated from a sequence of discrete amplitude samples. With digital synthesis, the frequency can be changed quickly and one can achieve a good resolution (small frequency increments). In addition, many modulated signals can be generated accurately. The disadvantages of direct digital synthesis are the generation of undesired frequency components due to quantization errors and the non-linearity of the DA-converter, and the limitation of the maximum frequency to half the clock frequency. We can obtain higher frequencies by mixing, or by connecting the DDS circuit into a phase-locked loop.

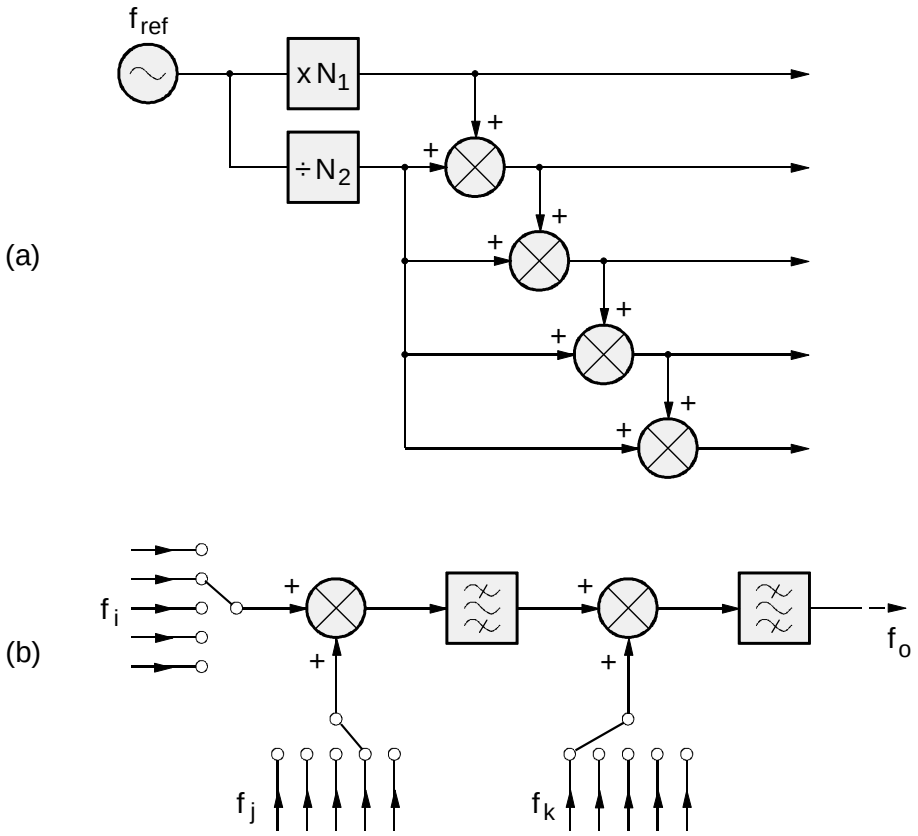

Figure 10. Direct frequency synthesis: (a) creation of an accurate frequency signal by multiplying, dividing and mixing, (b) combination of switch-selected frequencies with mixers and filters.

Indirect frequency synthesis

Indirect frequency synthesis is based on a feedback phase-locked loop (PLL) [11, 12], with which the signal of a voltage controlled oscillator (VCO) is locked to a reference oscillator. In Fig. 11 it is illustrated how the output signal f_o of a VCO is mixed down to $f_o - f_{LO}$ and divided by N_2. In the phase detector this signal is compared to the signal obtained by dividing the reference frequency f_{ref} by N_1. The phase detector can be a double-balanced mixer or a digital phase comparator. The output voltage of the phase detector is proportional to the phase difference of the two signals. After the loop filter (a low-pass filter) this voltage adjusts the VCO's frequency until the phase difference becomes constant. The loop filter determines the main PLL characteristics such as the lock-in range, lock-in speed, and partly also the phase noise. When the phase difference eventually is constant, i.e. the PLL is locked, the output frequency

$$f_o = \frac{N_2}{N_1} f_{ref} + f_{LO}\,.$$

The frequency is varied by changing factor N_2 and the local oscillator frequency. N_1 is usually constant. The local oscillator frequency f_{LO} is generated by multiplying the VCO frequency of a second phase-locked loop, whose mixer in turn gets its local oscillator frequency from a third phase-locked VCO. With a sufficient number of loops any desired number of different frequency signals with any frequency resolution can be generated.

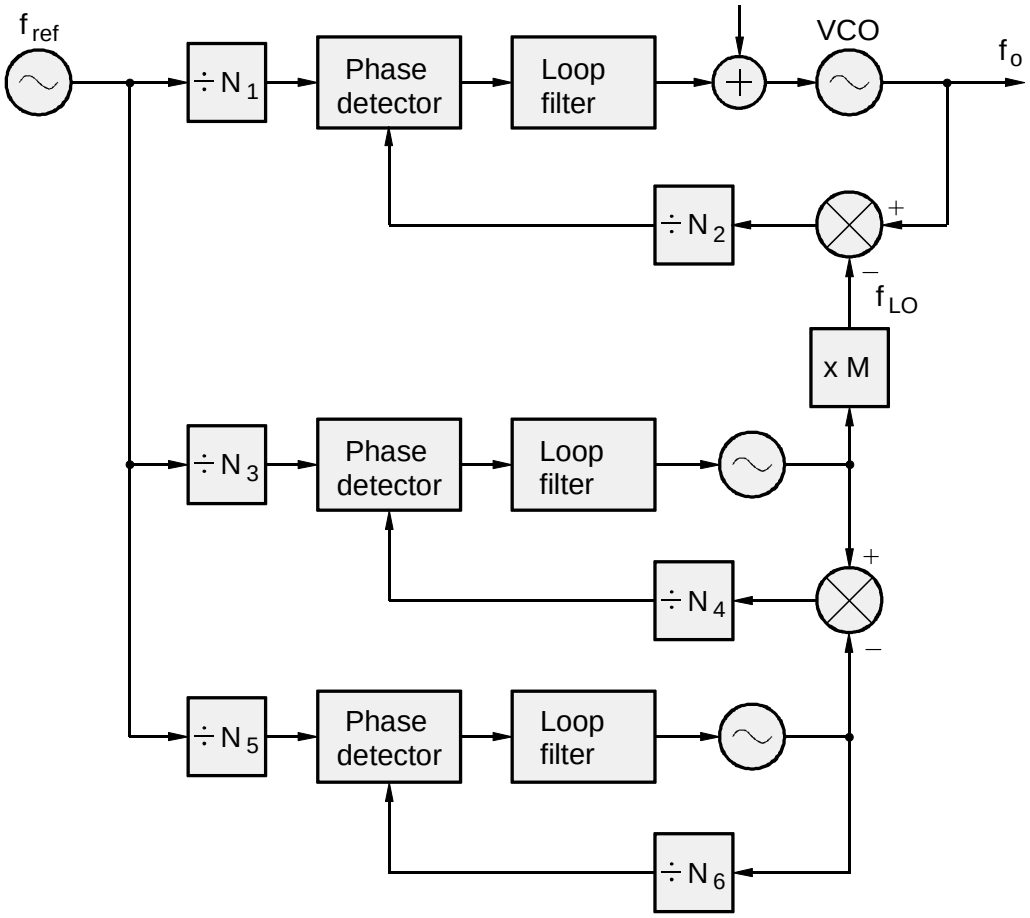

Figure 11. Signal generator based on indirect frequency synthesis.

All the *N* factors in Fig. 11 are integer numbers. However, even with a single phase-locker loop one can create small frequencies steps, namely by quickly alternating the division factor between *N* and *N* + 1, which results in a fractional average division factor N_{ave}.

The frequency sweeping in a synthesized frequency generator is now realized so that the generator goes through its possible densely distributed frequency points. The lock-in time for each frequency point is about 50 ms, which limits the sweeping speed. Alternatively, only the starting frequency is indirectly synthesized, and the sweeping is done like in a normal (non-synthesized) sweep generator.

The typical phase noise spectrum of a frequency synthesizer based on indirect synthesis is shown in Fig. 12. From the carrier wave frequency to the edge of the loop bandwidth the phase noise follows the reference oscillator phase-noise spectrum, while outside this frequency band the phase noise is determined by the VCO. Far from the carrier wave the reference oscillator phase noise is stronger than that of the VCO. This is due to the fact that the frequency division in the loop feedback branch corresponds to a frequency multiplication, and frequency multiplication increases the level of phase noise by 20log*N* dB. The phase noise spectrum is usually optimized by choosing the loop bandwidth so that at the according distance from the carrier wave frequency the oscillator noise values are identical.

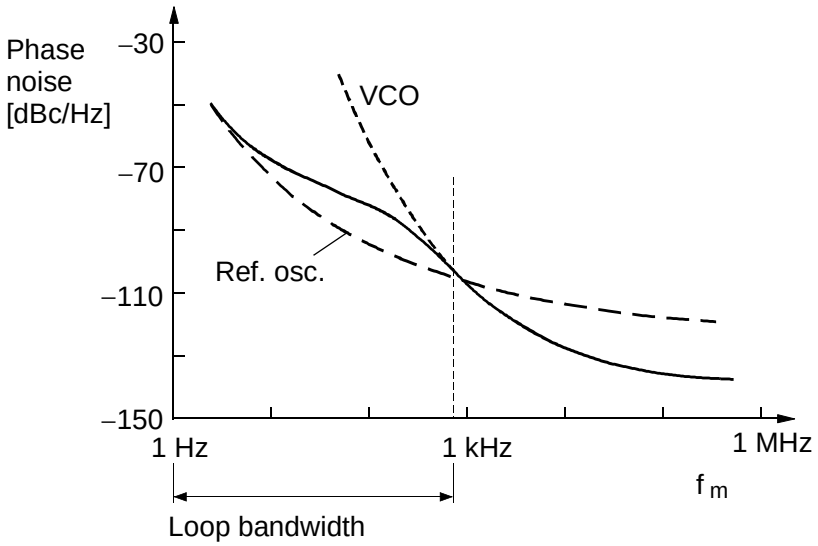

Figure 12. Phase noise spectrum of a frequency synthesizer based on indirect synthesis (solid line); f_m is the offset from the carrier wave frequency.

Vector signal generators

A vector signal generator produces both analogue modulated signals and complex, digitally modulated signals. Typically selectable digital modulation methods include, e.g. phase modulation schemes BPSK, QPSK, OQPSK, π/4-DQPSK and 8PSK, frequency modulation schemes FSK, 4FSK, 8FSK, MSK and GMSK, as well as QAM modulation schemes up to 256QAM.

Modulated signal parameters are user definable or preset according to the common communications systems standards. Among these parameters are digital baseband bit rate and filtering function, FM or FSK frequency deviation, AM or FM modulation frequency, and pulse shape and length of pulsed signals. Multi-purpose generators cover the standards of mobile communications networks and related standards such as GSM, EDGE, W-CDMA, CDMA2000, DECT, IEEE 802.11 and Bluetooth. A generator should be upgradeable to new standards (or to possible changes in existing standards).

For enhancing the tests of components and receivers a generator may provide a number of useful features. A generator might be able to produce two signals at different frequencies to measure the intermodulation of amplifiers. In tests of base stations and mobile devices we need test signals showing real-world behavior such as noise and fading. Being able to connect a bit-error analyzer to the generator facilitates testing the sensitivity and selectivity of receivers. Testing of code-sharing, i.e. CDMA systems (spread spectrum techniques) requires very complex signals. The generator needs to produce a number of carrier waves, where each carrier contains a number of coded channels.

The block diagram of a vector signal generator is shown in Fig. 13. The carrier wave is generated by the frequency synthesizer and the IQ modulator modulates the signal. The generator itself can produce a baseband signal (i.e. a bit string), or this can be externally fed to the generator's memory. The carrier wave is encoded in I- and Q-signals, which are filtered to obtain the desired signal spectrum. Also I- and Q-signals can be externally fed to the generator. For testing receivers the signal can be impaired, e.g. by adding IQ-imbalance, oscillator phase noise, noise or other interferences.

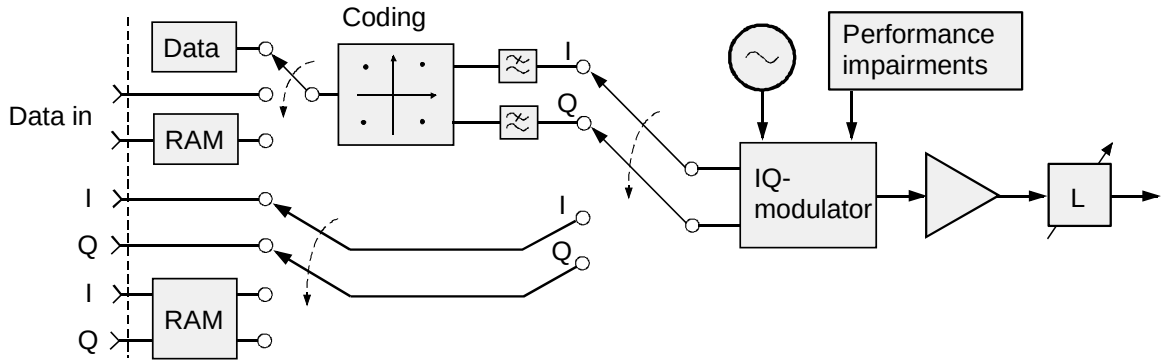

Figure 13. Vector signal generator.

[7] Bahl I., Bhartia P.: *Microwave Solid State Circuit Design*, Second Edition. Hoboken, New Jersey 2003, John Wiley & Sons.
[8] Manassewitsch V.: *Frequency Synthesizers: Theory and Design*, Third Edition. New York 1987, John Wiley & Sons.
[9] Crawford J. A.: *Frequency Synthesizer Design Handbook*. Norwood, Massachusetts 1994, Artech House.
[10] Rohde U. L.: *Microwave and Wireless Synthesizers: Theory and Design*. New York 1997, John Wiley & Sons.
[11] Gardner F. M.: *Phaselock Techniques*, 2nd Edition. New York 1979, John Wiley & Sons.
[12] Egan W. F.: *Phase-Lock Basics*. New York 1998, John Wiley & Sons.

Passive components

In microwave measurements we need a wide range of components such as cables, connectors, attenuators, etc., whose characteristics and correct use are important for successful measurements. These passive components are discussed extensively in microwave technology textbooks, e.g. [13-15].

Waveguides

Microwave devices and components are usually connected with coaxial cables. These can be used up to about 60 GHz. Rectangular waveguides are used in measurements starting from about 10 GHz to over 300 GHz (see Fig. 14.). When testing semiconductor components often a microstrip measurement jig, i.e. a special fixture is used.

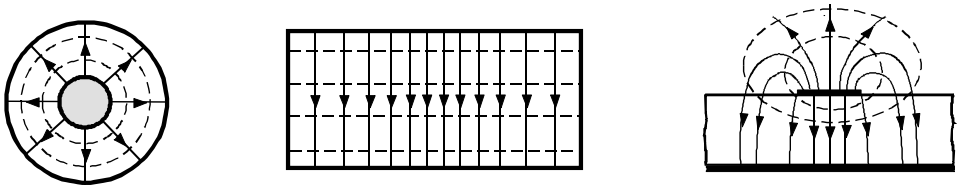

Figure 14. Cross sections of coaxial line, rectangular waveguide, and microstrip line, each with field pattern of lowest order wavemode (E-field depicted as solid line, H-field as dashed line).

Table 3. Characteristics of transmission lines used in microwave measurements.

Characteristic	Coaxial line	Rectangular waveguide	Microstrip line
Basic wavemode	TEM	TE_{10}	Quasi-TEM
Bandwidth	Wide	Average	Wide
Dispersion	None	Average	Small
Losses	Average	Small	Large
Power rating	Average	High	Small
Size	Average	Large	Small
Processibility	Average	Average	Good
Integratability	Poor	Poor	Good

Coaxial cables used in measurements are usually flexible or semi-rigid. In flexible cables the inner and outer conductor are made of braided copper wires, which are often silver-plated to reduce losses. Inner and outer conductors are separated by a low-loss insulation, such as polyethylene or Teflon. Flexible cables are used where repeated bending is required. Semi-rigid cables are solid and can be bent only a few times, generally not more than ten times without damaging the cable. Compared to flexible RF cables, semi-rigid ones have lower losses, better isolation and at the same time less leakage radiation. Therefore, in very accurate measurement semi-rigid cables are good to use. For high-quality coaxial cables one should expect a good phase stability, i.e. the electric length of the cable should remain constant when the cable is bent or the temperature changes.

The most important characteristics of coaxial cables (see Table 3) are characteristic impedance, bandwidth, losses and power handling capability. In microwave measurements one almost always uses 50-Ω cables. Only in few cases 75-Ω cables are used, e.g. the impedance of general (home) TV antenna cables is 75 Ω. Most common cable diameters are 3.58 mm and 6.35 mm. With an increasing diameter attenuation decreases and power rating increases (see Fig. 15), but the operating frequency decreases. To ensure that no higher wavemodes than TEM waves propagate in the coaxial cable, the wavelength λ inside the insulation must meet the condition $\lambda \geq \pi(r_o + r_i)$, where r_o is the inner radius of the outer conductor, and r_i is the radius of the centre conductor. The lowest order undesired (circular waveguide) waveform is TE_{11}.

Example: The dimensions of a semi-rigid SMA cable (with an outer diameter of 3.58 mm) are r_o = 1.49 mm, r_i = 0.455 mm; and the permittivity of Teflon ε_r = 2.07. Now the wavelength inside the insulation should be greater than 6.1 mm. The operating frequency range is therefore up to $c/(\sqrt{\varepsilon_r}\lambda)$ = 34 GHz. At higher frequencies TE- and TM-wavemodes will start to appear.

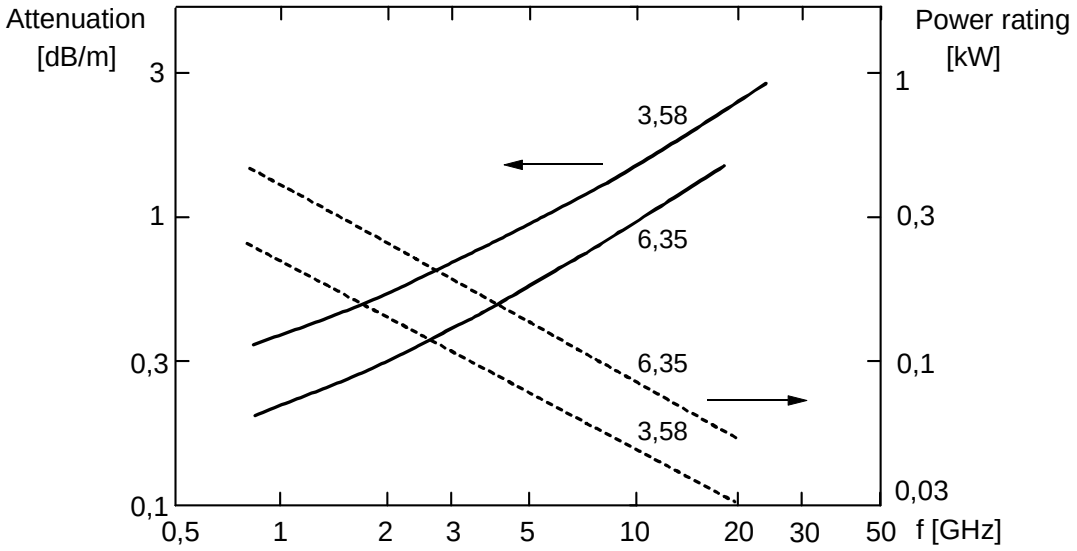

Figure 15. Attenuation and average power rating of a semi-rigid coaxial cable. Cable diameters 3,58 mm and 6,35 mm.

The useable frequency band of a rectangular waveguide is about 1.2–1.9 times the cut-off frequency $f_c = c/2a$ (c speed of light, a wider side of cross section). In this band only the TE_{10}-mode can propagate inside the waveguide. Below the cut-off frequency waves can not propagate at all, and slightly above this the attenuation is very large. At frequencies above $2f_c$ higher wavemode can propagate. To cover the whole microwave and millimeter wave region we need as many as twenty standard waveguide sizes. In Fig. 16 the attenuation and peak power rating of standard waveguide from 18 GHz to 300 GHz is shown. The peak power rating is limited by the breakdown field strength of air, the average power rating by the heating of the waveguide.

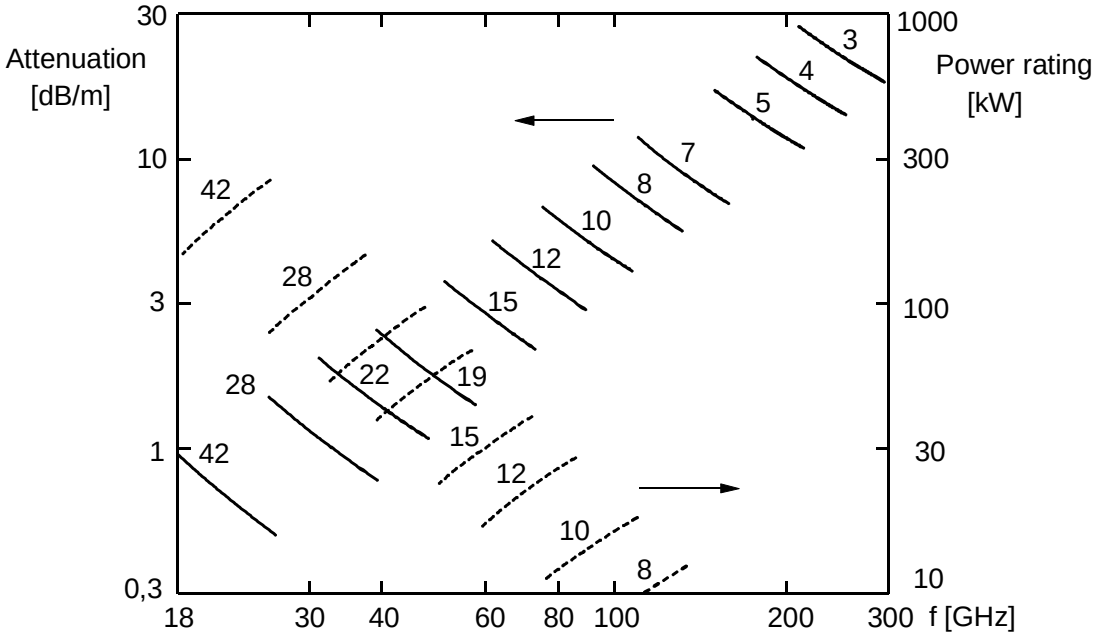

Figure 16. Attenuation and peak power rating of waveguides WR-3 through WR-42.

Connectors and Adapters

Microwave cables and devices are connected with a wide range of coaxial connectors. Common microwave connectors are APC-7, SMA, BNC and N connector (see Fig. 17). Also SMB, SMC and TNC connectors are suitable for microwaves. At higher frequencies the 3.5-mm (APC-3.5) and 2.4-mm (APC-2.4) connector are more common. Up to the lower end of the millimeter wave range we find the 2.92-mm K-connector and the 1.85-mm connector. There are also 1.0-mm connectors working up to 110 GHz. The maximum usable frequency of a connector increases when the diameter of the line decreases, as shown in Fig. 18.

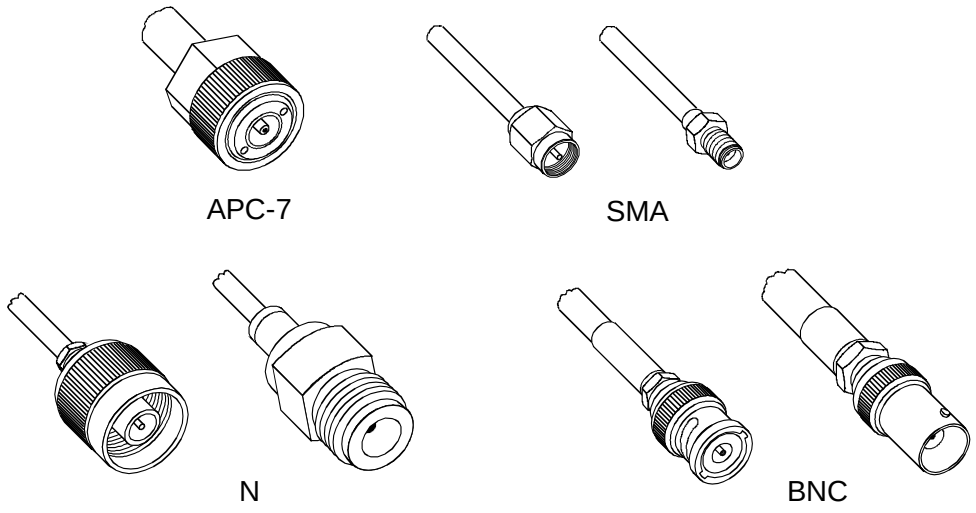

Figure 17. APC-7-, SMA-, N- and BNC-connectors.

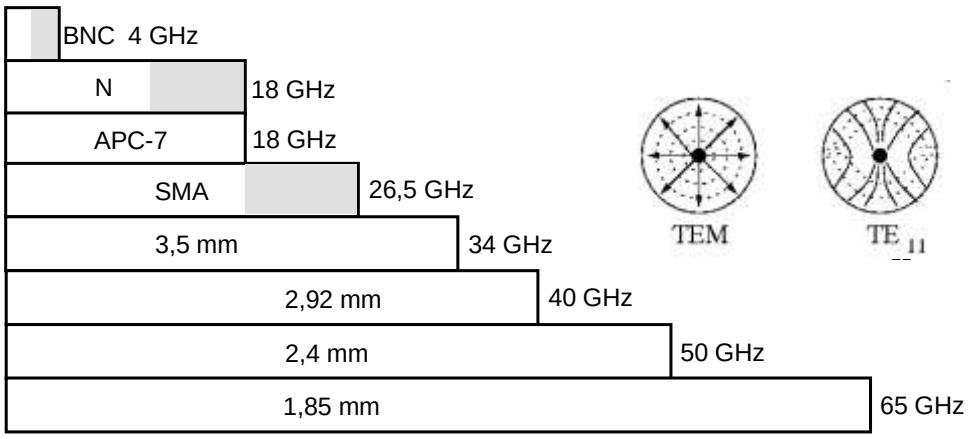

Figure 18. Typical frequency ranges of coaxial connector types. For some types the specified frequency range varies, e.g. the N-connector can be specified up to 11 or 18 GHz.

Connectors need to provide good repeatability and impedance matching. The APC-7 connector is a precision connector, used in measurements requiring very good repeatability. The APC-7 connector is a gender-less connector, whose outer conductor has an inner diameter of 7 mm. In most microwave measurements, though, we will use SMA and N connectors. These are of either male or female type. BNC connectors are only suitable for low-frequency (less than 2 GHz) measurements.

Since we often end up connecting components with different types of connectors (or same gender) a wide range of adapters is usually needed. Sometimes also a coaxial adapter with a 90 degree bend may facilitate some connections in a measurement setup.

Rectangular waveguides are connected through flanges. Pins ensure a correct alignment. An adapter for connecting a rectangular waveguide to a coaxial cable is illustrated in Fig. 19.

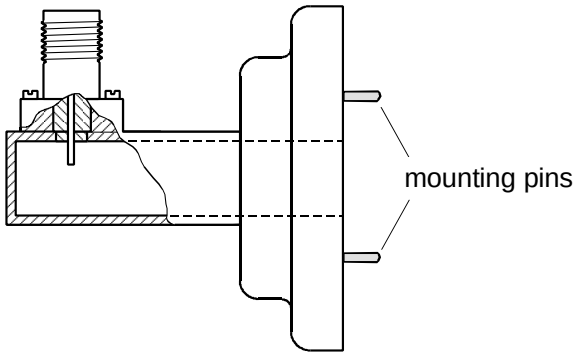

Figure 19. SMA-to-waveguide adapter.

Connectors and adapters have to be used correctly. Negligence easily damages microwave components. One of the most common cause of errors in microwave measurements is a faulty connector or adapter. Follow these instructions on the use of coaxial connectors:

- Store connectors carefully. Use plastic caps. Do not leave connectors rolling around in the storage boxes.
- Visually check each connector. Is the center pin straight and at the right height? Are the contact leads of the female connector okay? Are any contact surfaces damaged? Dispose off broken connectors (i.e. hand them to the lab supervisor).
- If necessary check the mechanical dimensions of the connector with a dial gauge. The correct location of the center pin relative to the outer conductor is important. Familiarize with the use of a dial gauge.

- Clean the connector if necessary. Compressed air in cans, cotton tips and cleaning liquids such as isopropanol, can be used.
- Tighten the connector correctly. Align the connectors carefully; press them against each other without twisting them sideways. Do not tighten the connector too tightly, but also a loose connection works badly. The suitable force is obtained with a torque wrench. Learn how to use a torque wrench.
- Tighten the connectors by rotating the connector *sleeve*, not the center parts. When rotating the center parts against each other the contact surfaces may be damaged.
- Do not let a device hang freely at the connector.
- Protect the terminals of expensive measurement equipment with a redundant adapter.

Attenuators

Attenuators are quite often needed in microwave measurements. One uses them to reduce power level to protect a measurement device, to improve impedance matching or when measuring power ratio. There are coaxial and waveguide attenuators, with a fixed attenuation, and continuously or incrementally adjustable attenuators.

Coaxial fixed attenuators are usually T- or π-type circuits made of resistive series and shunt elements according to Fig. 20. Attenuator standard values are 1, 2, 3, 6, 10, 20 and 30 dB. The accuracy at the frequency band 0-12 GHz would be e.g. $\pm$ 0.3 dB for a 3-dB attenuator, or $\pm$1 dB for a 30-dB attenuator. Often attenuators are calibrated at a few frequencies. A typical VWSR is less than 1.2 up to 10 GHz and slightly worse at higher frequencies. A typical power rating is 2 W. For high-power attenuators cooling plates are used.

Waveguide attenuators can be realized according to Figure 21(a), by placing a resistive sheet in the waveguide so that the sheet is coaligned with the electric field of the TE_{10} wavemode. To improve matching the resistive sheet is gradually widening into the waveguide. Attenuation can be tuned by adjusting the width of the sheet.

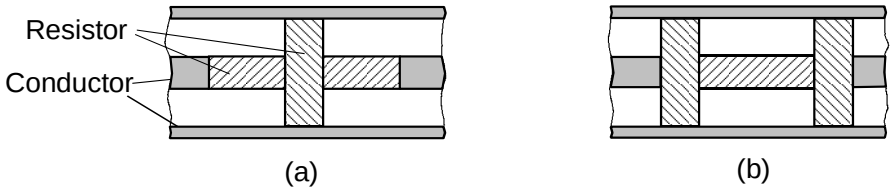

Figure 20. Coaxial attenuator design: (a) T-type and (b) π-type attenuator.

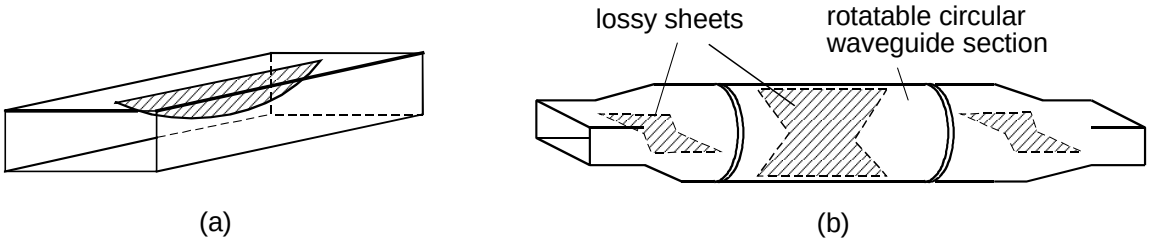

Figure 21. Waveguide attenuators: (a) fixed resistive sheet, (b) rotatable resistive sheet.

Figure 21(b) illustrates an adjustable waveguide attenuator with two rectangular waveguide ends and a rotatable circular waveguide section in between them. In all three sections are resistive sheets whose wedge-shape reduces reflections. When the TE_{10}-wavemode fields propagate from the rectangular to the circular waveguide section they transform into the TE_{11}-wavemode nearly unattenuated, because the first resistive sheet is oriented per-

pendicular to the incoming electric fields. When the sheet in the circular section is also in the horizontal orientation, the attenuation is very small. When the circular waveguide section is rotated, the electric fields of the TE_{11}-wavemode will have a component parallel to the resistive sheet, which then absorbs this component. The remaining perpendicular component propagates to the second transition, where its horizontal component is absorbed by the third resistive sheet. The remaining vertical fields transform back into the TE_{10}-wavemode. Suppression in decibels is derived from the formula $L = -40\log(\sin\theta)$ where θ is the angle between the electric field and the resistive plate in the circular waveguide section [13]. The adjustable attenuation range can be 0-60 dB.

Incrementally adjustable attenuators consist of a set of fixed attenuators, which can be switched between fixed contacts. Typical increments are 0.1, 1 or 10 dB. A good attenuator has a repeatability of 0.03 dB and a lifetime of more than one million switching operations. Incrementally adjustable attenuators are often electronically controlled.

Terminations

Matched, short-circuited and open terminations are components often needed in measurements. They are used as terminations e.g. in the calibration of vector network analyzers. An ideal matched termination absorbs all incident power, i.e. its reflection coefficient $\rho = 0$. This means the transmission line is terminated with its characteristic impedance. A matched waveguide termination can be realized by placing a wedge, plate or pyramid made of lossy material in the waveguide (see Fig. 22). The length of such a termination should be at least one wavelength.

The location of the matched termination along the transmission line is often adjustable. By repeating a measurement with the matched termination located at different locations one can eliminate the effect of small reflections from the termination. The VSWR of a good matched termination is under 1,01.

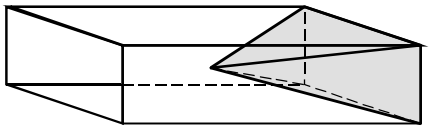

Figure 22. Pyramid made of resistive material acts as matched waveguide termination.

A short-circuited termination reflects all incident power, i.e. its reflection coefficient $\rho = -1$. There are fixed but also adjustable short circuits, with help of which one can create a desired reactance at the measurement plane. The impedance at the distance *d* from the short circuit is

$$Z = jZ_0 \tan\beta d ,$$

where the phase constant $\beta = 2\pi/\lambda_g$ (λ_g is the wavelength inside the transmission line). A tightly fitted metal piston is a simple adjustable short circuit; however, the disadvantage of this solution is the uneven contact between the piston and the outer conductor. A more reliable solution consists of several quarter-wavelength long waveguide sections as shown in Fig. 23. A transmission line of length $\lambda_g/4$ acts as an impedance transformer so that the resulting short-circuit impedance is

$$Z_s = \left(\frac{Z_1}{Z_2}\right)^2 Z_s' .$$

When Z_s' is small and Z_2 much larger than Z_l, Z_s is close to the ideal short-circuit impedance 0 Ω. This short circuit cannot be used over a wide frequency band since the length of the quarter-wavelength waveguide sections is $\lambda_g/4$ only at one frequency. The bandwidth is about 10 %.

Similar to the short circuit, also the open termination reflects all incident power, but ρ = +1. The field lines of an open coaxial cable extend beyond the physical end of the cable (fringing fields), so the electric length of the cable is slightly larger than the physical length. The extension of the cable can be modeled with an end capacitance. This capacitance can be calculated accurately and taken into account when using this open termination in a calibration. The open end of a waveguide does not suit as an open termination, since an open waveguide acts as an antenna.

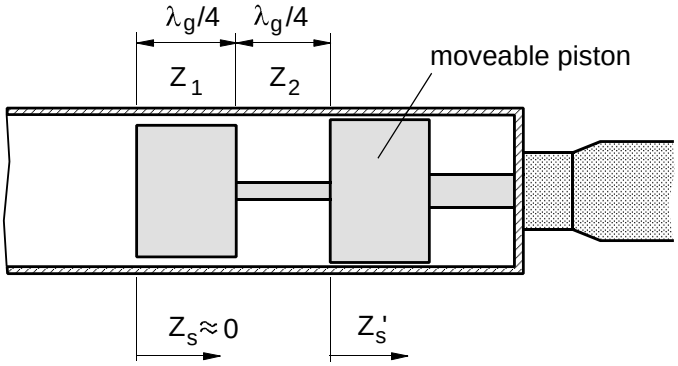

Figure 23. Adjustable short-circuited termination.

Directional couplers

With help of a directional coupler we can take a well-defined fractional sample of the power propagating in a transmission line. Directional couplers are important components in circuit analyzers when determining the reflection coefficient, generally for monitoring power levels or adjusting the power level in ALC-circuits, and finally for measuring the power without interfering with the signal propagating in the transmission line or when we need to reduced the power level to protect the actual measuring device.

When we feed power P_1 to the directional coupler illustrated in Fig. 24 and all the other ports are matched, then power P_4 is coupled to port 4. Due to non-idealities in a directional coupler some power P_3 goes to port 3. Coupling C and directivity D in decibels are:

$$C = 10 \log \frac{P_1}{P_4},$$

$$D = 10 \log \frac{P_4}{P_3}.$$

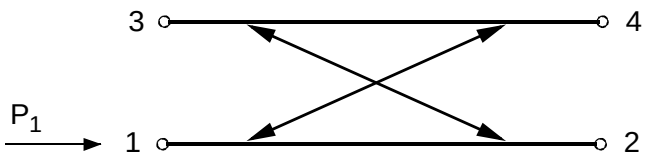

Figure 24. Directional coupler.

Standard coupling factors are 6, 10, 20 and 30 dB, and flatness of the coupling frequency response is typically ± 1 dB. The bandwidth of a coaxial directional coupler can be more than a decade. Good directional couplers can have a directivity of 30–40 dB and a VSWR of 1.1 – 1.2. The attenuation due to electrical losses is typically a few tenths of dB, somewhat depending on the frequency. Fig. 25 shows a broadband 13-dB directional coupler.

Figure 25. Wideband (1–18 GHz) 13-dB directional coupler.

Often, two directional couplers are connected back-to-back as shown in Fig. 26. With such a dual directional coupler we can measure simultaneously power P_p propagating towards the load and the reflected power P_r. The coefficients k_1 and k_2 in Fig. 26 are

$$k_1 = \left[1 - 10^{-C_2/10} - 10^{-(C_2+D_2)/10}\right] 10^{-C_1/10},$$
$$k_2 = \left[1 - 10^{-C_1/10} - 10^{-(C_1+D_1)/10}\right] 10^{-C_2/10},$$

if there are no losses or reflections. Also with a single directional coupler we could measure P_p and P_r but the accuracy would be quite poor if the matching of the ports is not perfect. If e.g. we measure P_p with a simple (i.e. unmatched) diode-detector P_r part of the reflected power would go directly to the power detector measuring P_r . With a dual directional coupler these problems are significantly reduced since the forth port of each coupler is terminated with a high-quality matched load.

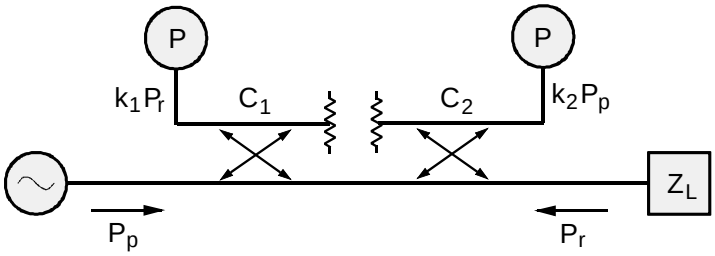

Figure 26. Dual directional coupler.

Example: We measure at port two the power reflected from the load at port three. The generator is at port one, and a matched load is connected to port four. The reflection coefficient of the attenuator under test is $|\rho| = 0{,}1$ i.e. the return loss $L_{retn} = 20\log(1/|\rho|) = 20$ dB. If the directivity $D = 30$ dB, the measurement result is in the range $L_{retn} = 17{,}6–23{,}3$ dB, depending on the phase difference of the wave that is to be measured and the additional wave caused by the non-zero directivity. If D was as large as L_{retn}, the measurement result for

L_{retn} would vary between −6 dB and ∞ dB. The mismatch of the matched load affects the measurement result in the same way. An accurate error analysis requires taking the mismatch also of the generator and the power meter into account, as well as all the multiple reflections that will occur.

Isolators

Isolators allow power to propagate only in one direction. Power propagating in the reverse direction is absorbed inside the isolator. Isolators are used to match the output of generators, and in measurements of oscillators or amplifiers that are very sensitive to the load impedance.

Waveguide-isolators can be based on the Faraday rotation, i.e. rotation of a propagating field orientation in a ferrite that is placed in a magnetic DC field. Isolators are available for frequencies up to over 200 GHz. Typical characteristics of a 100-GHz isolator are: insertion loss below 3 dB, and a ratio between reverse and forward insertion loss of over 25 dB, VSWR of less than 2.

Phase shifters

With a phase shifter we can change the phase difference between a signal and a reference signal. A tunable phase shifter can be realized in many ways. Coaxial and microstrip phase shifters are often based on the use of pin-diodes. Waveguide phase shifters can be realized e.g. by changing the magnetic field in a ferrite, by changing the biasing of a semiconductor placed on the wall of a waveguide (effective width of the waveguide changes), or by rotating a dielectric sheet inside a circular waveguide [13]. The latter looks similar to the attenuator illustrated in Fig. 21(b).

[13] Collin R.E.: *Foundations for Microwave Engineering*, 2nd Ed. New York 2001, IEEE Press.

[14] Pozar D. M.: *Microwave Engineering*, 2nd Edition. New York 1998, John Wiley & Sons.

[15] Räisänen A., Lehto A.: *Radio Engineering for Wireless Communication and Sensor Applications*. Norwood 2003, Artech House, Inc..

Diode detectors

The detection of a signal means to change it into a observable, usable form. In a diode detector, the high-frequency signal is transformed into a voltage (or current) that is proportional to the signal power. In case of an unmodulated signal the detector produces a DC voltage. From an amplitude-modulated signal the diode extracts the modulating signal. This low-frequency signal is known as a video signal.

When the voltage output of the detector is directly proportional to the RF power, a calibrated diode is well suited for power measurements (see p. 38ff.) and power monitoring. Often, the exact response of the detector does not need to be known. Those applications include, e.g., power control with automatic level (ALC) and gain control circuits (AGC), power ratio measurement with a tunable attenuator, and null-detector measurement bridges.

The same diodes that can act as detectors also work in superheterodyne receiver mixers. The sensitivity of diode detectors is much worse than that of superheterodyne receivers, but due to their simplicity these detectors are more useful in many applications.

Principle of diode detectors

Detectors are based on the nonlinearity of semiconductor diodes [16, 17]. If we connect a sinusoidal voltage to the diode, the current output contains additionally to the input signal the harmonics of the input RF signal, and a DC component that is proportional to the RF power.

The most common diode type used in detectors is the Schottky diode, which is based on a metal-semiconductor interface. The equivalent circuit of a Schottky-diode (see Fig. 27) includes junction resistance r_j, voltage dependent capacitance c_j and series resistance R_s. Of these, R_s and c_j are parasitic elements that weaken the diode characteristics. The current-voltage characteristic (Fig. 28) of the junction resistance is of exponential form:

$$i = I_s(e^{\alpha v_j} - 1),$$

$$\alpha = \frac{q}{\eta k T},$$

where v_j is the voltage across the junction, I_s is the saturation current of the diode, η is the ideality factor of the diode, q is the unit charge ($1{,}60\times10^{-19}$ C), k is the Boltzmann constant ($1{,}38\times10^{-23}$ J/K), and T is the temperature. The ideality factor of a good diode is slightly above 1. The junction resistance is

$$r_j = \left(\frac{di}{dv_j}\right)^{-1} = \frac{1}{\alpha(i + I_s)}.$$

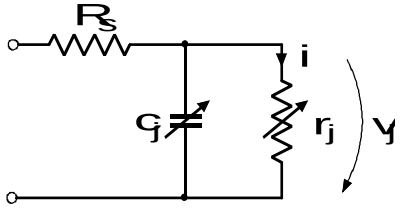

Figure 27. Equivalent circuit of a Schottky-diode.

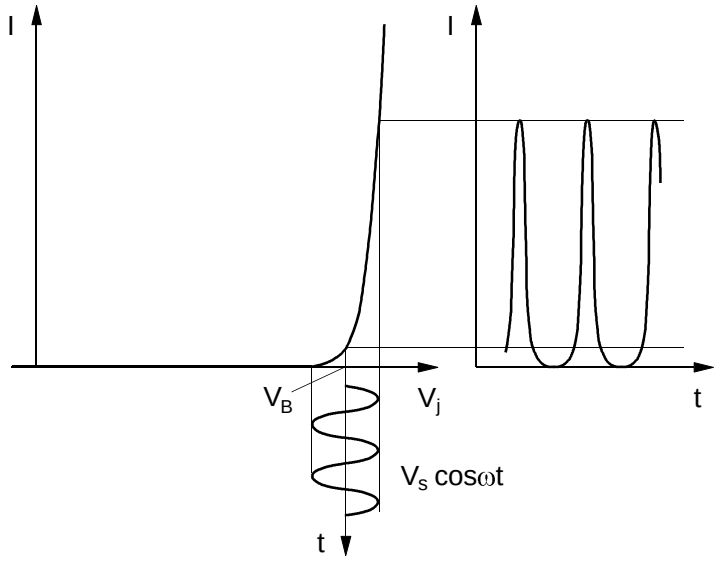

Figure 28. The exponential *I*–*V*-characteristic of a diode, and the current waveform *I*(*t*), when a biasing voltage along with a sinusoidal signal are applied across the diode.

Let us first omit R_s and c_j. When we apply a bias voltage V_B and a small sinusoidal signal $V_s\cos\omega t$ (Fig. 28) across the diode series resistance r_j, we can derive the Taylor series for the current:

$$i = I_B + \alpha^2(I_s + I_B)\frac{V_s^2}{4} + \alpha(I_s + I_B)V_s\cos\omega t + \alpha^2(I_s + I_B)\frac{V_s^2}{4}\cos 2\omega t + \ldots$$

The influence of the higher-order terms is small when $\alpha V_s << 1$. So, at room temperature $V_s << 1/\alpha \approx 25$ mV. I_B is the DC current $i(V_B)$ due to biasing.

The second term on the right side of the equation is the DC component that is proportional to the RF signal power, i.e. the actual desired output signal. We can consider the junction a voltage source with the voltage $V_o = \frac{\alpha V_s^2}{4}$ and the internal resistance r_j. We consider the diode a square-law detector, because ideally the desired output signal is directly proportional to the power of the measured RF signal, or in other words V_o is proportional to V_s^2.

The voltage sensitivity is the ratio of the output voltage change ΔV_o and the input RF power P_s, when the output load is open (or at least $R_L >> 50\ \Omega$). When the ideal diode is matched at the microwave frequency we get

$$\beta_v = \frac{\Delta V_o}{P_s} = \frac{\alpha r_j}{2}.$$

Accordingly, the current sensitivity, i.e. the ratio of the short-circuit current change and input RF power, is

$$\beta_i = \frac{\Delta I_o}{P_s} = \frac{\alpha}{2}.$$

The characteristic curve, and thus voltage and current sensitivity of the diode, depend strongly on temperature.

When we take the diode's parasitic components R_s and c_j , as well as the finite load at video frequency R_L (see Fig. 29) into account, we obtain as the loaded voltage sensitivity, i.e. the ratio between the change of the voltage across the load ΔV_L and the RF signal power P_s in the matched case [17]

$$\beta_{v,loaded} = \frac{\Delta V_L}{P_s} = \frac{\alpha r_j}{2} \times \frac{R_L}{\{1+(R_s/r_j)[1+(\omega c_j r_j)^2]\}(r_j + R_s + R_L)} .$$

Because of the series resistance R_s and the junction capacitance c_j , not all the power P_s goes into the junction resistance r_j. The voltage sensitivity decreases as frequency increases, because an increasing part of the current bypasses the junction resistance through the junction capacitance. The above equations are therefore valid only for small signal levels. A large-signal analysis of a diode is much more complex.

Without bias voltage the junction resistance of a normal Schottky diode can be more than 30 MΩ and the diode can hardly be matched to the circuit. In addition we saw above, that in order to maximize the voltage sensitivity, R_L should be much greater than r_j. Such a diode requires a small biasing (10 – 30 μA), so that r_j is tuned properly. Because of this, often Schottky diodes with a small barrier voltage (less than 0.2 V) are used in detectors, which accordingly also without biasing show a sufficiently small r_j of about 2 kΩ.

In Fig. 29 we can see how to connect a detector diode to an actual RF circuit. To ensure that all RF power is available at the diode, we need to match it to the transmission line, i.e. usually a 50-Ω coaxial cable or a waveguide. The additional stray inductance and capacitance caused by the packaging can be thought to be included in the matching circuit. Matching the diode across a wide frequency band is challenging. In practice one often desires flat frequency response and contents with smaller voltage sensitivity than discussed in the previous section. Then, a 50-Ω resistor in parallel with the diode can act as a wideband matching circuit. The low-pass filter allows the DC or low frequency video signal from the detector to reach the load while it prevents the high-frequency signal and its harmonics to do so. The coil and the capacitor are needed, so that both the DC and the high-frequency currents can reach the diode.

Fig. 30 illustrates the typical power response of a diode. When the power level increases above –20 dBm the operation does no longer follow the square-law but gradually behaves linearly. At the linear range the output voltage V_L is approximately proportional to the signal voltage V_s. At even higher output powers the diode saturates and is eventually destroyed. At low power levels noise limits the square-law operational region.

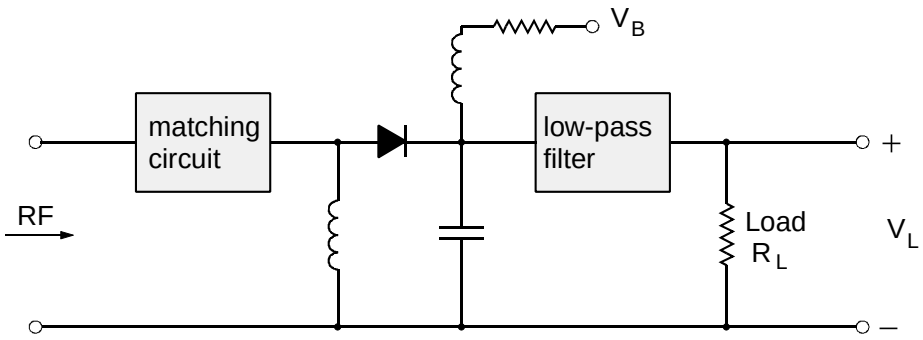

Figure 29. Diode-detector circuit.

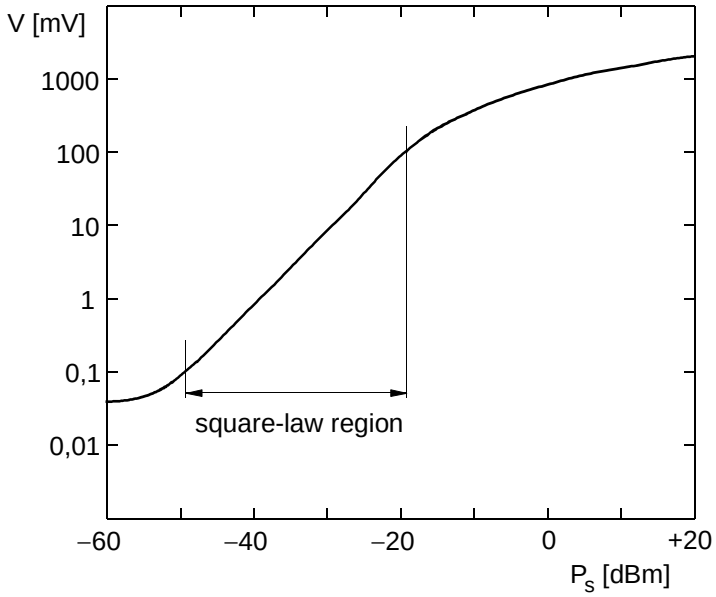

Figure 30. Diode-detector power response.

Noise in detector diodes

The noise of a detector diode consists of thermal noise from the series resistance, shot noise from the semiconductor junction, and low-frequency 1/*f*-noise. The sensitivity is also limited by the noise generated in the video amplifier that is often connected to the detector.

The sensitivity describes the ability to detect weak signals amidst noise, and should not be confused with voltage sensitivity. Typically the tangential sensitivity P_{tss} is used as a measure of the sensitivity of a detector. The detected signal power is equal to the tangential sensitivity, when the change in output voltage is equal to the peak-to-peak noise fluctuations (see Fig. 31). This is a subjective quantity, since there are no unambiguous extreme values for the random voltage variations. The voltage that corresponds to the tangential sensitivity is about 2.8 times the rms noise voltage. The tangential sensitivity is therefore $P_{tss} = \frac{2{,}8 v_n}{\beta_v}$ where v_n is the rms noise voltage and β_v is the voltage sensitivity. Typically the sensitivity of a detector diode is –50... –55 dBm for a video bandwidth *B* of 10 MHz.

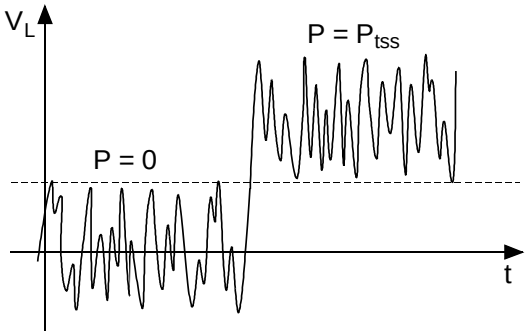

Figure 31. Tangential sensitivity.

Detectors in practice

Commercial coaxial diode detectors (see Fig. 32) can be found up to 50 GHz. These already contain the RF matching circuit and a video signal filter. Biasing is not needed. Typical characteristics are voltage sensitivity ≥ 500 μV/μW, frequency response flatness ±0.3 dB, voltage standing wave ratio <1.3 and a maximum input power 200 mW. It should be noted that the voltage sensitivity is given for the case that the load is open ($R_L = \infty$). If e.g. a 50-Ω oscilloscope is connected, the reading is lower than the value calculated based on the voltage sensitivity. One can find a variety of input connectors, and output voltage is, depending on the diode orientation, either positive or negative. The detector can be optimized for broadband operation (resistive matching), which means a compromise with the sensitivity, or for narrow-band operation (reactive adaptation), with the highest possible sensitivity.

Waveguide detectors are often tunable, i.e. the location of the short circuit can be tuned to optimally match the diode at each frequency. This will provide better sensitivity but a narrower bandwidth than with fixed-frequency detectors.

At this point it is good to remember that static charges easily destroy a diode detector! It is therefore best to ground oneself before handling any detectors in the measurement.

Figure 32. Diode detector.

[16] Räisänen A., Lehto A.: *Radio Engineering for Wireless Communication and Sensor Applications*. Norwood 2003, Artech House, Inc..

[17] Bahl I., Bhartia P.: *Microwave Solid State Circuit Design*, Second Edition. Hoboken, New Jersey 2003, John Wiley & Sons.

Measurement of power, attenuation and gain

Power is one of the main parameters describing the operation of radio equipment and systems. The transmit power affects radio link span and quality, while in the receiver the input and output power levels of each circuit block should be within well-defined limits for the receiver to operate properly. Measuring the power [1] is one of the most common microwave measurements - power must be measured accurately and repeatedly. We also look at the measurement of power ratios, i.e. of attenuation and of gain.

Microwave power can not be measured like at low frequencies by measuring the voltage and current separately and calculating their product. Although in microwave circuits we find electromagnetic fields created by voltages and currents there is no instrument to measure these. In addition, voltage and current depend on location when the transmission line is not terminated with a matched load. In waveguides voltage and current are not even unambiguous parameters.

Power is often measured by absorption. The power is absorbed in a suitable component whose characteristics change; e.g. the resistance changes with temperature. Bolometer, thermocouple and calorimeter are power sensors working with the absorption principle. Also calibrated diode detectors are well suited for power measurements.

Typically, a power meter consists of an exchangeable sensor element - the sensor head, and the meter section, which includes the electronics and display. The sensor heads of coaxial and waveguide types are built so that they can be combined cover the entire microwave region, often up to millimeter-wave range, using the same meter. A meter may have two channels, one channel measuring e.g. the input power of the device and the other at the same time the output power. Often only a well-defined portion of the actual power is measured using a directional coupler, in which case not all the power is turned into heat during the measurement.

Bolometers

In a bolometer power is absorbed in a temperature-dependent resistor, i.e. a thermistor. Thermistors consist of a piece of semiconductor whose resistance decreases with increasing temperature. By biasing the resistance of a thermistor can be adjusted over a wide range, so it can easily be matched to a circuit. Thermistors are quite slow: the time constant is 10^{-2}...1 s. The measurement range is typically from –30 dBm to +10 dBm, with a maximum input power of 15 dBm. Commercially available are coaxial sensor heads for frequencies between 10 MHz and 18 GHz, and waveguide sensor heads up to 110–170 GHz.

The power dependence of the resistance is very nonlinear and will vary from one thermistor to another. Therefore, thermistors are used so that high-frequency power is replaced by DC power (or low-frequency power). This way the operating point of the thermistor is always the same and is not affected by the non-linearity. This is achieved by placing a thermistor in self-balancing Wheatstone bridge (see Fig. 33) in one branch, which provides the bias current to the thermistor. When the microwave power heats up the thermistors, its resistance R_T decreases and the bridge will have an imbalance. Therefore the bias current decreases and the thermistor cools down until the bridge is again balanced. The DC power that corresponds to the reduction in bias current is then equal to the measured microwave power.

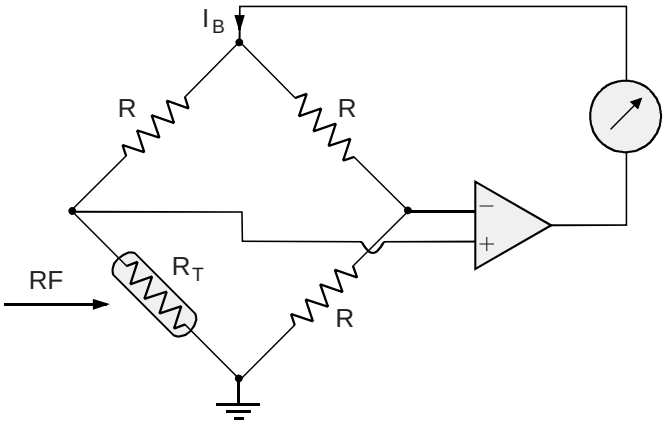

Figure 33. Thermistor in a self-balancing Wheatstone bridge.

To facilitate the connection of thermistors to RF circuits, they are used in pairs as shown in Fig. 34. For the microwave signal the thermistors are in parallel, for the DC they are in series. If the resistance of each thermistor R_T = 100 Ω the microwave resistance is 50 Ω whereas the DC resistance is 200 Ω. The thermistor resistances should be equal, otherwise the microwave power is divided differently between the thermistors than the DC power. Capacitances C_c and C_b are almost short-circuits at microwave frequencies. Therefore the two ports of the DC power supply both appear as RF ground. With only a single thermistor also a coil would be needed to prevent the microwave signal from leaking into the meter section, and a coil would significantly limit the frequency range of the sensor head.

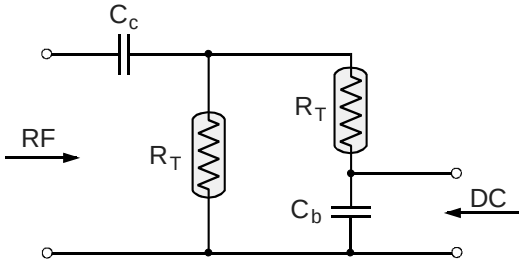

Figure 34. Pair of identical thermistors.

Thermistors are not able to distinguish between a change in RF power and a change in ambient temperature. To compensate for temperature drift a second thermistor pair is added to the first pair, but in the second pair no RF power gets absorbed. This thermistor is in its own bridge as shown in Fig. 35. The RF bridge is automatically balanced by tuning voltage V_{rf}. The other bridge will remain in balance, regardless of temperature changes by tuning voltage V_c. Initially the power meter is set to zero, without any RF power present, so that $V_c = V_{rf0}$. When the RF power is allowed to reach the thermistor, V_{rf} decreases. The change in DC power equals the RF power:

$$P_{rf} = \frac{V_c^2}{4R} - \frac{V_{rf}^2}{4R} = \frac{1}{4R}(V_c - V_{rf})(V_c + V_{rf}),$$

where R is the resistance of the thermistor pair ($2R_T$) in the balanced state. The meter unit multiplies the sum and difference of voltages. The accuracy can improved if the output voltages are directly measured with an accurate digital meter and the power calculated from these values.

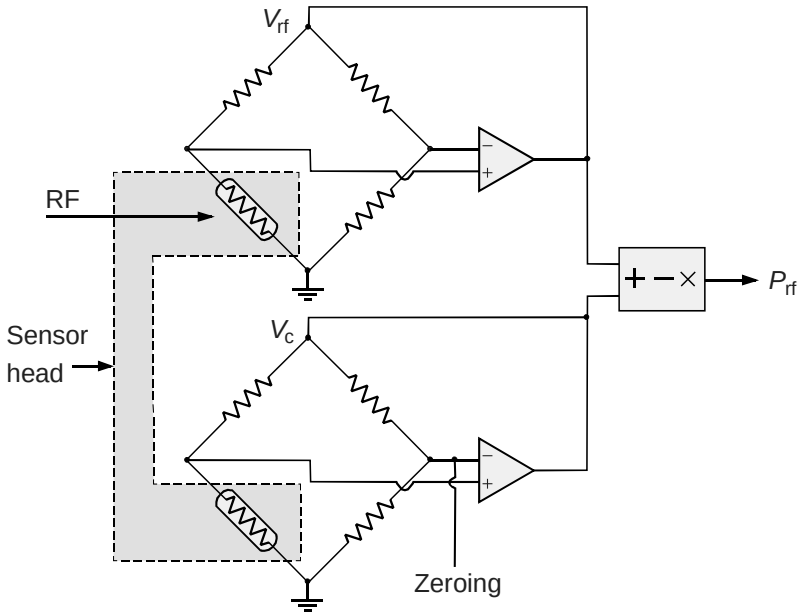

Figure 35. Simplified diagram of a temperature-compensated power sensor head [1].

Thermocouples

A thermocouple is formed when two wires made from different metals are joined at one end and the remaining two wire ends are connected to a meter. If the joint is at a different temperature than the end with the meter, there will be a thermoelectric voltage proportional to the temperature difference. A suitable metal pair is antimony and bismuth. Finally a resistor is attached to the joint which absorbs RF power and matches the joint to the transmission line.

Fig. 36 shows a semiconductor-based thermocouple. It consists of two thermocouples, which are combined in parallel at microwave frequency, and in series at DC, similar to the thermistor shown in Fig. 34. On the silicon chip a thin-film resistors is attached inside which microwave power is absorbed and converted into heat. The resistors have a resistance of 100 Ω so they form a matched 50-Ω load to the transmission line. The resistor and the n-type silicon form the hot junction, whereas the edge of the chip stays cold. This temperature difference creates the thermoelectric voltage. Since the two elements are in series, their voltages sum up. With gold leads this chip can be connected e.g. to a coplanar waveguide. The measurement sensor has a small reflection coefficient and the output voltage is 100 – 200 μV/mW. Such a sensor is less sensitive to ambient temperature changes than a thermistor.

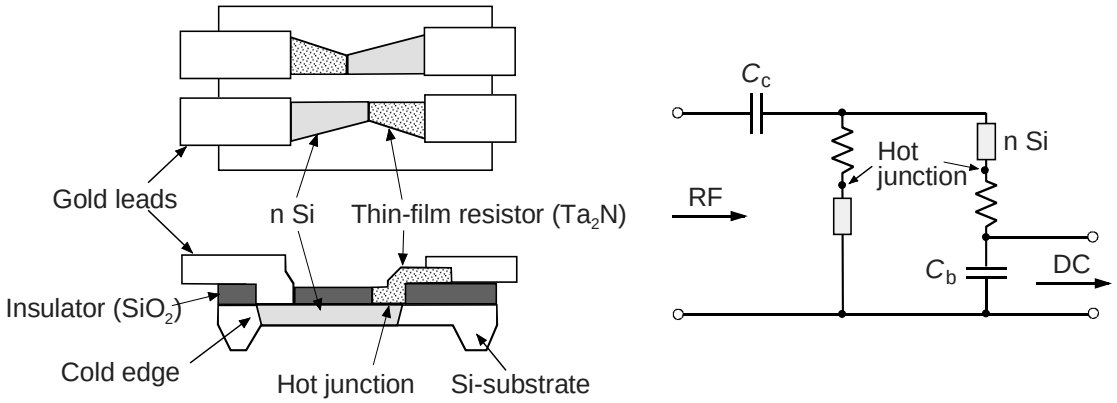

Figure 36. Thermocouple based on semiconductor- technology [18].

Fig. 37 shows a a block diagram of a thermocouple and the according power meter. The meter also works with diode sensors, since the output signal of both types of sensor is a low DC voltage. Transferring small DC voltages from the sensor to the meter is not that easy, since basically at all connections a thermoelectric voltage is created, hereby increasing uncertainty. One solution is to chop the DC voltage into a carrier wave, which is amplified before being fed into the cable. In Fig. 37 the chop frequency is 220 Hz, and FETs operate as switches. The signal is further amplified in the meter section before the actual detection. In modern meters a synchronous detector is used, and instead of the analog circuits we find an A/D converter and DSP circuits downstream, which also simplify the display of measurement result in different formats.

Before a measurement the display is zeroed without and RF signal and the calibration factor (see p 40ff.) is chosen. Since the thermocouple method is not based on measuring DC power, such as thermistor bridges, an accurate reference oscillator is needed to correct the sensor drift. The reference oscillator frequency is 50 MHz and the power 1 mW.

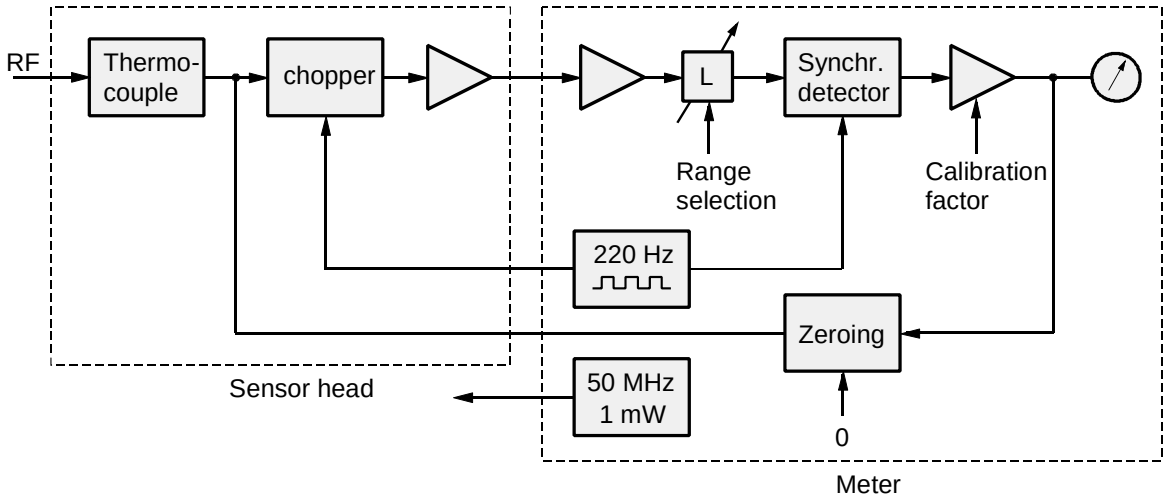

Figure 37. Thermocouple-based sensor head and power meter.

The typical measurement range of a thermocouple is from –40...–30 dBm up to +20 dBm, and a maximum input power of +25 dBm. The time constant of the sensor head is for small power in the range of seconds and for higher powers down to 0,1s. Coaxial sensor heads are available up to 50 GHz and waveguide-based up to up to some 90–140 GHz.

Diode sensors

The operating principles of diode detectors were already discussed. A calibrated diode detector can be used to measure power, but the accuracy is not very good. Diode detectors are e.g. sensitive to temperature changes.

Fig. 38 shows a diode sensor for power measurement, which works better than simple diode detectors and can be used in connection with the meter section shown in Fig. 37. Since the impedance of the diode depends on temperature and power level, the sensor would be poorly matched if the diode was directly connected to the transmission line. Therefore, a 50-Ω resistor acts as the load and the high-impedance diodes are in parallel, resulting in good matching. Two diodes form a balanced structure. This way, possible thermoelectric voltages at the connections and effects of even-numbered harmonics are avoided. From Fig. 37 we are already familiar with the sensor's FET-based chopping unit and amplifier. The sensor head can also feature two separate paths after the diodes for the measurement of large and small powers.

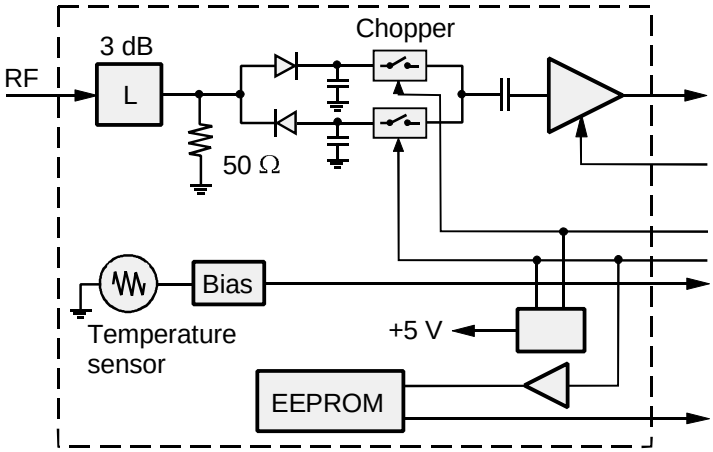

Figure 38. Diode sensor [1].

The square-law region of the sensor, in which the output voltage is directly proportional to the RF power, ranges from about –70 dBm to –20 dBm. This power range can be extended to +20 dBm by calibrating the diode's power response. The sensor head contains the memory (EEPROM, Electrically-erasable programmable read-only memory), in which is stored the calibration data as a function of power level, frequency and temperature. This data is unique for each sensor, and they can be updated during the regular calibration of a sensor. Whenever the power meter is switched on or the sensor head is changed the sensor data is downloaded to the meter's memory. Through the temperature sensor in the sensor head, the meter obtains the ambient temperature. Before each measurement one calibration is done using the reference oscillator of the meter. A high-precision 30-dB attenuator is to be connected between the oscillator and the sensor to reduce the power from 0 dBm to -30 dBm (i.e. stay within in square-law region).

The time constant of the diode sensor is short. Combined with a suitable meter, fast diode sensors can be used for non-recurring or repetitive power variations, and in the measurements of peak power. In this case, the meter is part of an A/D converter, which then takes an output voltage sample e.g. tens of millions of times per second. From those samples the meter can determine power characteristics of pulsed signals (e.g. GSM-pulses). The power can also be measured during an adjustable time gate, e.g. near the rise or fall of the pulse, or during the whole pulse duration.

Coaxial diode sensors are available up to 50 GHz, and waveguide-based up to 110 GHz.

Calorimeters

Calorimetry is a fundamental method of measuring the power, because it is directly links electrical power to the measurable heat flow, i.e. to the definition of watts, joules per second. Through calorimetry microwave power standards are traceable to DC standards. Traceability means that the measurement results are linked in an uninterrupted reference chain to national or international measurement standards. In standard laboratories calorimeters are in fact used as measurement standards to allow the calibration of other types of power sensors.

A calorimeter measures the increase in temperature due to microwave radiation e.g. in a liquid. Calorimetry can be divided into static and flow calorimeters (Fig. 39), where the static calorimeter transfers heat from the RF load by conduction and the latter by flow. In compensated calorimeters the temperature is kept constant by reducing the low-frequency power, so that the measured RF power is equal to the reduction in low-frequency power. In differential calorimeters one of two identical calorimeters is heated by the RF load and the second by an equal low-frequency power so that the temperature difference is zero.

Calorimeters are suitable for high power levels up to more than 1 kW. Their slowness is their weakness: stabilizing of the temperature usually takes minutes; especially when measuring small power levels the time constant is long. Commercially calorimeters are available up to 300 GHz.

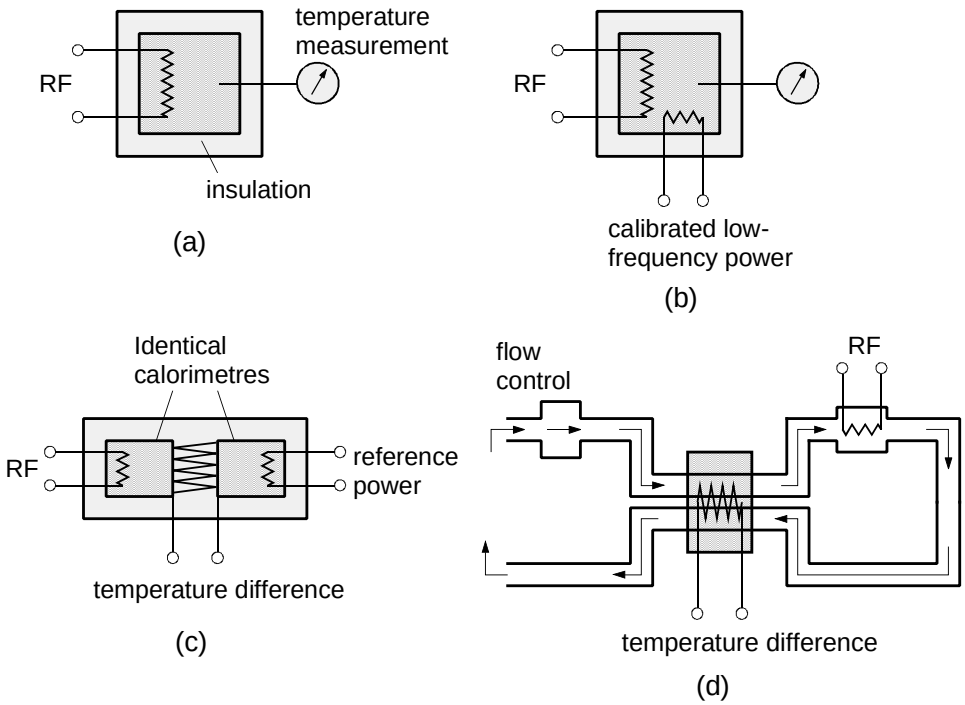

Figure 39. Different calorimeter types: (a) direct, (b) compensated, (c) differential, (d) flow.

Comparison of power sensors

Which power sensor is best suitable always depends on the type of measurement, since the importance of various factors, such as frequency and power range, measurement speed, accuracy and repeatability varies from one measurement to another.

In Fig. 40 we compare the power ranges of the different sensor types. Thermistors and thermocouples are well suited for measuring the power levels normally encountered in microwave measurements and are often fast enough. Thermocouples have partly replaced thermistors due to their better matching, thermocouple also have a wider measurement range. Diode sensors have a largest measurement range, up to 90 dB, and they are fast. Thermocouples and diode sensors are

thus the most used sensor types. With help of attenuators the power ranges of these sensors can be extended to higher power levels. Calorimeters are suitable for measuring high power and cover the entire millimeter-wave region, but they are very slow.

Thermocouples and thermistors are based on RF power absorption and the according temperature rise. Therefore, they are also suitable for measuring average powers of complex modulated and amplitude-varying signals. They are fairly slow, so instantaneous powers can not be measured. For pulsed signals the peak power can be calculated, if the pulse ratio is known.

The diode sensor detects the RF load voltage and is calibrated with an unmodulated signal, so it may give incorrect results when measuring the average power of modulated signals. Since diode sensors are fast, they are suitable for measuring the instantaneous power of pulsed signals.

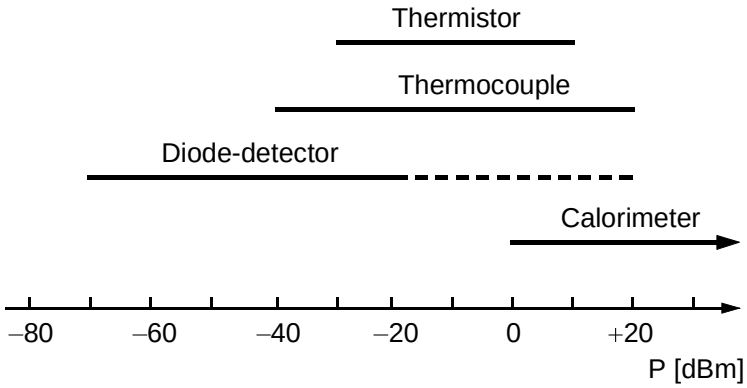

Figure 40. Measurement ranges of different sensor types.

Measurement errors

Error sources in power measurements are reflections caused by mismatch, sensor calibration errors, error in the meter, attenuation by measurement cables, noise, and incompatible signals.

In many cases, mismatch is the main error source in power measurement. Let's have a look at the example in Fig. 41 where we measure the available power from the device or generator $P_{S,a}$. The reflection coefficient between generator and transmission line is ρ_S and the reflection coefficient between the transmission line and the sensor is ρ_L. These mismatch locations cause multiple reflections. Only if the impedance of the generator and the impedance of the sensor seen through the transmission line complex-conjugate, all power is transferred from generator to sensor. Otherwise, some power does not reach the sensor. In general, the power transmission efficiency, i.e. the ratio of the power absorbed in the sensor to the available power from the source is

$$\eta_{SL} = \frac{P_L}{P_{S,a}} = \frac{(1-|\rho_S|^2)(1-|\rho_L|^2)}{\left|1-\rho_S\rho_L e^{-j\beta 2l}\right|^2},$$

where l is the length of the transmission line and β the phase constant; the transmission line is assumed lossless. The limits of the power transmission efficiency η_{SL} in decibels as a function of generator and load *VSWR*s is shown in Fig. 42.

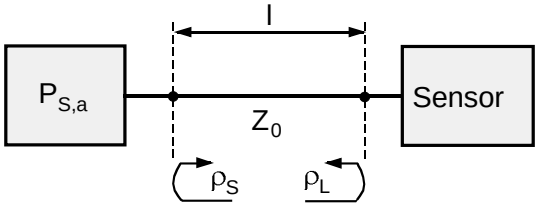

Figure 41. Mismatch in power measurement.

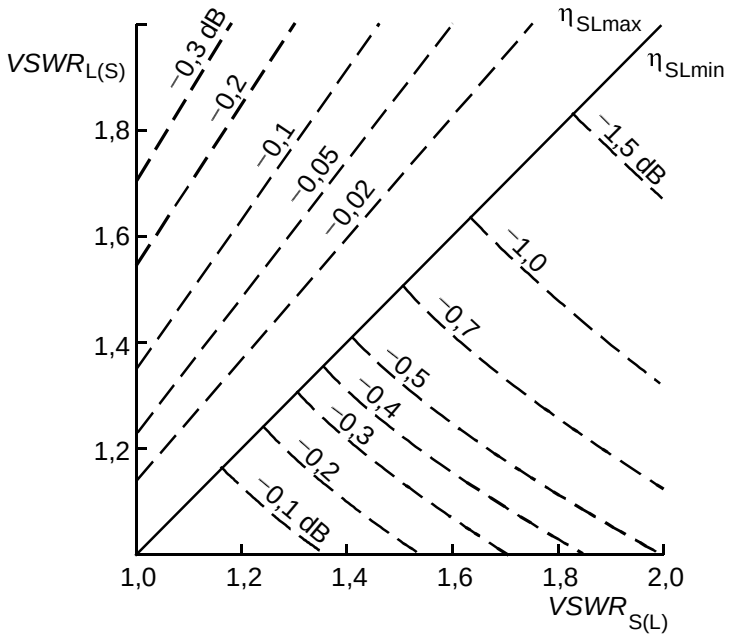

Figure 42. Power loss between generator and load due to mismatch η_{SL} [dB]. When reading $VSWR_S$ from the horizontal and $VSWR_L$ from the vertical axis, we get η_{SLmin} ($VSWR_S > VSWR_L$) or η_{SLmax} ($VSWR_S < VSWR_L$). When reading $VSWR_S$ from the vertical and $VSWR_L$ from the horizontal axis, we get η_{SLmax} ($VSWR_S > VSWR_L$) or η_{SLmin} ($VSWR_S < VSWR_L$).

Often want to know the power that the generator feeds into the transmission line, which is matched at the other end; i.e. the load impedance equals the characteristic impedance Z_0 of the line. We obtain the ration of the power absorbed by the sensor to the power that would propagate in the perfectly matched transmission line by

$$\eta_{Z0} = \frac{P_L}{P_{Z0}} = \frac{\eta_{SL}}{1-|\rho_S|^2} = \frac{1-|\rho_L|^2}{\left|1-\rho_S\rho_L e^{-j\beta 2l}\right|^2}$$

Generally we know only the absolute values of the reflection coefficients (or the *VSWR*). The unknown phase relation and electrical length of the line cause uncertainty in the transmission efficiency. The maximum efficiency we obtain when the denominator in the above equations is $(1-|\rho_S||\rho_L|)^2$, and the minimum efficiency when the denominator is $(1+|\rho_S||\rho_L|)^2$. So, when we take the measured power and calculate the available power $P_{S,a}$ or the power fed to the transmission line P_{Z0}, the uncertainty limits due to mismatch are in dB

$$U_{max,min} = 10\log(1\pm|\rho_S||\rho_L|)^2 .$$

Example: Generator $VSWR_S$ = 2,0 ($|\rho_S|$ = 0,33) and sensor $VSWR_L$ = 1,5 ($|\rho_L|$ = 0,20). When we take η_{SL} and Fig. 42 we can see that the measured power is 0,09–1,25 dB lower than $P_{S,a}$. With η_{Z0} we find that the measured power can be 0,42 dB higher or 0,74 dB lower than P_{Z0}. The range is in both cases 1,16 dB, i.e. when using the measured power P_L the uncertainty of the derived powers $P_{S,a}$ and P_{Z0} is –0,60...+0,56 dB, or about ±14 %. It should actually also be taken into account that there is already some uncertainty in the *VSWR* values.

Uncertainty caused by the multiple reflections can be reduced by adding a well-matched isolator to the output of the measured device. The uncertainty of the insertion loss of the isolator if, however, a new error source. An attenuator also reduces mismatch significantly, but the meas-

ured power level decreases. If in the previous example we add well-characterized and well-matched 10-dB attenuator to the transmission line the measured signal decreases by 10 dB, but the first reflection at the sensor is reduced by 30 dB. The measured power (corrected for the 10 dB attenuator) is now 0.63–0.75 dB lower then that available power from the generator, , i.e. the uncertainty range is now only 0.12 dB.

If the generator is matched to the transmission line $1-|\rho_L|^2$ of the power reaches the sensor. However, part of this power is absorbed in the sensor enclosure, is radiated, or leaks past the sensor. The actual ratio of power absorbed by the sensor element is described by the effective efficiency η_e. In thermistors the microwave power distributes somewhat differently than the DC or low-frequency power fed to it in the bridge circuitry. This difference is also included in the effective efficiency. The sensor calibration coefficient *K* describes the ratio of power absorbed by sensor element, or in the thermistor case DC-power change, to the power incident at the sensor $K = (1-|\rho_L|^2)\eta_e$. If there are no multiple reflections, the power *P* propagating in the transmission line can be derived from the measured power P_m as follows: $P = P_m / K$.

Sensor manufacturers specify the calibration factors at a number of frequencies. The sensor is calibrated by comparing it with a standard instrument, whose calibration is in turn based on the calibration standard of an accredited measurement standard laboratory. Each additional calibration stage in this chain increases uncertainty of the calibration factor *K*. Typically, the uncertainty of *K* is 1–5%, and it depends e.g. on frequency.

If the sensor is slow, one must wait long enough to obtain a correct measurement reading. One also has to wait when the sensor was held in the hand, since the hand warms up the sensor, affecting the measured result until the sensor has cooled down again.

In the meter part error sources are e.g. the reference oscillator, zeroing accuracy, and the so-called instrumentation uncertainty. A reference oscillator is used with diode and thermocouple sensors, but not with thermistors. Its output power uncertainty is typically around 1%. The mismatch of the reference oscillator adds its own uncertainty during calibration at the start of the measurement. The power meter is zeroed without RF power, but only the idle deflection voltage present, which itself is not exact. Instrumentation uncertainty is caused e.g. by inaccuracies in the range-selection attenuators, non-linearity of amplifiers, and display errors. A typical instrumentation uncertainty is 0.5–1%.

The uncertainty of the attenuation of the cable between source and sensor causes errors in power measurement. The significance of this error source usually increases with Frequency.

Noise limits the measurement of small power levels. The measured signal can be noisy, but also the sensor's own noise can be significant compared to the measured signal. In the latter case, the signal can be amplified to increase the power range of the meter, but the uncertainty of the gain is a new error source.

Since power meters are broadband, one has to ensure that only the desired signal is measured. At the generator output we often have harmonic frequency components in addition to the fundamental frequency. Measurement errors due to harmonic frequencies can be removed by using a band-pass or low-pass filter.

Finally, the total uncertainty of power measurements depends e.g. on measurement set-up, frequency and power level. For normal laboratory measurements in the microwave region a good value is 2–3% of the absolute power measured. Uncertainties for measuring relative power levels can be considerably more accurate, because many error sources then cancel out.

Measurements of attenuation and gain

In [20] five different definitions of gain are described. For attenuation we can use the same definitions because attenuation is simply a gain smaller than unity.

The inverse of the maximum available power gain (if this is <1) is often called attenuation L, indicating the attenuation caused by absorption. IEEE refers to it as inherent insertion loss [19]. When measuring this value for a two-port, its input is conjugate matched to generator and the output is conjugate matched to the load:

$$L = \frac{P_{in}(\rho_S = \rho_{in}^*)}{P_{out}(\rho_L = \rho_{out}^*)}.$$

Insertion loss is the ratio of power available directly from the generator P_d to the power available from the two-port:

$$L_i = \frac{P_d}{P_L}$$

i.e. in this case we do not require matched impedances at either end. Similarly, we can define the insertion gain for an amplifier.

The insertion loss or gain of a component can be obtained by measuring the power without and with the component. Another way is to add an adjustable high-precision attenuator and adjust the power level so that the output is identical without and with the component, as shown in Fig. 43 (a). The attenuation or gain of the two-port is then the difference (in decibels) of the attenuator readings. The advantage of this method is that the accuracy of the power sensor is not relevant, but the whole accuracy depends on the adjustable attenuator.

Attenuation or gain can also be measured at the intermediate-frequency or audio frequency. At low frequencies more accurate adjustable attenuators can be manufactured than at microwave or millimeter-wave region. In Fig. 43 (b) the microwave signal is mixed down to the intermediate frequency and the measurement is then performed with an adjustable attenuator. The power level should be low enough for the mixer to operate within its linear range. With a sensitive mixer a very wide dynamic range can be achieved: up to more than 100 dB. This requires that the leakage radiation is very small. In Fig. 43 (c) the microwave signal is amplitude modulated with 1 kHz. The signal is detected by a diode in square-law mode, which at its output provides a 1-kHz signal which is used to determine the attenuation. The RF attenuation is half the change in attenuation at low-frequency in decibels, because the diode output voltage is proportional to the RF power. The square-law region of the diode restricts the dynamic range.

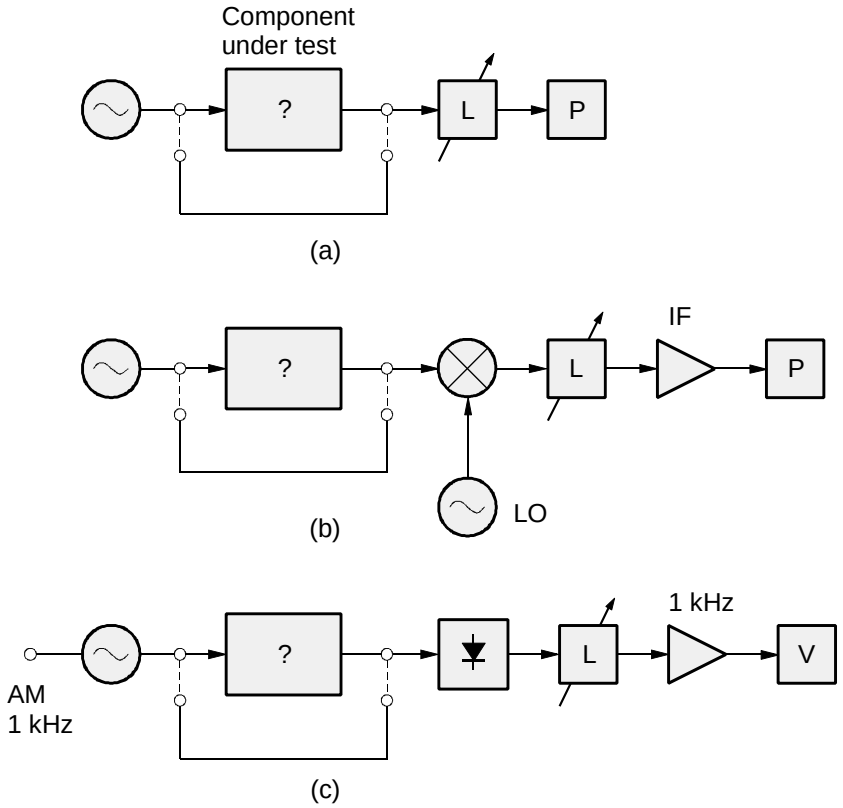

Figure 43. Attenuation or gain measurement at (a) microwave frequency, (b) an intermediate frequency, and (c) audio frequency.

As for the measurement of power, also in measuring attenuation or gain the mismatch between generator, component under test and load/sensor may be a significant error source, depending on the type of measurement method is used. Errors can often be reduced by adding on both ends of the component well-matched attenuators.

[18] *Fundamentals of RF and microwave power measurements*. Agilent Technologies, Application Note 64-1C, April 2001.

[19] *IEEE Standard Dictionary of Electrical and Electronics Terms*, Second Edition. New York 1978, The Institute of Electrical and Electronics Engineers.

[20] Räisänen A., Lehto A.: *Radio Engineering for Wireless Communication and Sensor Applications*. Norwood 2003, Artech House, Inc., pp. 111 ff..

Frequency measurements

To ensure that radio devices do not interfere with each other, they must operate at precisely defined frequencies. We can measure the frequency with electronic frequency counters, and with selective, tunable cavities. Other ways are to measure the frequency with a spectrum analyzer or to calculate it from the wavelength measured e.g. with a slotted-line.

Frequency counters

Frequency can be measured by calculating the number of signal cycles during the known duration that the gate of the frequency counter is open. For low frequencies, better accuracy is achieved by counting the frequency-counter clock pulses during one signal period (Fig. 44).

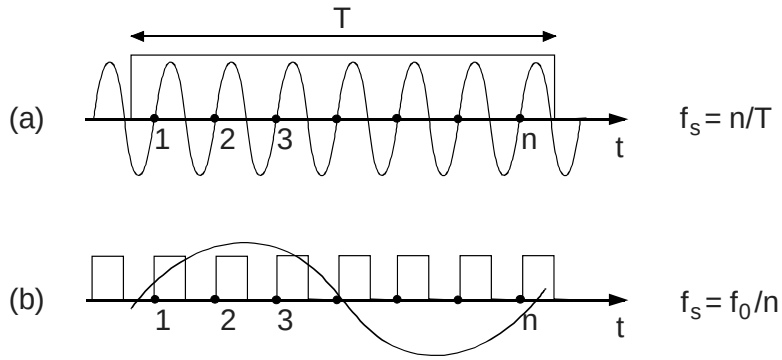

Figure 44. (a) Counting the signal cycles during one gate time *T*, (b) counting the frequency-counter clock cycles during one signal period; signal frequency is f_s and clock frequency is f_0.

Frequency counters with high-speed digital circuits can directly measure frequencies up to a few GHz. The upper limit is basically set by the speed at which digital circuits can switch between zero and one. Higher frequency can be measured by adding a divider circuit at the input (prescaling) and counting the resulting lower frequency. The division ratio *N* is typically 2–16. At millimeter-wave region, frequency counters can be realized by mixing the signal first to a lower frequency, or by phase-locking one harmonic frequency component of a low-frequency oscillator to the measured signal.

The characteristics of frequency counters are bandwidth, accuracy, measurement speed, sensitivity, dynamic range, signal-to-noise ratio (*SNR*), frequency and amplitude modulation tolerance, and amplitude discrimination.

The accuracy is affected by quantization error (only complete cycles are counted), and clock oscillator accuracy. Generally we use 1, 5 or 10 MHz crystal oscillators. The short-term stability of the oscillator is about 10^{-9}–10^{-10} 1/s, which limits the repeatability of a measurement. To stabilize the frequency the crystal is kept in an oven at a constant temperature and the supply voltage is stabilized. The absolute accuracy of frequency counter readings may be better than 10^{-7}. Aging of the crystal reduces absolute accuracy, which for a good crystal is in the range of 5×10^{-10} per day. Even better accuracies (10^{-12}) can be achieved with time standards such as rubidium or cesium clocks.

Measurement speed is affected by the signal acquisition time, which depends on the structure of the counter and which is of the order of 100 ms, and gate time, which depends on the desired resolution. The obtainable resolution is reciprocal to the gate time, e.g. with a 1-ms measurement time we can not achieve better than a 1-kHz resolution.

The typical sensitivity, i.e. the smallest signal power level at which a measurement is possible it is –20...–25 dBm; best values are about –40 dBm. A typical maximum input power is +7 dBm and dynamic range is 30–40 dB.

For a frequency counter the signal-to-noise ratio means how much a signal level must be above the noise level, so that the frequency can be measured. For frequency counters based on phase-locked loops an *SNR* of 20 dB is sufficient, those based on heterodyne converters require a signal-to-noise ratio of 40 dB.

FM and AM tolerance describes the ability of the frequency counter to detect the carrier wave frequency of a modulated signal. A typical maximum deviation of an FM-modulated signal is 10–50 MHz and the maximum modulation frequency 50 MHz. The ability to measure a signal in the presence of other lower-level signals is called amplitude discrimination.

Heterodyne frequency counter

The signal frequency can be converted down to the frequency band of a normal digital frequency counter by mixing it with the signal of a very stable local oscillator. Heterodyne-conversion-based frequency counters are available up to at least 26.5 GHz.

The block diagram of a heterodyne-conversion-based frequency counter is presented in Fig. 45. The required local oscillator signal is generated from the clock oscillator by first multiplying the frequency to f_{in} (100–500 MHz), and then feeding this to a harmonic (comb) generator. The harmonic generator, e.g. a step-recovery diode, generates a comb spectrum covering the entire frequency range of the frequency counter, with spectral-line intervals of f_{in}. One of these spectral lines Kf_{in} is selected with a filter and fed to the mixer. The filter can be a YIG filter or a filter group, where pin-diodes select the appropriate filter. At the output of the mixer we obtain the frequency difference $f_x - Kf_{in}$ (f_x is the frequency of the measured signal). This signal is fed to a digital frequency counter, and from this result the correct frequency is calculated. In practice, a microprocessor takes care of the control and of the calculations. The processor starts from $K = 1$ and step by step checks each harmonic frequency until the signal detector indicates that the signal is within that measurement band. Only then, the actual measurement is started. The processor can also first step through the whole frequency band, save the output voltages of the signal detector in the memory, and then select the strongest signal for measurement.

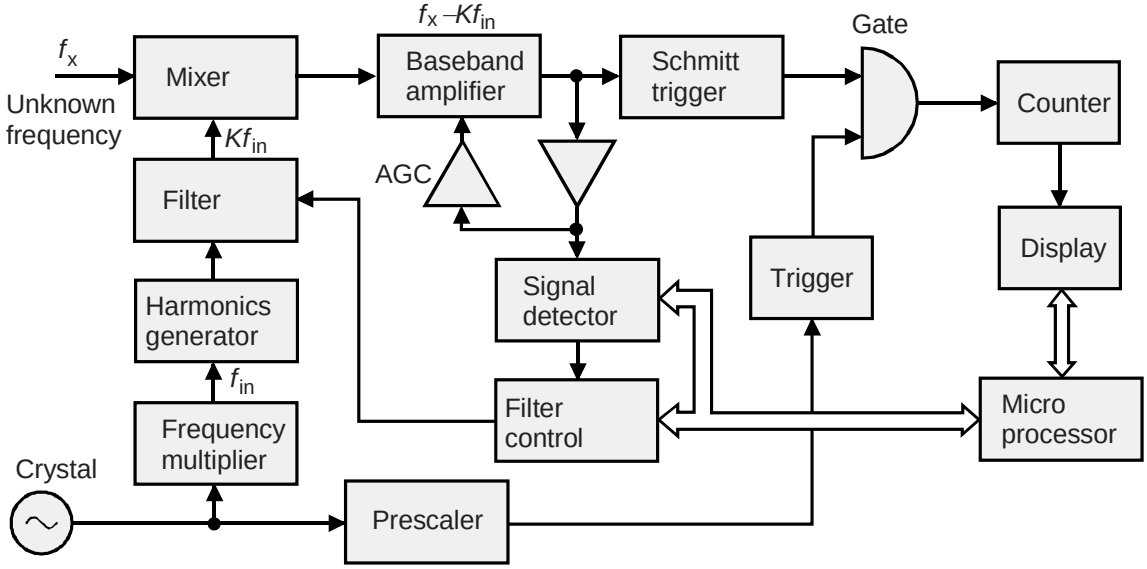

Figure 45. Frequency counter based on heterodyne conversion.

Phase-locked-loop frequency counters

A low-frequency oscillator can be phase locked to the high-frequency signal to be measured. This low-frequency signal is measured by a standard frequency counter, and the final result is obtained by multiplying with the ratio N or by multiplying the gate time by N.

The principle of a phase-locked-loop(PLL)-based frequency counter is presented in Fig. 46. The frequency of a voltage controlled oscillator (VCO) f_0 is multiplied with a harmonic generator to obtain nf_0 (n = 1, 2, 3,...). This is fed to the mixer. If the mixer output deviates from zero, the phase-lock-loop controls the VCO voltage and thereby its frequency. When the loop is locked, $f_x = Nf_{01}$, and f_{01} is measured with the frequency counter. To determine N the VCO frequency is increased until again a zero is detected at the mixer output. Now $f_x = (N - 1) f_{02}$, and $N = f_{02}/(f_{02} - f_{01})$

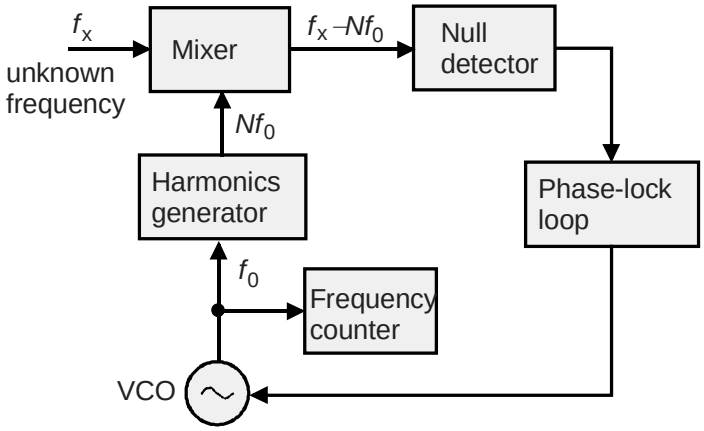

Figure 46. Operation principle of a frequency counter based on a phase-locked loop.

The disadvantage of this method is that the quantization error grows with N, or accordingly the measurement time must be N-times in order to obtain the same accuracy as with heterodyne conversion. Another disadvantage is that locking to a frequency-modulated signal can be quite difficult. PLL-based frequency counters are available up to 100 GHz. Often, they use a combination of two counter techniques described above.

Wavemeters

Cavity resonators can be used for frequency measurements, then they are called wavemeters. The operation is based on the fact that the cavity resonance is manually tuned to the frequency of the unknown signal and from the calibrated scale then reads the frequency. This frequency measurement method is very rarely used.

The cavity resonance frequency f_0 depends on the geometry and the waveform, which is excited in the cavity. To achieve good selectivity the loaded quality factor Q_L, i.e. the ratio of f_0 and half-power (3-dB) bandwidth, must be sufficiently high. Q_L is given by (see e.g. [21, 22])

$$\frac{1}{Q_L} = \frac{1}{Q_0} + \frac{1}{Q_e}$$

where Q_0 is the unloaded quality factor, and Q_e is the external quality factor. Q_0 depends on the cavity dimensions and losses. Losses are reduced by improving the conductivity of the cavity wall with silver-plating. Q_e is large if the cavity coupling to the connected waveguide is weak.

Fig. 47 shows three examples of tunable resonators. The power is coupled to the coaxial half-wave resonator through current loops (Fig. 47 (a)). The cavity length is changed by moving

the short-circuiting disk. The resonance frequency f_0 cannot be precisely calculated from the length, because the coupling loops affect the resonance. In the resonator in Fig. 47 (b) the capacitance between resonator wall and center-conductor end-capacitor is changed by moving the center conductor, and hereby tuning f_0.

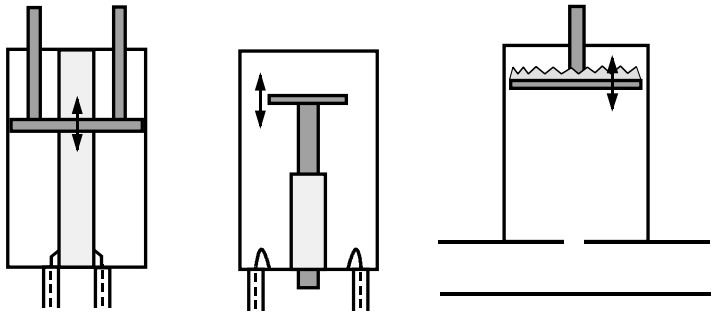

Figure 47. Adjustable resonators used as wavemeters: (a) coaxial half-wave resonator, (b) coaxial resonator with tunable end-capacitor, (c) TE_{011}-mode cylindrical cavity resonator.

In the cylindrical cavity resonator in Fig. 47 (c) the power is coupled from the waveguide to the cavity through a small hole. By moving the short-circuit f_0 changes. The resonator dimensions support the TE_{011}-wavemode. Then Q_0 is very large, up to 10^5, and there are no axial currents on the cavity wall. This facilitates the construction of the short circuit, because there is no need for galvanic contact between adjustable plate and cylinder wall. If the wavemeter was used at a wavemode with axial currents along the cavity wall, the short circuit would have to be even and lossless. Undesired wavemodes are suppressed by absorbing material behind the short circuit.

In practice, the accuracy of a wavemeter is not generally affected by the quality factor. Instead, changes in f_0 due to thermal expansion and humidity, mechanical tolerances, and scale Resolution are significant error sources. The accuracy of coaxial wavemeters may be better than 0.2% and that of cylindrical wavemeters 0.1%. Wavemeters operating only over a narrow frequency band (<10%) come closer to the accuracy limit set by the quality factor.

In Fig. 48 two ways to connect the wavemeter are illustrated. Transmission-type wavemeters allow the sample that directional couplers takes from the signal to reach the diode detector, whose current is monitored with a meter. Coaxial wavemeters can be used this way. Reflection-type wavemeters absorb at their resonance frequency part of the power propagating in the transmission line, so at the load a power drop can be measured. The wavemeter in Fig. 47(c) operates in this way and causes a power drop of ~1 dB.

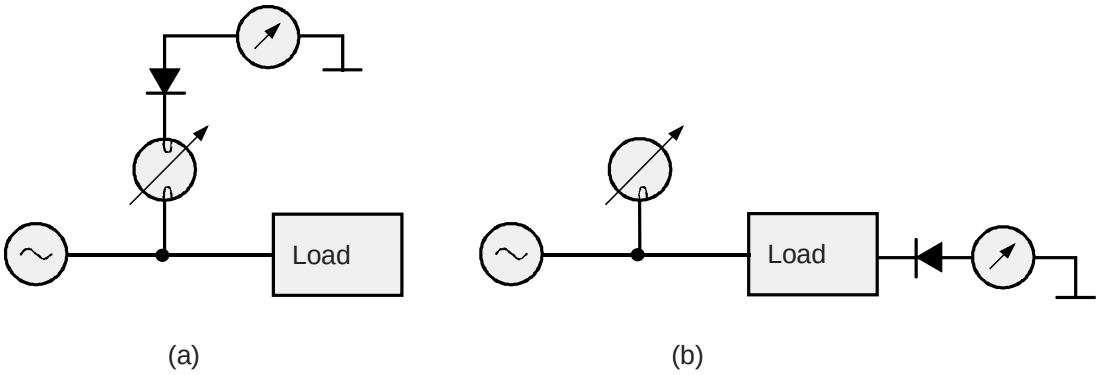

Figure 48. Wave meter Switching scheme: (a) mainstreaming, (b) the reaction principle.

[21] Collin R.E.: *Foundations for Microwave Engineering,* 2nd Edition. New York 2001, IEEE Press.
[22] Räisänen A., Lehto A.: *Radio Engineering for Wireless Communication and Sensor Applications*. Norwood 2003, Artech House, Inc..

Impedance measurements

If the impedances of individual blocks or components in a radio device are not well matched to each other, the device performance suffers. Signals are not transferred efficiently from one block to another, and the microwaves are reflected between the blocks. Therefore, impedance is an important variable, which is often measured in design phase of a circuit or device.

Let's first take a look at impedance measurements using a slotted-line, an impedance bridge, and the six-port method. Normally, however, the impedance is measured with an automatic vector network analyzer (VNA), which will be handled later.

Impedance measurement with a slotted line

The role of slotted-line measurements is nowadays very small, since this method is slow and complicated compared to the use of atomized vector network analyzers. However, one important use of the slotted-line is in education, because it can illustrate one basic concept of radio technology: the standing wave caused by a reflected wave.

The slotted line is a waveguide or coaxial cable with a narrow slot along the line. The slot is located so that it causes as small changes to the fields inside the transmission line as possible. For example, for the TE_{10}-wavemode in a rectangular waveguide (see Fig. 49) a slot along the wide sidewall of the waveguide does not interfere with the fields, since the surface currents do not flow across this center line. A movable probe inside the slot picks up a voltage proportional to the electric field. A scale indicates the exact position of the probe along the slot. The induced voltage is detected by a square-law diode, amplified and transmitted to the meter.

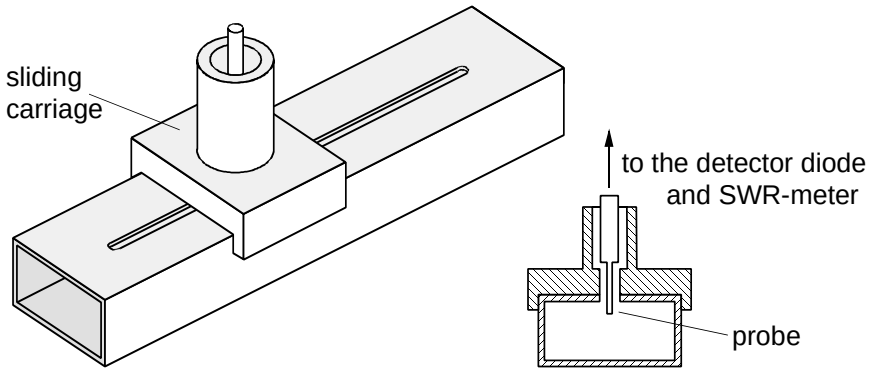

Figure 49. Slotted-line setup made from a rectangular waveguide.

If the slotted line is terminated at location $z = 0$ with a load whose reflection coefficient is $|\rho_L|$ and phase ϕ, the absolute value of the electric field E in the transmission line is [23]

$$|E| = |E^+|\sqrt{1 + 2|\rho_L|\cos(\phi + 2\beta z) + |\rho_L|^2}\ .$$

$|E^+|$ is the amplitude of the wave travelling towards the load in positive z-direction, and $\beta = 2\pi/\lambda_g$ is the phase constant in the lossless transmission line. The output voltage of the diode operating in the square-law region V_{det} is proportional to the power coupling to the probe, i.e. to the square of the field $|E|^2$ in the slotted line. From the curve of the measured voltage as a function of the location $V_{det}(z)$ we can determine the impedance of the load. The absolute value of the reflection coefficient is calculated from the ratio of the standing wave which we get from the maximum and minimum of $V_{det}(z)$:

$$VSWR = \frac{|E_{max}|}{|E_{min}|} = \frac{1+|\rho_L|}{1-|\rho_L|} = \sqrt{\frac{V_{det,max}}{V_{det,min}}} .$$

The phase ϕ can be determined from the condition

$$\phi + 2\beta z_{min} = (1-2n)\pi ,$$

where z_{min} is the n:th minimum location when counting from the load. The phase constant is obtained from the distance between two minima $\lambda_g/2$.

Generally the scale of the slotted line is not calibrated, i.e. the load is not at location $z = 0$. Since the impedance pattern is periodical with half-wavelength period, we can determine the load impedance also at $z = m\lambda_g/2$. The locations of these reference planes can be measured when the slotted line is terminated with a short circuit. Then the absolute value of the electric field along the line is

$$|E| = 2\left|E^+ \sin(\beta z_s)\right| ,$$

where z_s is the distance from the short circuit. So, the electric field is zero at distances $m\lambda_g/2$ from the short circuit.

Slotted-line measurements can be made with an unmodulated CW signal, but in order to decrease external interferences and to increase sensitivity, an amplitude-modulated signal with 1-kHz square wave is used (see Fig. 50). Then the output of the detector contains a 1-kHz component. This voltage is then fed to a special VSWR-meter. At the input of this meter is a narrow-band filter tuned to 1 kHz. The pre-amplification can be adjusted e.g. between 0–60 dB, either in 10-dB steps or continuously.

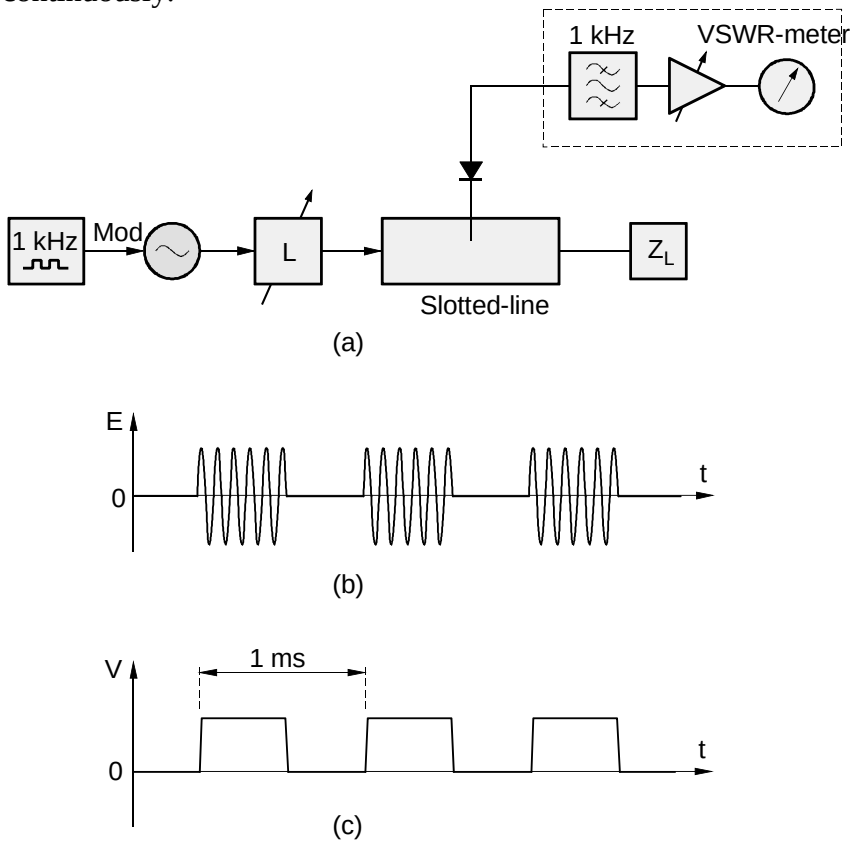

Fig. 50. Impedance measurements with a slotted-line: (a) measurement setup, (b) modulated signal, (c) diode output voltage.

Errors in the slotted-line measurement are cause by e.g. the non-ideal square law behavior of the diode, the loading of the line by the probe, the inaccurate movement of the probe, and the *VSWR* of the slotted line itself. In order to ensure square-law operation, the measurements are performed at a low enough power level. Or by using an high-precision attenuator we can eliminate the non-linearity of the diode detector: With the attenuator the detected power levels at the maximum and the minimum of the standing wave are adjusted to be identical, so that the difference in attenuation ΔL in dB gives us $VSWR = 10^{\Delta L/20}$.

The loading by the probe changes the field pattern in the line, so that the location of the minimum and the *VSWR* change. The loading can be reduced by inserting the probe as little as possible into the line, while ensuring a sufficient sensitivity.

Impedance bridges

Fig. 51 illustrates an impedance bridge for measuring complex-valued reflection coefficients/ impedances. The signal is divided with directional couplers into a reference and a test branch. The bridge is calibrated (zeroed) while a short-circuit is connected to the test port. We adjust the high-precision attenuator and phase shifter so that the meter shows zero, i.e. the signals in the two branches are of equal amplitude and opposite phase. Then the unknown impedance is connected to the test port and the meter zeroed again. From the difference in the attenuator and phase shifter readings, $\Delta L = L_{cal} - L_{meas}$ and $\Delta\phi = \phi_{cal} - \phi_{meas}$ we obtain the reflection coefficient absolute value and phase from

$$|\rho| = 10^{-\Delta L/20},$$

$$\phi = -\pi - \Delta\phi.$$

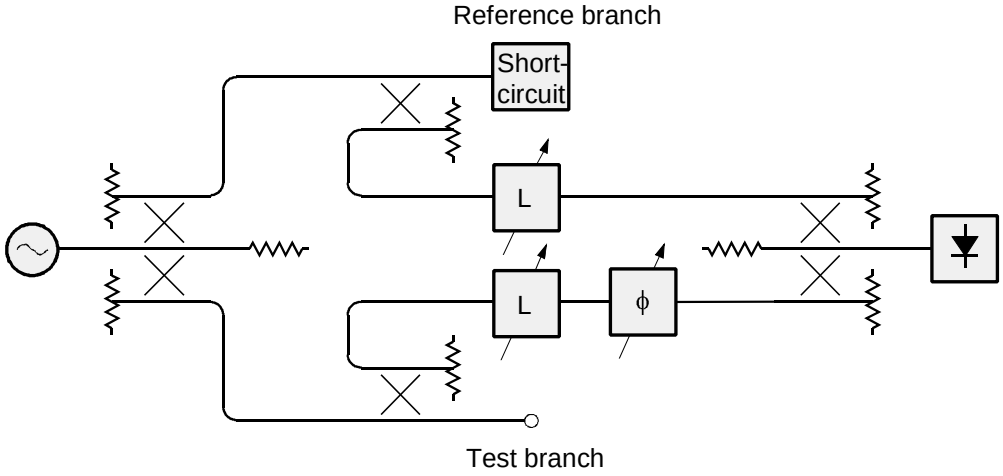

Figure 51. Impedance bridge for determining complex-valued impedances.

Six-port method

We could measure the complex reflection coefficient with the dual directional coupler shown in Fig. 26 by measuring the forward and reflected wave amplitudes and phases, a total of four quantities. Basically, the same information can be obtained by measuring four other independent quantities. This is realized in the six-port method, which measures only signal amplitudes [24].

The six-port method is basically quite simple: the generator is connected to one port and the unknown impedance to another port, by measuring the power at the four remaining ports one can determine the unknown impedance, as long as the six-port characteristics are known. Theoreti-

cally, almost any six port can be used to uniquely determine the impedance. However, finding the optimal six-port structure is difficult.

The advantages of the six-port method are the simplicity of the high-frequency part, a good accuracy, and usability up to 100 GHz and more. Phase measurements require a quite complex heterodyne receiver; now simple diode detectors or power meters are sufficient. The method does not require a very accurate frequency, so there is no need for phase-locked generators. Measurement accuracy depends on the accuracy of the calibration and on the ability to accurately measure power ratios. Weaknesses of the six-port method are the time-consuming calibration and the necessary substantial calculations needed to obtain the results [25].

In Fig. 52 we see a six-port with the generator connected to port 1 and the unknown load to port 2. The complex amplitudes of the waves propagating from and to port 2 are b_2 and a_2, respectively. Since the absolute phase is not relevant, we need three independent parameters to describe the situation at the port. It can be shown, for example, that the measured amplitude at port 3 is a linear combination of the two waves at port 2, i.e. $b_3 = Aa_2 + Bb_2$. Accordingly, the power values at all the ports are:

$$P_3 = \left|Aa_2 + Bb_2\right|^2,$$

$$P_4 = \left|Ca_2 + Db_2\right|^2,$$

$$P_5 = \left|Ea_2 + Fb_2\right|^2,$$

$$P_6 = \left|Ga_2 + Hb_2\right|^2,$$

where coefficients A ... H are complex valued constants that depend on the six-port circuit. By designing the circuit so that e.g. $C = 0$, we get $P_4 = |D|^2|b_2|^2$, i.e. the power measured at port 4 is directly proportional to the power propagating towards the load in port 2. Now we can eliminate $|b_2|^2$ from the other three equations and obtain

$$\left|\rho_L - q_3\right|^2 = \frac{\left|D\right|^2 P_3}{\left|A\right|^2 P_4},$$

$$\left|\rho_L - q_5\right|^2 = \frac{\left|D\right|^2 P_5}{\left|E\right|^2 P_4},$$

$$\left|\rho_L - q_6\right|^2 = \frac{\left|D\right|^2 P_6}{\left|G\right|^2 P_4},$$

where the reflection coefficient of the load $\rho_L = a_2/b_2$, $q_3 = -B/A$, $q_5 = -F/E$ and $q_6 = -H/G$. This set of equations has only two unknowns: amplitude and phase of the reflection coefficient. The six-port characteristics can be found by calibration, i.e. by connecting several precisely known impedances to port 2. When we graphically interpret those three equations they represent circles on the complex reflection coefficient plane. The centre points q_3, q_5 and q_6 can be obtained from the six-port characteristics, and the according radii from the measured power values and the six-port characteristics. The unknown reflection coefficient ρ_L is at the intersection of the three circles, i.e. with help of the three circles we can uniquely determine ρ_L (Fig. 53). Due to uncertainties in the measurements the circles do not usually intersect at the same point. In this case statistical methods are used to determine the most probable value for ρ_L.

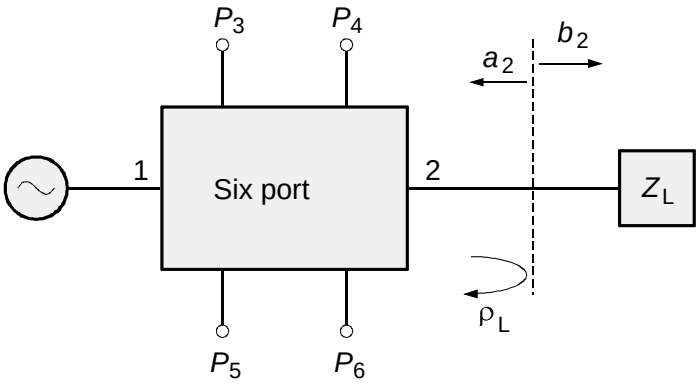

Figure 52. Six-port for impedance measurements.

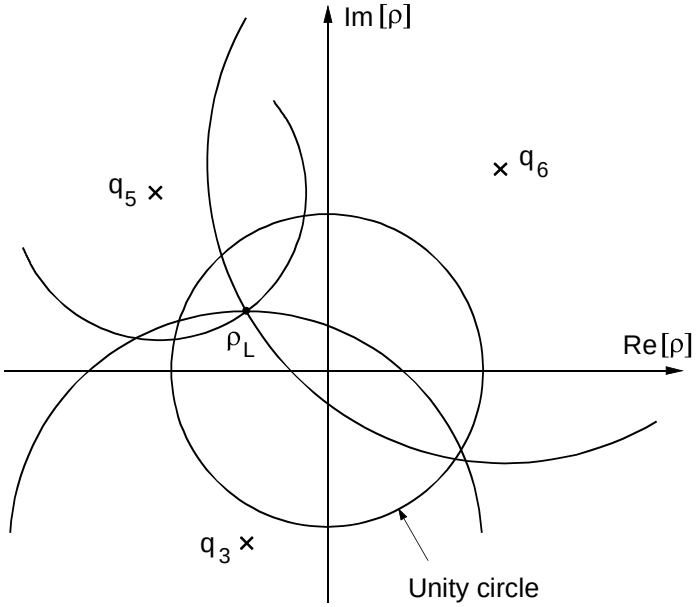

Figure 53. Determining the reflection coefficient of the load with help of three circles.

The best accuracy is achieved if the circles intersect at as large angles as possible. It has been suggested that the best placement of the center points q_3, q_5 and q_6 is on an equilateral triangle, whose center is the origin, and the absolute value $|q|$ should be between 0.5 and 1.5. In practice, a six port can be implemented e.g. with directional couplers [26]. The six-port diode detectors can also be integrated into a monolithic circuit [27, 28].

The six-port method requires measuring power at four ports, yielding three power ratios. With the four-port reflectometer illustrated in Fig. 54 we can measure only one power ratio. However, by moving the partially-reflecting termination in the fourth port of the second directional coupler, we can simulate a six-port. For three different location of the termination we obtain three power ratios, which corresponds to the measurement with a six-port. Calibration loads have to be connected to the test port for each location of the movable termination. Adding different location of the termination can potentially improve accuracy. An advantage of this method is its simple components, a disadvantage the cumbersome and slow measurement procedure.

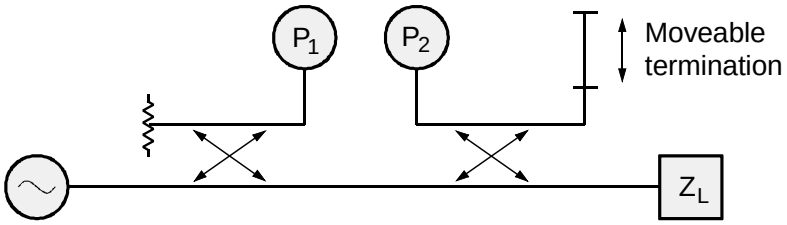

Figure 54. Four-port for impedance measurements.

[23] Räisänen A., Lehto A.: *Radio Engineering for Wireless Communication and Sensor Applications*. Norwood 2003, Artech House, Inc., pp. 69 ff.

[24] Engen G. F.: The six-port reflectometer: an alternative network analyzer. *IEEE Transactions on Microwave Theory and Techniques*, 25(1977)12, pp. 1075–1080.

[25] Engen G. F.: Calibrating the six-port reflectometer by means of sliding terminations. *IEEE Transactions on Microwave Theory and Techniques*, 26(1978)12, pp. 951–957.

[26] Engen G. F.: An improved circuit for implementing the six-port technique of microwave measurements. *IEEE Transactions on Microwave Theory and Techniques*, 25(1977)12, pp. 1080–1083.

[27] Wiedmann F., Huyart B., Bergeault E., Jallet L.: New structure for a six-port reflectometer in monolithic microwave integrated-circuit technology. *IEEE Transactions on Instrumentation and Measurement*, 46(1997)2, pp. 527–530.

[28] Wiedmann F., Huyart B., Bergeault E., Jallet L.: A new robust method for six-port reflectometer calibration. *IEEE Transactions on Instrumentation and Measurement*, 48(1999)5, pp. 927–931.

Network Analyzers

In the development work and the production of circuits it is important to be able to quickly measure circuit characteristics, such as matching, gain and scattering parameters, as a function of frequency. Automatic network analyzers are best suited for such measurements.

Network analyzer architecture

Network analyzer can be divided into scalar and vector network analyzers, depending on whether only the signal amplitude ratios are measured or also the phase differences.

In network analyzers we usually find a test unit as shown in Fig. 55, which samples the wave propagating towards the circuit, the reflected wave, and the through wave, and a receiver or analyzer unit in which these signals are processed. The test unit may have more than on port, which allows to measure enable multi-port circuits [29]. The common integration of signal generator, test unit and analyzer unit into a single unit facilitates the use of the network analyzer.

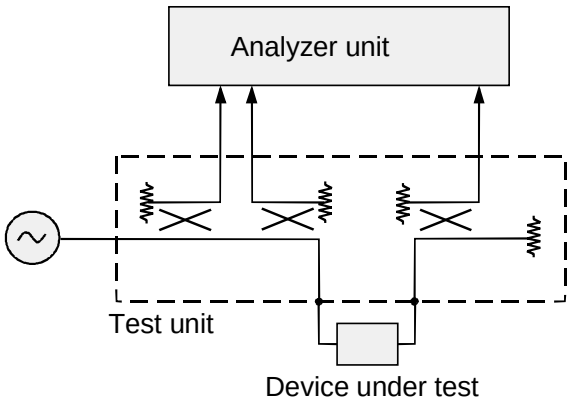

Figure 55. Simplified block diagram of a network analyzer.

In scalar network analyzers signals is usually detected by diodes, and the result is the absolute value of reflection coefficient, attenuation or gain of the test circuit. In vector network analyzers (VNA) the signals are mixed to a lower frequency, where also the phase difference is measured. This will result in impedance and scattering parameters of the circuit. The advantage of the vector network analyzer is its wide dynamic measurement range. However, for many measurements scalar network analyzers are perfectly adequate, and the diode detector is simpler and cheaper compared to the heterodyne technique.

Also a vector network analyzer can be based on the measurement of only amplitudes. In Fig. 56 a vector network analyzer is illustrated that is made up of two six-ports [30, 31]. The simplicity of amplitude-only measurements is, however, often outweighed by the complexity of calibration and calculations.

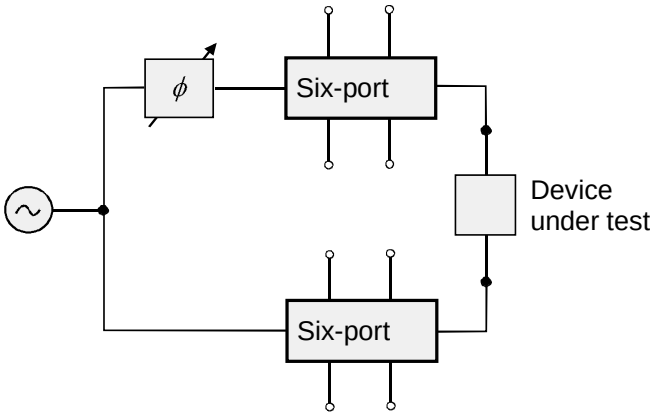

Figure 56. Vector network analyzer based on two six-ports.

Scalar network analyzers

Fig. 57 shows a scalar network analyzer, which can measure the power propagating to the circuit under test, the reflected and transmitted power, or then the according power ratios. Measuring the incoming power is not always needed because during the calibration we already obtain the power levels, to which reflected and transmitted power levels are compared. A detector measuring the incoming power can also be connected to the automatic level control of the sweep generator.

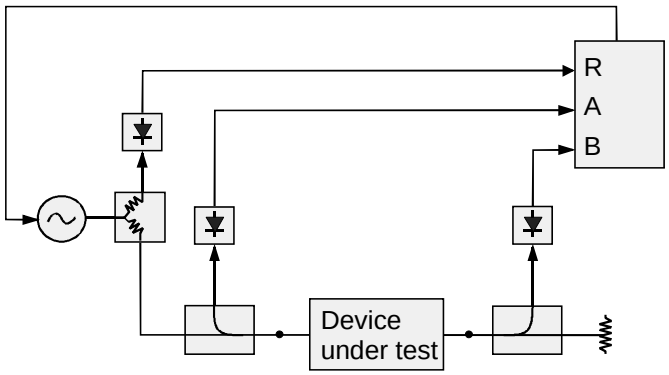

Figure 57. Scalar network analyzer.

The analyser unit controls the sweep generator. First the power is divided, a quarter of the power (–6 dB) goes to the reference channel (R) and another quarter to the tested device (half of the power is dissipated in the resistive power divider). The reflected and the transmitted signals go through resistive directional bridges to channels A and B, respectively. In the resistive bridge (see Fig. 58) the detector diode is connected between terminal 1 and 2. A quarter of the power of the incoming (forward) wave between terminals 3 and 4 is coupled to the transmission line between terminals 2 and 4, whose characteristic impedance is R_0. No forward power can couple to the diode. But a quarter of the power reflected from the load is detected by the diode. The total insertion loss of the bridge, from the generator to the diode, is thus 12 dB. Such a bridge has good matching and directivity over a very wide frequency band. The losses in power dividers and resistive bridges also reduce possible errors caused by multiple reflections.

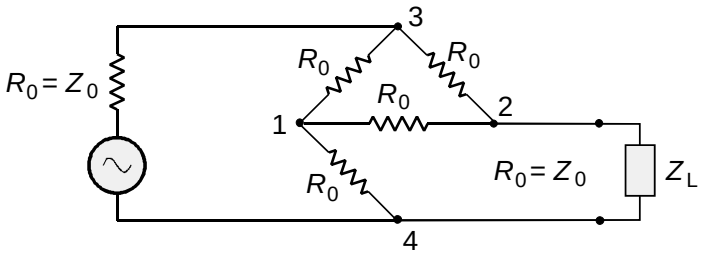

Figure 58. Directional bridge.

The output signals from the diode detectors are fed to DC amplifiers in the analyzer unit. Voltages are transformed to digital format in the earliest possible stage. With look-up-tables in memory one can then correct non-idealities of the detector, i.e. deviations from the ideal square-law and temperature dependencies. Results can also be presented in any desired format (dB, dBm, *VSWR*, etc.). The dynamic range of such a network analyser (measurement range from noise level to compression point) can be from –60 dBm to 16 dBm.

For reflection measurements we calibrate the analyzer by replacing the tested device with a short circuit or an open termination. In addition to the desired reflection the diode detects also multiple reflections caused by some mismatch in the components and connectors in the measurement setup. In addition, some of the forward power reaches the diode due to the finite directivity of the resistive bridges. The phase difference between these waves changes as a function of frequency, causing ripples in the measured power level. Since the ripples measured with a short circuit and an open terminal are roughly of opposite phase, calibration accuracy is improve by averaging these curves (see Fig. 59). If a sufficiently long high-precision cable, usually an air-insulated coaxial cable, is connected between the test port and the short circuit, the ripple as a function of frequency is so rapid that the average can be estimated by one measured curve.

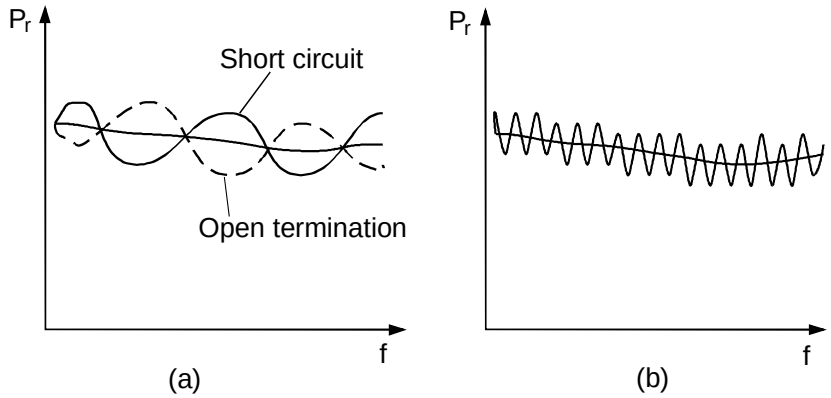

Figure 59. Calibration of reflection measurement (a) with short circuit and open termination, and (b) with a short circuit at the end of a long high-precision cable.

When measuring small reflection coefficients, the finite directivity of the resistive bridge is the most significant error source. If the directivity is known, e.g. 40 dB, we can still measure reflections whose power is lower than the power leaking to the diode due to the directivity. This is based on the fact that the leakage signal can be used as a reference signal, to which the measured reflected waves cause a ripple. The reflection coefficient is then calculated from the ripple amplitude. Sufficiently rapid ripples are obtained when using an air-insulated coaxial cable. This method is limited by the reflection from the cable, which for a good air-insulated cable is about –60 dB.

For transmission measurement the calibration is done by directly connecting the two measurement ports. Multiple reflections between the diode detector and the components and connect-

ors in the measurement setup cause calibration and measurement errors. Also here we can use air-insulated cables and the ripple-averaging method described above.

VNA architecture

Let's now have a look to the structure and operation of a typical vector network analyser. The basic components are the signal generator, the test unit and the actual analyser unit. The signal generator is a programmable, possibly synthesized, sweep generator. Measurements are performed during the sweep at e.g. 51, 101, 201 or 401 discrete frequency points.

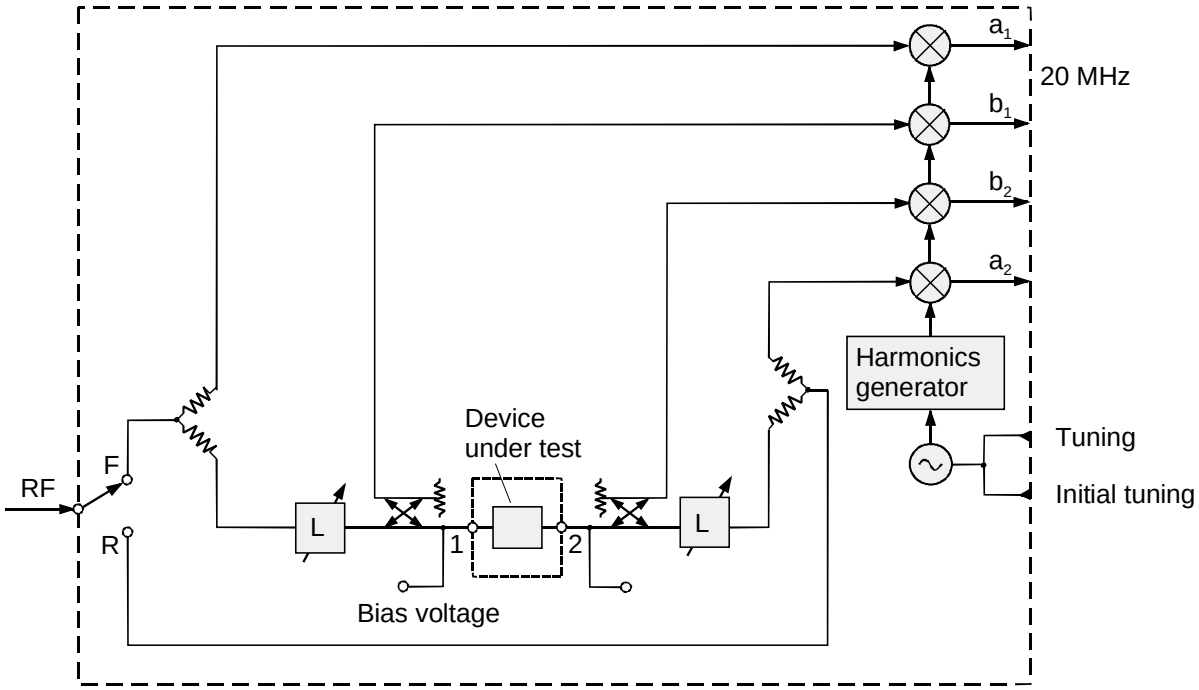

Figure 60. Test unit of a VNA.

With the test unit shown in Fig. 60 reflection and transmission measurements can be performed in both forward and reverse direction. When the input switch is in forward position (F), the signal is divided into the reference channel a_1 and through the step attenuator to the device under test. The biasing of active devices is fed through a T-circuit. The reflected signal is connected to the test channel b_1 and the transmitted signal to channel b_2. By comparing the signals at these channels a_1, b_1 and b_2 we obtain the scattering parameters S_{11} and S_{21}. When the input switch is in reverse position (R), the reference channel is a_2, the reflected signal goes to b_2, and the transmitted signal to b_1. From those we can calculate S_{22} and S_{12}.

The block diagram of the whole VNA is shown in Figure 61. The first mixing is done in the test unit and to a fixed 20-MHz frequency. The local oscillator of the mixer is a phase-locked voltage controlled oscillator (VCO), whose frequency f_{LO} is tuned between 50–300 MHz. This frequency is multiplied with a step-recovery diode. Through the bus the analyzer unit controls the signal generator so that the difference between the N-th harmonic of the local oscillator and the signal frequency at the beginning of the sweep is in the range of 20 MHz. The phase-lock circuit accurately fine-tunes the frequency by comparing the intermediate frequency to the 20-MHz crystal oscillator, and the phase-locked-loop follows the signal during the whole sweep.

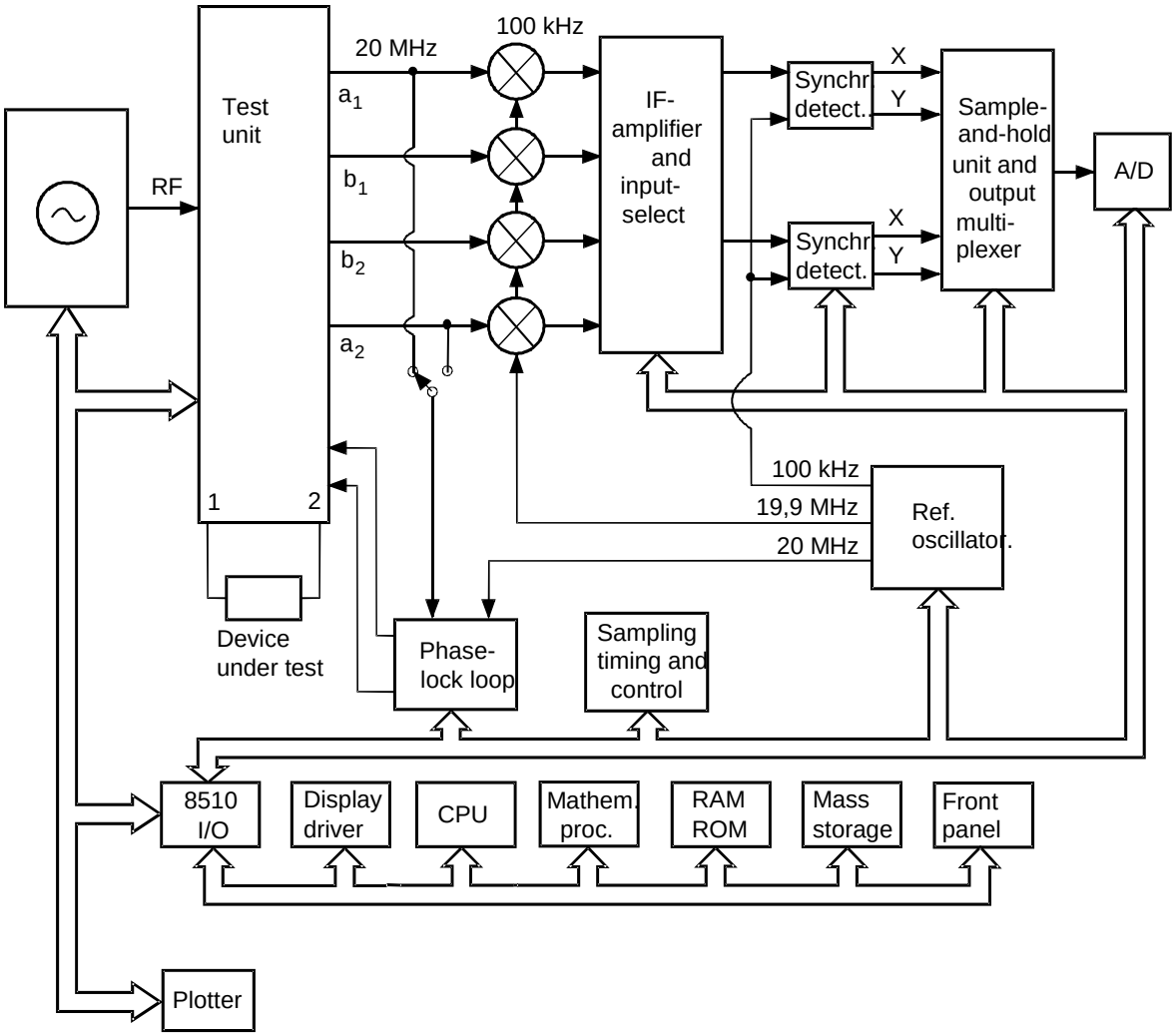

Figure 61. Block diagram of a vector network analyser (VNA).

After the second mixing stage the frequency is 100 kHz. Two synchronous detectors compare two of the channels at a time with the 100-kHz reference. Each synchronous detector consists of two mixers, both of which are connected to the measured signal (one of channels a_1...b_2) and the reference signal (see Fig. 62). The reference signal is connected to the second mixer with a 90° phase shift. If the phase difference between measured and reference signal is ϕ, the output after the low-pass filter is $X = M\cos\phi$ and $Y = M\sin\phi$, where M is proportional to the signal amplitude. X and Y are A/D-converted into numeric format and transferred to the central processing unit. The microprocessor calculates for each channel the signal amplitude $M = \sqrt{X^2 + Y^2}$ and the phase difference $\phi = \arctan(Y/X)$ relative to the reference signal. From these the S-parameters are calculated, which are then corrected with the error model presented in the next section.

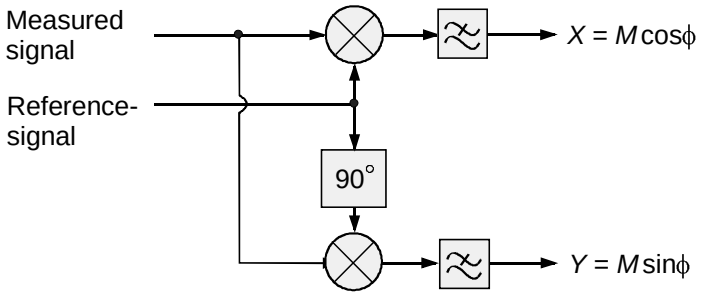

Figure 62. Synchronous detector.

The results are stored in the memory and are output in the desired format to the screen or to an external plotter. The measurement and computation time for a 401-frequency-point S-parameter measurement is less than 1 s, so the results can be obtained almost in real time, and the device under test can easily be tuned. Accuracy can be improved by averaging up to 4096 measurements. The analyzer can also calculate the time-domain response by performing an inverse Fourier transform of the frequency response.

The frequency range of the measurement system in Fig. 63 is from 45 MHz to 50 GHz, and with additional waveguide modules it can be extended up to 110 GHz. When the measured channel is e.g. 50 dB below the level of the reference channel, the uncertainty of the measured amplitude is less than ±0.1 dB and of the phase ±1.0°. The dynamic range is 80–100 dB. Commercially available are vector network analyzer operating up to over 1000 GHz.

Figure 63: HP8510 VNA with optional WR-10 waveguide measurement modules.

Measurement errors and error models

Sources of random errors in network-analyzer measurements are noise, temperature drift, imperfect repeatability of connections with RF connectors, etc. The effect of those errors is reduced by averaging a large number of measurements.

Typical sources of systematic errors are mismatch of the test-unit components, finite directivity of directional couplers or bridges, crosstalk between channels and different amplitude and phase responses of the channels.

- The combined effect of the finite directivity and the reflections occurring between directional coupler and device under test is called effective directivity. Due to effective directivity part of the forward signal couples to the channel measuring the reflection, where it is added to the reflected signal of the tested device.
- Any deviation of the test unit's output-port impedance (port 1 in Forward mode) from 50 Ω causes an error both in reflection and transmission measurements, and the error depends on the input impedance of the device under test.

– Any deviation of the test unit's input-port impedance (port 2 in Forward mode) from 50 Ω causes an error in transmission measurements and in reflection measurements of low-loss devices. The error depends on the output impedance of the device.
– Crosstalk occurs between measurement and reference channels both at the high-frequency and the intermediate frequency. This can be a significant source of error in transmission measurements of equipment with very high insertion loss.
– The frequency responses of test and reference channels differ due to different mismatch levels, lengths, and losses. This causes errors in both reflection and transmission measurements.

Systematic errors are usually the most significant source of uncertainty in the measurement results. The impact of those errors can be significantly reduced with help of an error model. Around the device under test we create an imaginary error circuit containing all errors, and the rest of the test unit is ideal. Errors can be described e.g. using a flowchart. The network analyzer is calibrated, i.e. all the terms of the error model are determined at all measurement frequencies, by replacing the device under test with precisely known loads. Thereafter, the measurement results of the device under test can be corrected with this error model [32].

Fig. 64 illustrates a 12-term error model used when measuring reflection and transmission in both directions. A 6-term error model is sufficient if the device is measured only in one direction or the device is turned around to measure S_{12} and S_{22}. To measure the reflection of a one-port only three terms are needed. The error terms in the model depend primarily on the effective directivities (E_{DF}, E_{DR}), the test port mismatch (E_{SF}, E_{SR}, E_{LF}, E_{LR}), the differences between reference and test channels (E_{RF}, E_{RR}, E_{TF}, E_{TR}) and crosstalk (E_{XF}, E_{XR}).

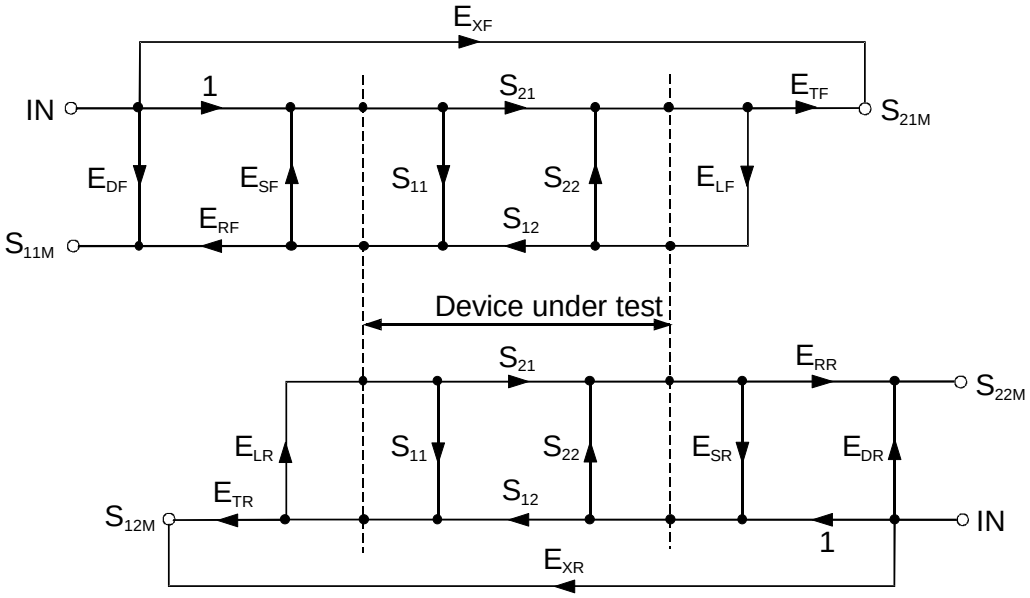

Figure 64. The 12-term error model.

The calibration terminations for reflection measurements are matched load, short circuit and open termination. They are one after another connected to each test port (or preferably to the adapters sitting on the test ports to protect these from wear and tear) and S_{11} and S_{22} are measured. The corresponding flow charts for the case of port 1 are shown in Fig. 65. The accuracy of the terminations depends on how careful they are handled, and on the accuracy of their calibration with national standards. Matched loads are either fixed broadband loads or sliding airline loads. The reflection coefficient measured with an ideal matched load is E_{DF}. In reality the small reflection δE of a non-ideal matched load adds to E_{DF}. If the matched load is moved further from the test port, the phase difference between E_{DF} and δE changes. When this measurement is re-

peated at a sufficient number (4–8) of locations, E_{DF} can be determined as the center of the obtained error circle as illustrated in Fig. 66. The accuracy of a short-circuit depends mainly on the repeatability of its mechanical connection. The small end capacitance of a coaxial open-terminal has an impact on the calibration. This capacitance can be calculated accurately and taken into account with a frequency-dependant phase angle θ as seen in Fig. 65. If the calibration is done with waveguides, the open end is replaced with a short circuit attached to a λ/4-long waveguide section.

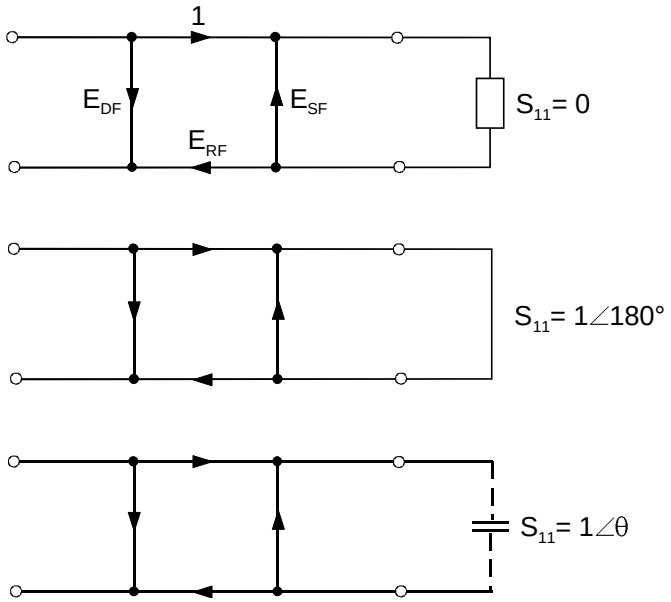

Figure 65. Calibration of a reflection measurement.

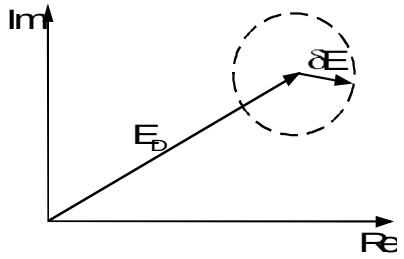

Figure 66. Correction of the reflections from a non-ideal matched load.

The measured reflection coefficient S_{11M} is obtained from the correct reflection coefficient S_{11} based on the flowchart

$$S_{11M} = E_{DF} + \frac{S_{11}E_{RF}}{1 - E_{SF}S_{11}}.$$

With help of three known loads (S_{11}) we get a system of three equations that allows solving for E_{DF}, E_{RF} and E_{SF}. The reflection coefficient of an unknown one-port we can then obtain with:

$$S_{11} = \frac{S_{11M} - E_{DF}}{E_{SF}(S_{11M} - E_{DF}) + E_{RF}}.$$

We obtain E_{XF} by directly measuring the transmission from port 2 when all ports are terminated with a matched load, see Fig. 67(a). When the ports are connected directly to each other and we measure reflection and transmission, the situation is as shown in Fig 67(b). From the

flowchart we can get expressions for S_{11M} and S_{21M}. From the obtained two equations we solve for E_{LF} and E_{TF}.

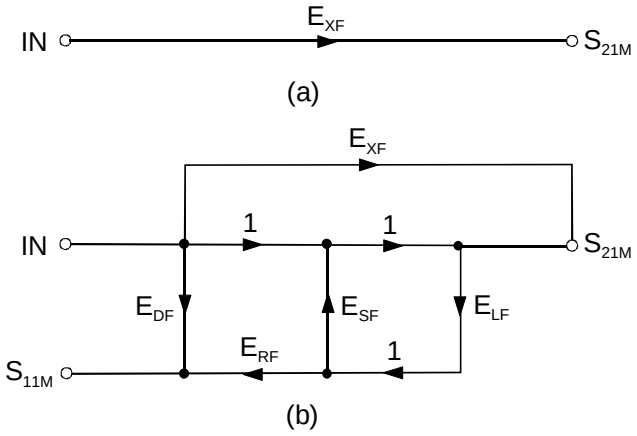

Figure 67. Flowcharts of (a) crosstalk and (b) transmission calibration measurements.

So far we have found six forward-mode error terms through six calibration measurements. By repeating the same measurements in reverse mode we obtain the missing terms for the twelve-term error model. The S-parameters of the device under test we then get from

$$S_{11} = \frac{(1+DE_{SR})A - E_{LF}BC}{(1+AE_{SF})(1+DE_{SR}) - CBE_{LR}E_{LF}},$$

$$S_{12} = \frac{[1+A(E_{SF}-E_{LR})]C}{(1+AE_{SF})(1+DE_{SR}) - CBE_{LR}E_{LF}},$$

$$S_{21} = \frac{[1+D(E_{SR}-E_{LF})]B}{(1+AE_{SF})(1+DE_{SR}) - CBE_{LR}E_{LF}},$$

$$S_{22} = \frac{(1+AE_{SF})D - E_{LR}BC}{(1+AE_{SF})(1+DE_{SR}) - CBE_{LR}E_{LF}},$$

where $A = (S_{11M} - E_{DF})/E_{RF}$, $B = (S_{21M} - E_{XF})/E_{TF}$, $C = (S_{12M} - E_{XR})/E_{TR}$ and $D = (S_{22M} - E_{DR})/E_{RR}$. So, in order to determine any one of the S-parameters of a two-port, all other *S*-parameters and all twelve error terms need to be known.

In practice, the analyzer's microprocessor manages calibration and measurement control. The analyzer prompts the user to define the measurement quantities, frequencies and connector types, then to connect the necessary calibration terminations, and finally the device under test. With help of the error model the vector analyzer readily calculates the corrected S-parameters.

Other calibration methods

The above described calibration is based on short, open and matched load as well and a through connection (SOLT-calibration, short-open-load-thru). It is well suited for coaxial systems, whereas for many other types of transmission lines the realization of those three impedance standards is difficult. E.g. on microstrip lines it is difficult to make a broadband matched load or at high frequencies a good short-circuit.

A number of other methods have been developed besides the SOLT-calibration [33–35]. They are based on simple, well-known or partially known standards. By measuring the reflection and transmission from both test ports of the network analyzer we obtain the combined scattering

matrix of the calibration standard and the surrounding imaginary two-port error circuit. When the measurement is done using three standards we obtain a system of equations, with which we can solve for the three complex terms needed to describe each error circuit.

- In the TSD calibration (thru-short-delay) the measurement reference planes are first connected directly each another, then terminated by short circuits, and finally a transmission line is connected between the reference planes (see Fig. 68). The transmission line should be reflectionless, e.g. a coaxial air line or a waveguide. An exact characterization of the line length is not needed, as long as its electrical length is not a multiple of 180°. The optimal length would be an odd multiple of 90°.
- The TRL calibration (thru-reflect-line) differs from the TSD-calibration in as much as reflective terminations are used instead of short circuits (they can also be short circuits or open terminations). Knowledge of the reflection coefficients of the terminations is not needed, but they must be identical. As the number of measurements is larger than the number of unknowns in the error model, the calibration also yields the electric length of the line and reflection coefficients of the terminations.
- In the LRL-calibration (line-reflect-line) two reflection-less transmission lines and reflective terminations (see Fig. 69) are used. For one transmission line the electrical length and attenuation must be known. The optimal case is if the electrical lengths of the two lines differ by an odd multiple of 90°. Sufficiently good accuracy can be achieved if the phase difference is 18–162°. When calibrating over a very wide frequency band, at least a third transmission line is therefore needed. As for the TRL calibration, the reflection coefficients of the terminations does not need to be known, but they must be identical.
- In the LRM-calibration (line-reflect-match), the reference planes are connected to each other, then terminated with reflective terminations (reflection coefficients unknown but identical) and finally with matched loads.

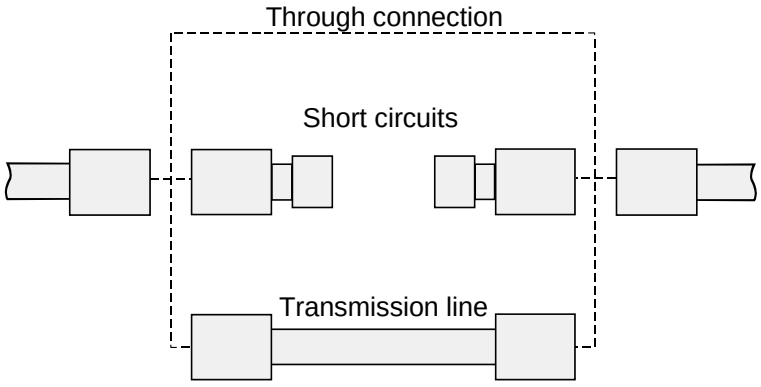

Figure 68. The TSD-calibration.

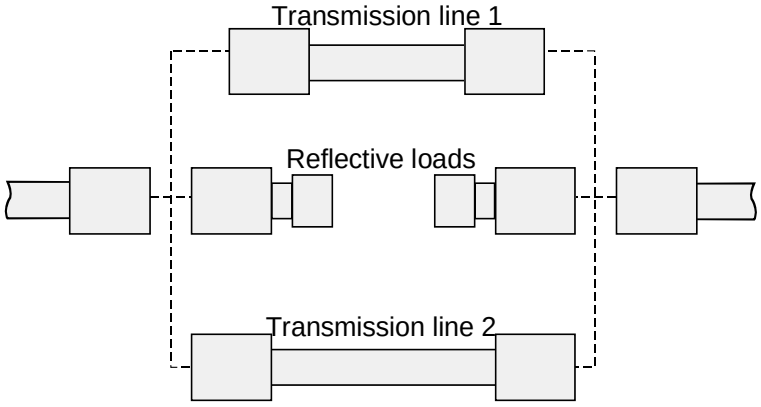

Figure 69. The LRL- calibration.

If the connectors of the device under test are of the same gender, the analyzer ports can not be directly connected for the through measurement. This problem can be solved by using two electrically identical adapters (same impedance, loss and electrical length). For example, if the tested device has female connectors and the analyzer ports male connectors, during the through calibration a female–female adapter is used, and in the measurement of the device a (redundant) male–female adapter. In the LRL-calibration these problems do not exist, as long as the calibration standards and the device under test have connectors of the same gender. If the tested device has two different types of connectors, the LRL calibration can be performed with standards for both connector types. With the first calibration measurement we get the error model on the input side of the device, and with the second calibration the error model on the output side.

Measuring microwave components

Using a network analyzer to measure the S-parameters of components such as transistors requires a special technique when the component does not have any connectors with which it could be connected to the analyzer ports. In this case, a test fixture or jig is used, in which the component is fixed. Semiconductor devices and MMIC circuits can be measured also directly on the wafer with specially designed measurement probes.

Measurements with a test fixture

When using measurement fixtures or jigs they should correspond to the typical application environment of the measured component, so e.g. a microstrip-line-based transistor is measured on a microstrip fixture. Connections of the component to the fixture must be well reproducible.

The fixture affects the measurement results: it causes mismatch, attenuation and phase shift. Mismatch is mainly due the transition from the coaxial cable or waveguide of the network analyzer to the e.g. microstrip line. The effect of the fixture on the results can be removed by de-embedding, yielding the actual S-parameters of the component [36]. For de-embedding, the error model of the circuitry between the analyzer and the component should be known. The error model is obtained from the calibration with three standards. Fig. 70 shows an example of a microstrip jig TRL calibration method. The setup can be calibrated by connecting the standards to the test fixture. Another possibility is to first calibrate the ports of the network analyzer with its own standards, and then to measure the standards connected to the jig. If the fixture consists of two identical halves, two standards are sufficient for calibration. If a good circuit model of the jig is already available, one calibration measurement is enough which is then fitted to the model.

Using the error model the actual S-parameters can be calculated in closed form. If a circuit model of the jig is available, de-embedding can be done by adding negative components to the measurement results.

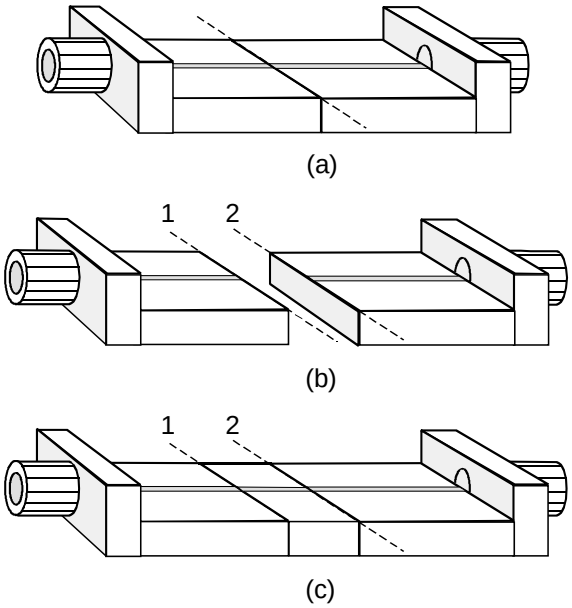

Figure 70. TRL calibration of a microstrip jig: (a) through, (b) open termination, (c) line. The component under test is placed between plane 1 and 2.

Measurements directly on the wafer

Testing a monolithic microwave integrated circuit (MMIC) on a test fixture is a long process: the circuit is cut out of the wafer, bonded with thin wires to its individual fixture, which is then connected to the test fixture, and finally the effect of the fixture is removed from the results. We can safe time and effort by measuring the characteristics of an MMIC directly to the wafer [37]. All circuits on the wafer can be tested, and damage to the circuits is avoided. Also, feedback for e.g. adjusting the manufacturing process is obtained quickly.

The measurement is performed with a vector network analyzer and coplanar probes (see Fig. 71). Such a probe has three contacts. A coplanar line can also be tapered so that there are two contact points. The probe has to fit the dimensions of contact points on the MMIC. These contact points are typically patches with a side length of about 100 μm. Since the characteristic impedance of a coplanar line depends mainly on the width of the centre line and its distance to the ground plane, a 50-Ω coplanar line can easily be tapered into very small dimensions. At the other end of the probe is a coaxial cable or waveguide. Commercially coaxial probes are available up to 60 GHz and waveguide probes up to 110 GHz. Both probe types can also be used for noise measurements of MMIC circuits.

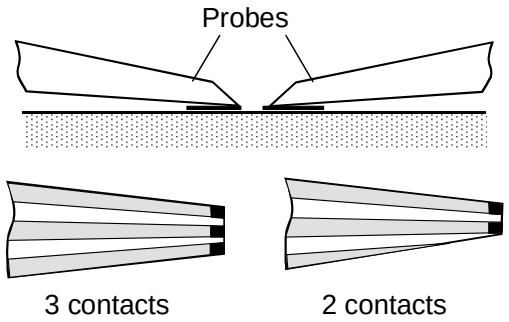

Figure 71. Co-planar probes.

The calibration of on-wafer measurements is less accurate than that of normal network analyzer measurements since the calibration standards are not as accurate. Substrates with various

calibration terminations based on thin-film technology are available. The size of the calibration loads should correspond to the dimensions of the MMIC circuits. On the same substrate can also be known loads with which a successful calibration can then be verified.

During a SOLT-calibration the probes are short-circuited, left open, terminated with 50-Ω loads, and interconnected. The open termination is realized by lifting the probe into the air. Through connection should be as short as possible, or with a cable of exactly known length, so that the calibration can be corrected for the effect of the cable length. Also the TRL and LRM calibration methods are well suited for the calibration of on-wafer measurements.

Because of radiation leakage of the probes the 16-term error model is more accurate than the 12-term model [38]. The error terms can be determined by measuring for five two-ports the reflections and transmissions in both directions (20 measurements in total). These five two-port circuits may be, e.g. through connection, matched terminations in both ports, reflective terminations in both ports, matched termination at port 1 and reflective termination at port 2, and finally matched termination at port 2 and reflective termination at port 1.

Measurement of mixers with vector network analyzers

The measurement of conversion loss or gain of mixers or other frequency translated devices cannot be done by simply connecting the device between the test ports of a normal vector network analyzer. Fig. 72 illustrates two test methods for the measurement of mixers.

In a relative measurement as shown in Fig. 72 (a) the frequency response of a mixer is compared to that of a standard mixer. A third mixer (marked *Ref.* in the figure) converts the analyzer reference signal to the intermediate frequency. The analyzer measures at the intermediate frequency the amplitude and phase differences between the mixer under test and the standard mixer. If the frequency response of the standard mixer is known, the frequency response of the mixer under test is obtained from the difference in measured frequency responses.

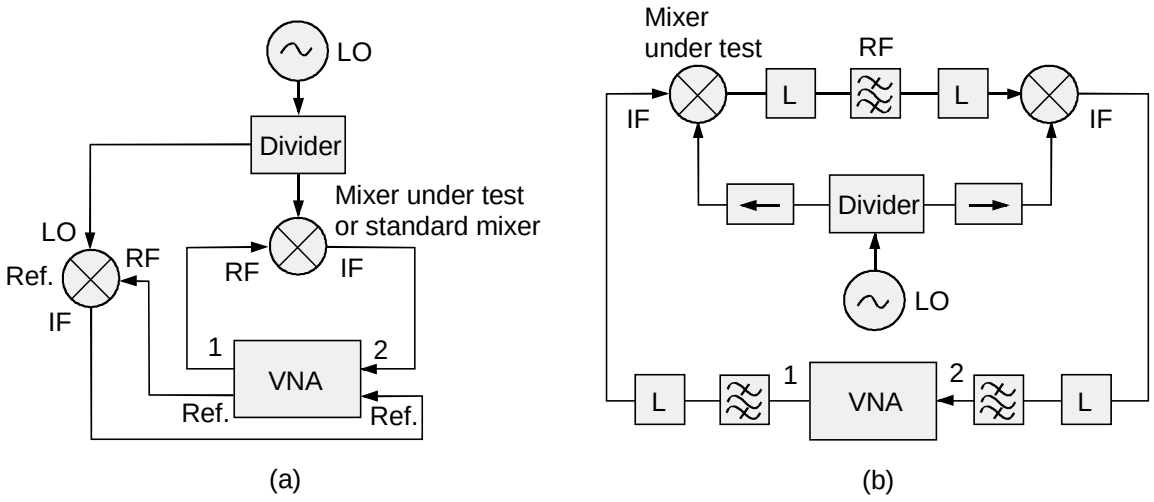

Figure 72. Measuring mixers with a vector network analyser:
(a) relative measurement, (b), absolute measurement.

An absolute measurement set-up is shown in Fig. 72 (b) [39]. It is is based on the fact that a mixer converts the RF signal to IF the same way as the IF to RF, i.e., the frequency response is the same in both directions (not all mixers operate this way). The the two mixers are connected between the ports of the network analyzer so that the second mixer converts back to the original frequency, which can be either IF or RF. By measuring three mixers in pairs (1-2, 1-3 and 2-3) we get a set of equations, from which the frequency responses of all three mixers can be ob-

tained. At least one of the mixer must show the reciprocal frequency response described above. The attenuators and filter between the mixers filter reduces errors due to reflections and additional frequency components created by the mixer.

[29] Rolfes I., Schiek B.: Multiport method for the measurement of the scattering parameters of *N*-ports. *IEEE Transactions on Microwave Theory and Techniques*, 53(2005)6, s. 1990–1996.

[30] Hoer C. A.: A network analyzer incorporating two six-port reflectometers. *IEEE Transactions on Microwave Theory and Techniques*, 25(1977)12, s. 1070–1074.

[31] Cronson H. M., Susman L.: A dual six-port automatic network analyzer. *IEEE Transactions on Microwave Theory and Techniques*, 29(1981)4, s. 372–378.

[32] Williams J.: Accuracy enhancement fundamentals for vector network analyzers. *Microwave Journal*, March 1989, s. 99–114.

[33] Maury M. A., Simpson G. R.: LRL calibration of vector automatic network analyzers. *Microwave Journal*, May 1987, s. 387–392.

[34] Lautzenhiser S., Davidson A., Jones K.: Improve accuracy of on-wafer tests via LRM calibration. *Microwaves & RF*, January 1990, s. 105–109.

[35] Silvonen K. J.: A general approach to network analyzer calibration. *IEEE Transactions on Microwave Theory and Techniques*, 40(1992)4, s. 754–759.

[36] Silvonen K. J.: Calibration of test fixtures using at least two standards. *IEEE Transactions on Microwave Theory and Techniques*, 39(1991)4, s. 624–630.

[37] Eddison I. G., Goyal R.: On wafer testing of MMICs. Luku 9 (s. 683–734) kirjassa Goyal R. (toim.): *Monolithic Microwave Integrated Circuits: Technology & Design*. Norwood, Massachusetts 1989, Artech House.

[38] Butler J. V., Rytting D. K., Iskander M. F., Pollard R. D., Bossche M. V.: 16-term error model and calibration procedure for on-wafer network analysis measurements. *IEEE Transactions on Microwave Theory and Techniques*, 39(1991)12, s. 2211– 2217.

[39] Clark C. J., Moulthrop A. A., Muha M. S., Silva C. P.: Transmission response measurements of frequency-translating devices using a vector network analyzer. *IEEE Transactions on Microwave Theory and Techniques*, 44(1996)12, s. 2724–2737.

Spectrum analyzers and vector signal analyzers

The spectrum analyzer is one of the most useful microwave measurement equipment [40]. It can be used for measuring power, frequency, and characteristics of modulated signals and noise. It can also be used to examine different non-idealities which harm the optimal use of the limited RF spectrum, such as harmonic distortion, intermodulation distortion, spread of modulated signal spectrum, and electromagnetic interference (see Fig. 73). Also the frequency response of devices can be measured with a spectrum analyzer. Naturally we obtain more accurate and reliable results with equipment specially designed for one purpose, but often also diverse measurement can be made with a good spectrum analyzer with sufficient accuracy.

The vector signal analyzer is a relatively new measurement device, suitable for many of the same measurements as the spectrum analyzer. It can, unlike a spectrum analyzer, be used to examine the characteristics also of digitally modulated signals.

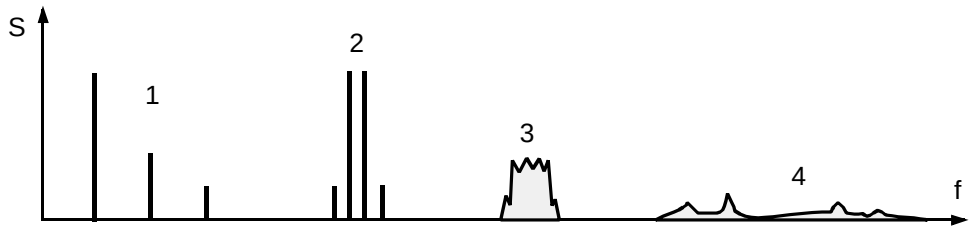

Figure 73. Typical measurements done with spectrum analyzers: 1) harmonic distortion, 2) intermodulation distortion, 3) spectrum of modulated signal, 4) electromagnetic interference.

Signal analysis

RF signals are usually characterized either in time domain or frequency domain. In Fig. 74 a signal is presented in a three-dimensional amplitude–time–frequency coordinate system. Also the signal amplitude $A(t)$ as a function of time and the amplitude $A(f)$ as a function of frequency are shown. The signal contains a fundamental sinusoidal component as well as the second harmonic frequency component.

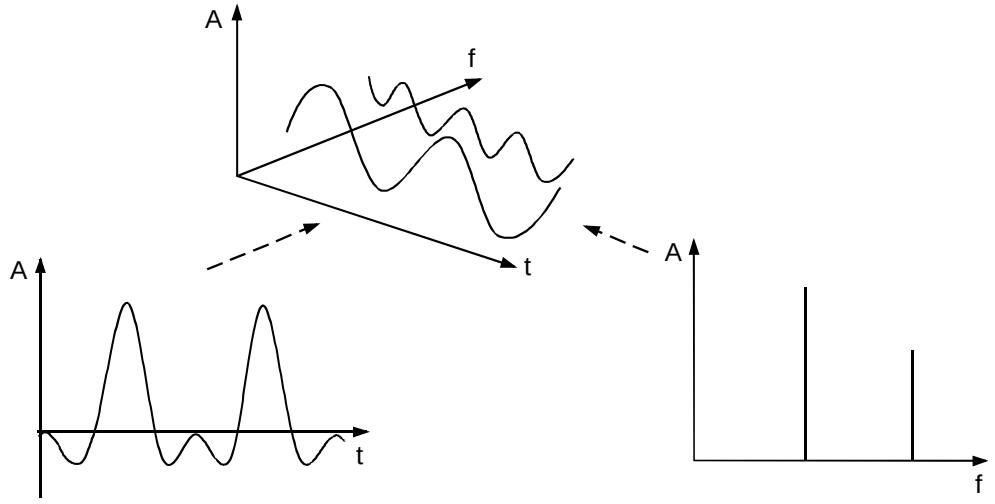

Figure 74. Three-dimensional, time-domain and frequency-domain representations of a signal.

A power meter gives an incomplete picture of such signals, because it measures the total output power and is unable to distinguish between several frequency components. A frequency counter might show incorrect readings because of the harmonic spectrum. An oscilloscope can

measure the amplitude as a function of time, but the interpretation of a complex signal in time domain is very difficult. In addition, the sensitivity of oscilloscopes is quite poor because of their wide frequency range.

The above problems are avoided when using a spectrum analyzer, or spectrometer, to analyze the signal. A spectrum analyzer is a sensitive receiver, which can be used to measure the signal amplitude as a function of frequency. However, any phase information is lost, i.e. the phase difference between the frequency components shown in Fig. 74 is not available. In other words, with the spectrum analyzer we basically measure the absolute value of the Fourier transform of the signal. The Fourier transform of function $g(t)$ is complex:

$$F(f) = \int_{-\infty}^{\infty} g(t)e^{-j2\pi ft}dt = F_{re}(f) + jF_{im}(f).$$

The absolute value of this is $|F(f)| = \sqrt{F_{re}^2(f) + F_{im}^2(f)}$. Now, in a vector signal analyzer the signal is quadrature-mixed to a lower frequency, where the phase information is preserved. Here, there are two versions of the local-oscillator signal, with a phase difference of 90°. As a mixing result we get an I- and a Q-signal from which the phase of the test signal can be calculated.

Types of spectrum analyzers

The signal spectrum can be measured in many ways. Spectrum analyzers are divided into real-time, Fourier transform-based and swept-frequency analyzers.

With real-time spectrum analyzers all frequencies in the measurement range are measured simultaneously. In this case, the analysis of a fast-changing signal is possible. Examples of real-time analyzers are filter-array, autocorrelation and acousto-optical spectrometers [41].

By dividing the signal with narrowband filters into several channels, each with an amplifier and detector, the signal spectrum can be measured (see Fig. 75). The distance between two channel centre frequencies should be in the order of the half-power bandwidth of the filters. The necessary amount of components is large, if the spectrum analyzer should operate across a wide frequency range and the resolution should be good.

An autocorrelation spectrometer calculates the signal autocorrelation as a function of delay. The Fourier transform of this function gives the power spectrum of the signal. Figure 76 illustrates a simple digital autocorrelation spectrometer. The input signal is chopped so that only the polarity changes are detected. The maximum output from channel n is obtained when the channel delay $n\Delta\tau$ equals the signal period $T = 1/f$. The frequency range of autocorrelation spectrometers is currently limited to around 1 GHz.

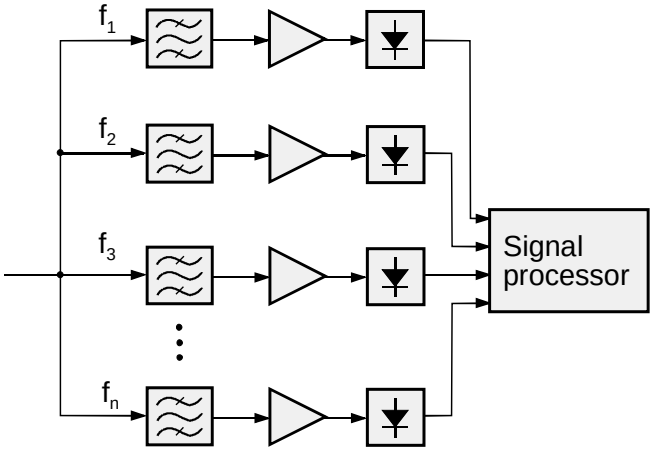

Figure 75. Spectrum analyzer based on filter group.

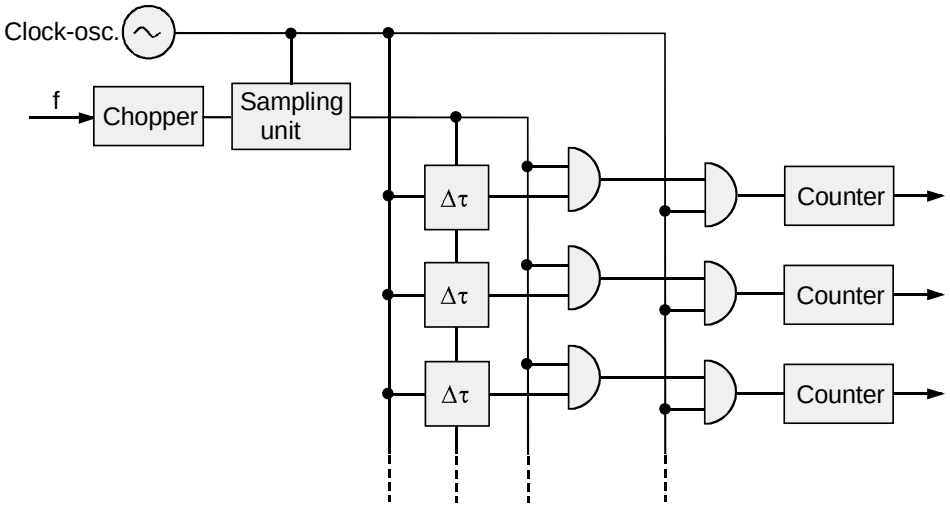

Figure 76. Autocorrelation spectrometer.

In acousto-optical spectrometers (Fig. 77) the electromagnetic wave is piezo-electrically transformed into an acoustic wave, which is fed to a so-called Bragg element. The Bragg element contains a medium, such as water or LiNbO3, in which the acoustic wave generates density fluctuations. A collimated laser beam (i.e. a beam with low divergence) gets diffracted by the lattice structure formed by these density fluctuations. The diffraction angle depends on the lattice period and thus on the frequency of the electromagnetic wave. The transmitted laser beam is refocused by lenses onto a photodiode array or a CCD (charge-coupled device), which is read by a computer. Acousto-optic spectrometers work up to over 2 GHz and are used e.g. in receivers for radio astronomy to analyze intermediate-frequency signals [42].

In Fourier or FFT analyzers, the signal is sampled for a certain period. From the samples the Fourier transform is then calculated, resulting in the signal spectrum. In the calculation an FFT algorithm (Fast Fourier Transform) is used [43], enabling a fast transformation. FFT analyzers operate up to around 1 GHz.

A sweep spectrum analyzer is basically a band-pass filter, whose center frequency is swept as a function of time [44]. It is the most widely used analyzer type used in microwave measurements, and is discussed in more detail next.

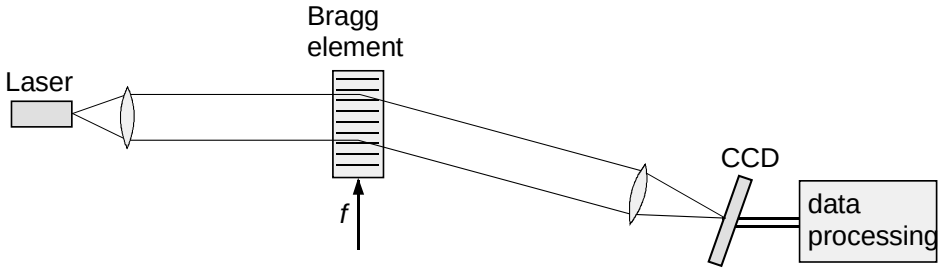

Figure 77. Acusto-optical spectrometer.

Swept-frequency spectrum analyzers

Architecture and operation

For microwave measurements beyond 1 GHz basically only sweep spectrum analyzers are available. These are superheterodyne receivers with a narrow measurement frequency band that is continuously swept over the measurement frequency range. For this reason, quickly changing signals can not be measured accurately with a sweeping spectrum analyzer.

Fig. 78 shows a simplified block diagram of a sweeping spectrum analyzer. After the input attenuator is a mixer, whose LO is a voltage controlled oscillator (VCO). The measurement frequency f_s differs from the VCO frequency f_{LO} by the intermediate frequency f_{IF}, so it will sweep at the same pace as the VCO frequency. The instantaneous measurement bandwidth is determined by the bandpass filter at the intermediate frequency. Intermediate frequency signal is detected by an envelope detector and the resulting low-frequency video signal controls the vertical deflection on the screen. If the intermediate frequency amplifier is logarithmic, the display is in decibel scale. The VCO is controlled by a ramp generator whose voltage also controls the horizontal deflection on the screen. In this way, the signal amplitude can be presented as a function of frequency. Still many additional functions are needed to make the spectrum analyzer useful.

Since the mixer is a nonlinear element, it generates from the input frequencies f_1 and f_2 the sum and difference frequencies $mf_1 \pm nf_2$, i.e. the mixing products. Spurious responses are other signals mixed to the intermediate frequency f_{IF}, which seem to come from the same frequency as the measured signal ($f_s = f_{LO} + f_{IF}$ or $f_{LO} - f_{IF}$), and therefore interfere with the measurement (see Fig. 79). Spurious responses are created by the image frequency ($f_i = f_{LO} - f_{IF}$ or $f_{LO} + f_{IF}$), signals on the harmonics sidebands of the VCO ($f_n = nf_{LO} \pm f_{IF}$, n = 2, 3, ...), and intermodulation products of two signals ($f_a = f_b \pm f_{IF}$). Also signals at the frequency f_{IF} that may leak through the mixer into the intermediate-frequency port are harmful. Such a leakage signal causes a general rise in measured signal levels at all frequencies.

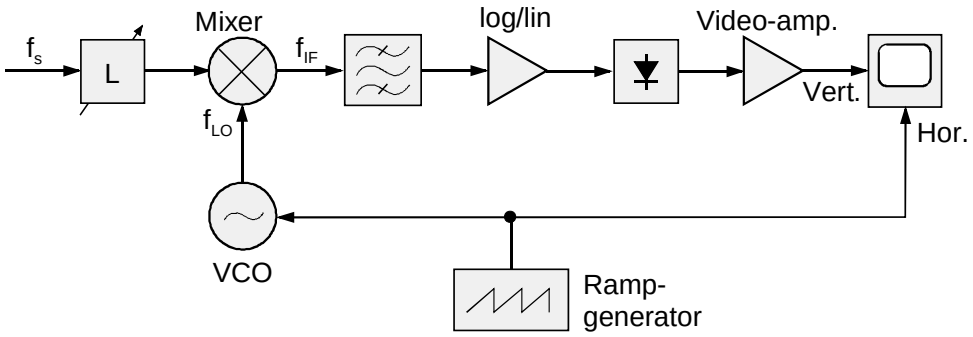

Figure 78. Sweeping spectrum analyzer.

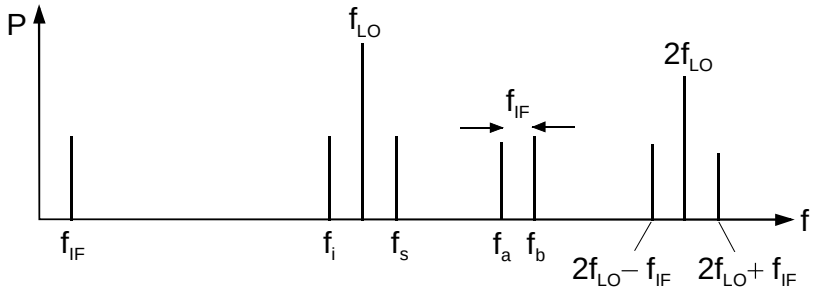

Figure 79. Spurious frequencies in the output spectrum of a spectrum analyzer.

Generation of spurious signals can be reduced through pre-filtering. The center frequency of the pre-filter is swept with the local oscillator in such a manner, that the frequencies differ by the intermediate frequency. This way only the measured signal reaches the mixer. As a pre-filter generally YIG filters are used, whose frequency can be adjusted through a DC magnetic field.

There are also ways to distinguish the measured signal from spurious signals in a spectrum analyzer by identifying the correct signal. This can be implemented, e.g., by adding a slight offset in the LO frequency between successive sweeps. The spurious frequencies are then offset in the opposite direction than the measured signal, or by the wrong frequency offset.

In practice, more than one mixing stage are used - the signal is mixed twice, thrice or even more often, to lower and lower intermediate frequencies. This way good selectivity is achieved. The center frequency of the sweep along with its width and speed, as well as bandwidths of the intermediate-frequency and video filter should be adjustable, to ensure that the analyzer can be used for characterizing a large variety of signals. The bandwidth can automatically be adjusted with the chosen sweeping speed.

In modern spectrum analyzers the output voltage of the diode detector is digitally stored in memory. In this case also very slow sweeping is possible and processing of measurement data easy. One can display, e.g. the difference of two sweeps, the average or peak value of multiple sweeps, or simply several measurements.

The display of a spectrum analyzer is usually similar to Fig. 83: the screen is divided by thin horizontal lines into 10 parts and vertically into 8 or 10 parts. The horizontal frequency range is selected by setting the sweep center frequency and width accordingly. The vertical scale is set to either logarithmic or linear scale, and the vertical reference level (e.g. –20 dBm, 0 dBm or 10 dBm) and sensitivity (e.g. 1 or 10 dB/division) are set according to the measurement situation. The analyzer can automatically seek and mark the highest displayed signal. Markers are usually also available on the screen to accurately read the signal amplitude, amplitude differences, the frequency or difference between frequencies. Often, a frequency counter is connected to the spectrum analyzer, and then the frequency can be measured precisely.

Spectrum Analyzer characteristics

Important characteristics of spectrum analyzers include frequency range, frequency accuracy, stability and resolution, amplitude accuracy, sensitivity and dynamic range. Let us have a look at those in detail.

Frequency range

The frequency range of a sweep spectrum analyzer can extend from tens of Hz to 40 GHz and with additional external, usually harmonic mixers more than 300 GHz can be reached. Since most of the microwave measurements are made with low intermediate frequencies, also low-frequency spectrum analyzers are needed. With those, obtaining a good frequency resolution can be easier and cheaper than in analyzers covering a broad frequency spectrum.

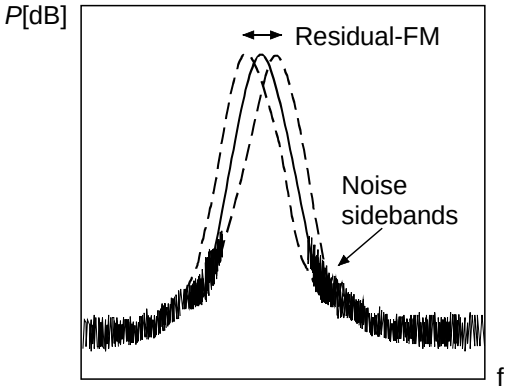

Figure 80. Residual FM and noise sidebands on the display of a spectrum analyzer.

Frequency accuracy

The absolute frequency accuracy depends on the accuracy and aging of the local oscillator as well as the width of the sweep. If the LO is a standard voltage-controlled oscillator, the frequen-

cy error may be several MHz. In synthesized local oscillators we get close to the accuracy of the crystal. Fast fluctuations in the frequency are denoted as residual-FM. Good residual-FM values are a few or a few tens of Hz. Noise sidebands are caused by phase noise in the local oscillator. They are expressed in decibels relative to the power of the carrier wave, at a certain distance from the center frequency, in a given bandwidth, e.g. −70 dBc at 50 kHz from the carrier in a 1-kHz band. Fig. 80 illustrates how residual-FM and noise affect the readings of the analyzer when measuring a pure sinusoidal signal. These instabilities reduce the ability of the analyzer to distinguish between two frequency-wise closely spaced signals.

Frequency resolution

Factors that affect the frequency selectivity, i.e. the resolution, are the stability of a local oscillator, as well as the bandwidth and shape of the intermediate-frequency band, i.e. the resolution band (Fig. 81). The resolution bandwidth can usually be chosen in increments of half a decade e.g. from 10 Hz to 1 MHz. When measuring two equally strong signals these can only be separated if their frequency difference is at least equal to the resolution bandwidth. A narrow resolution bandwidth can be used only if the residual-FM bandwidth is smaller. The selectivity factor $B_{\text{-60dB}}/B_{\text{-3dB}}$ of a filter is the ratio of the −60-dB and the −3-dB bandwidth. The smaller the selectivity factor the better the analyzer can distinguish a weak signal next to a strong one. A good selectivity factor is about 10. In such a case, e.g. with a 1-kHz filter a weak signal can be distinguished 5 kHz away from a 60 dB stronger signal, provided the weak signal is not below the noise sideband level.

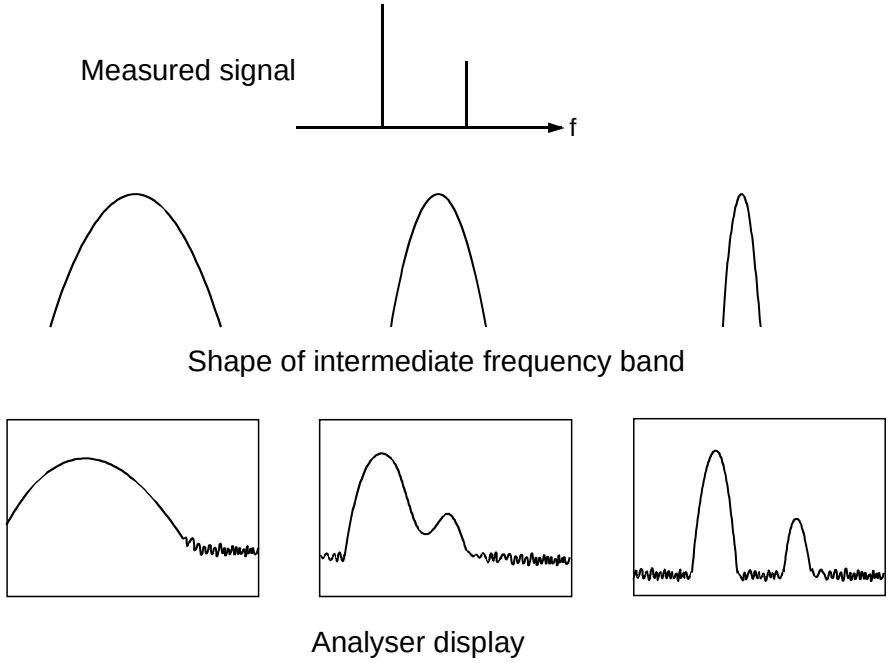

Figure 81. Influence of bandwidth of intermediate frequency on frequency selectivity.

The sweep time needed for a band of the same width as the resolution band can not be shorter than the rise time of the intermediate-frequency filter. This ensured that the filter has time to fully respond to a changing signal. Otherwise, the amplitude reading is smaller than the actual signal level, and the displayed frequency of the signal changes. The time t_s required for scanning the measurement band Δf when using resolution bandwidth B is at least

$$t_s = K \frac{\Delta f}{B^2}$$

where K is a factor that depends on the shape of the intermediate-frequency filter: For a Gaussian filter K = ≈ 3, for a rectangular filter K = 10–20. For example, the scanning time required for 1 GHz bandwidth using a Gaussian 1-kHz filter is about an hour, whereas with a 1-MHz filter it is only 3 ms. Typically, spectrum analyzers alert the user if the sweep time is too long in respect of the bandwidth, or they automatically adjust bandwidth or sweep time. The sweep span can vary from 0 Hz to the entire frequency range, and sweep time can be chosen e.g. between 50 ms – 100 s.

Amplitude accuracy

The amplitude accuracy depends on the frequency response of the whole system (input attenuator, intermediate-frequency amplifier, detector, etc.), on the sweep time, and on amplitude-calibration accuracy. Typical amplitude errors of good analyzers are below ± 2 dB.

Sensitivity

The sensitivity of a spectrum analyzer, i.e. the smallest detectable signal level, depends on the attenuation L_{in} before the mixer (input attenuator and YIG filter), the noise temperature of the mixer T_{M}, mixing loss L_{M}, the noise temperature of the intermediate-frequency part T_{IF}, and the noise bandwidth B_{n} of the device (see next paragraph for definition of B_{n}), which is approximately equal to the resolution bandwidth. The power reduced to the input P_{in} that corresponds to a signal–to–noise ratio of one (when the noise power of the source is $kT_{in}B$) is

$$P_{in} = kL_{in}(T_{in} + T_M + L_M T_{IF})B_n,$$

where k is the Boltzmann constant and T the ambient temperature. By reducing the resolution bandwidth by a tenth, the sensitivity is thus improved by 10 dB (but the sweep time is hundredfold!). A typical sensitivity is –120 dBm with a bandwidth of 1 kHz.

Sensitivity can be improved by using an external preamplifier. This can be significantly more sensitive than the internal mixer of the spectrum analyzer, which must operate over a very wide frequency range.

Dynamic range

The dynamic range of a spectrum analyzer can be defined in several ways (Fig. 82). The absolute upper limit of the measurement range is the power level where the mixer gets damaged, which is typically +13 dBm; when using an input attenuator the limit is about +30 dBm. The dynamic range of the display is the largest difference between two signal amplitudes whose readings can still be displayed at a specific resolution. The dynamic range of the analyzer can be defined as the difference between the 1-dB compression point and the noise level. When the signal level increases, the mixer starts to deviate from the linear response and the output saturates gradually. Between the level of about –3 dBm and –10 dBm the gain has decreased i.e. compressed by 1 dB.

Long before the mixer goes into compression, second- and third-order distortion components start to rise above the noise level: when the signal level is –40 dBm, both distortion components are approximately at a level of –120 dBm. When the signal level increases by 1 dB, second-order distortion increases by 2 dB and third-order distortion by 3 dB. The widest distortion-free dynamic range is therefore obtained when the signal level is approximately –40 dBm. At this level the power range down to the noise level can be, e.g. 70 dB. If the signal level is larger than –40 dBm, the dynamic range can be increased by reducing the power level at the mixer with an input attenuator, though the noise level is then increased.

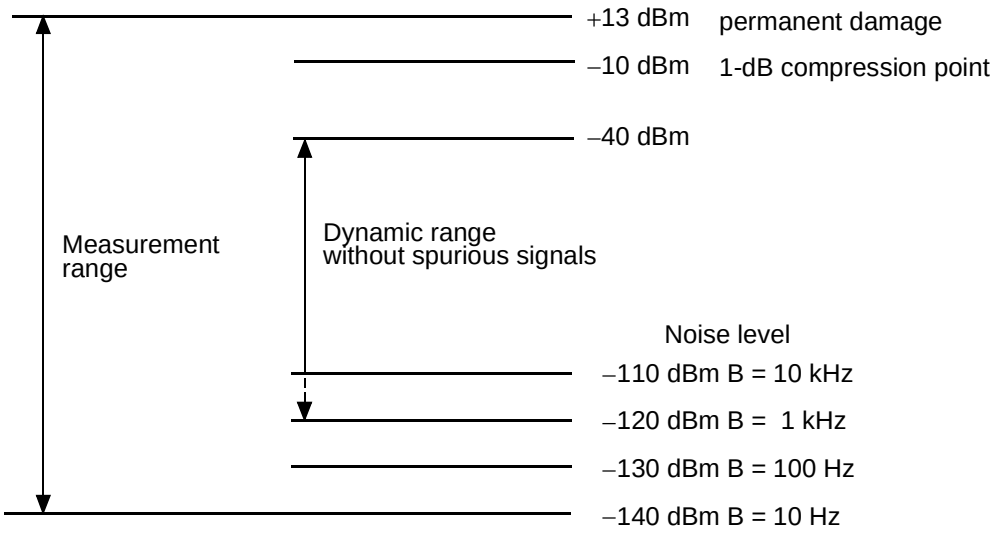

Figure 82. Typical measurement power range of a spectrum analyser.

Spectrum analyzer measurements

Let's now have a look at a few measurements, for which a spectrum analyzer is suitable. Some more specialized measurements using a spectrum analyzer will be discussed later.

Power and frequency response measurement

Spectrum analyzers are suitable for both relative and absolute power measurements. For measuring absolute power the display of the spectrum analyzer should be calibrated.

When measuring signal levels close to the noise level of a spectrum analyzer, accuracy can be improved by correcting the measurement result. It should be taken into account that the power of a sinusoidal signal and of random noise do not directly add up. The correction coefficients as a function of measured signal–to–noise ratios are presented in literature [45].

A spectrum analyzer can substitute for scalar network analyzer. With a signal generator that can accurately follow the frequency sweep of the spectrum analyzer, the frequency response can be measured with a dynamic range of more than 100 dB. The frequency tracking accuracy should be so good that the signal remains in the center of the resolution band. To reach this accuracy modern spectrum analyzers have a built-in generator (tracking generator), which keeps an offset to the LO frequency equaling the intermediate frequency. With scalar network analyzers one cannot reach this large dynamic range, because their wide-band detectors receive also the harmonic frequencies of the generator and noise over a wide frequency band.

For measuring the insertion loss, first the generator is connected directly to the analyzer input, which saves the frequency response in its memory. The device under test, e.g. a filter, is then connected between the generator and the analyzer and the response is measured. The difference between the two levels is the measured attenuation, and the result is presented normalized as shown in Fig. 83 (b). The return loss or absolute value of the reflection coefficient of a device is measured with a directional coupler or a directional bridge, by comparing the power reflected by the device to the power reflected by a short circuit.

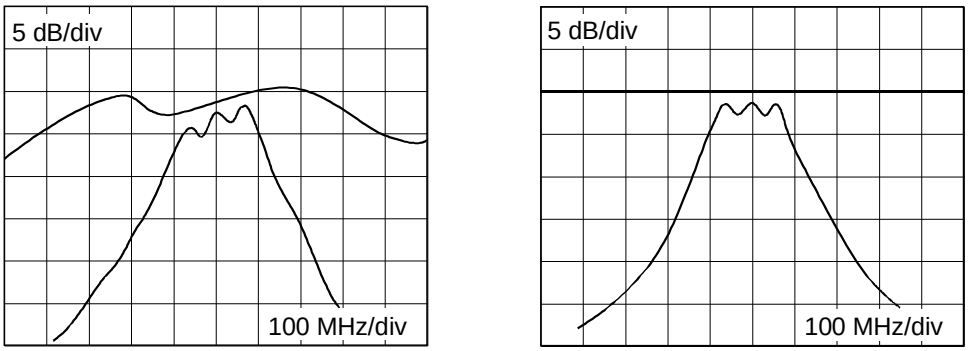

Figure 83. Filter response measured with spectrum analyzer (a) without, (b) with normalization.

Frequency measurements

Measuring frequencies fairly accurately (e.g. ±100 Hz at 10 GHz) with a spectrum analyzer has become possible thanks to improved resolution, marker frequencies, digital displays and in particular synthesized local oscillators. Spectrum analyzers can also contain a frequency counter.

A frequency can be determined as illustrated in Fig. 84 with help of an external reference signal or a calibration signal provided by the spectrum analyzer, either by tuning the marker frequency to the unknown signal, or by tuning the unknown signal into the center of the screen. Error sources in frequency measurements are the uncertainty in correctly defining the peak value of the spectral line due to the resolution bandwidth, uncertainties in displayed center frequency and division lines, and the inaccuracy of the oscillator on which the marker frequency is based.

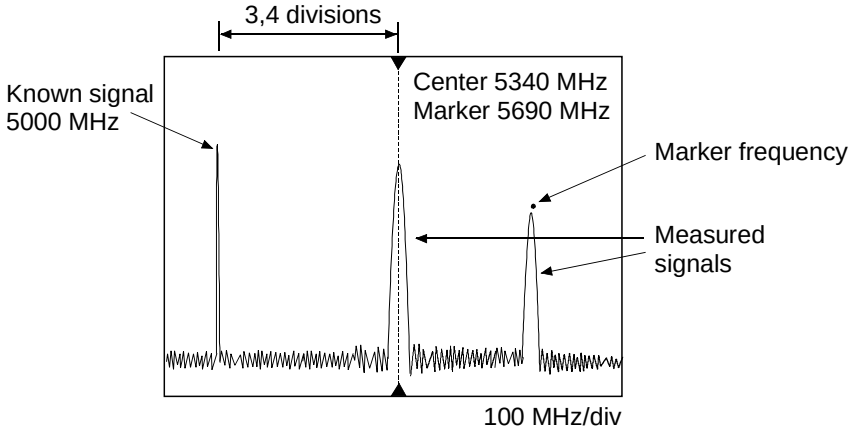

Figure 84. Measuring frequency with help of a known frequency, marker and center frequency.

Noise measurements

With a spectrum analyzer we can measure the noise factor or the noise power density conveniently across a wide frequency range, as long as we do not require a particularly good accuracy. More accurate noise measurement methods are discussed in the next chapter.

The noise factor is the ratio of the signal-to-noise ratios of input and output signals of a device:

$$F = \frac{P_{Si} / P_{Ni}}{P_{So} / P_{No}} .$$

By definition, across the noise bandwidth B_n of a device the input noise power is $P_{Ni} = kT_0B_n$, where T_0 = 290 K. Since the gain of the device is $G = P_{So}/P_{Si}$, the noise factor can be written as

$$F = \frac{P_{No}}{GkT_0B_n},$$

where k is the Boltzmann constant and P_{No} the available noise power at the output. By measuring P_{No}, G and B_n the noise factor F can be calculated.

The noise bandwidth can be selected by setting the resolution bandwidth of the analyzer so narrow that it determines the measurement bandwidth. The typical noise bandwidth of the filter of a spectrum analyzer is about 1.2 times the resolution bandwidth. The gain of the device can be determined by feeding the test signal to the signal analyzer first through the device and then bypassing the device, and finally calculating the difference between the readings in decibels.

If the noise level of the device is not large enough (> 6 dB) to be above the noise level of the analyzer, a low-noise preamplifier can be added between the device and the analyzer. The gain and noise factor of the pre-amplifier must be known. Since noise powers sum up directly, the measurement results can also be corrected by subtracting the noise power of the analyzer. When measuring noise a reduction of resolution bandwidth does not improve the signal–to–noise ratio.

When measuring the noise power P_{No} one must take into account that the detector in the spectrum analyzer is an envelope detector, i.e. the output voltage is proportional to the signal peak value, and the display scale is calibrated to display the *rms* value of a sinusoidal signal. The readings are fluctuating, since peak values of noise vary with a certain probability. Therefore, a narrow-band video filter is switched on when measuring noise, so that the reading is proportional to the average noise voltage. Averaging can also be done digitally. The video filter bandwidth should not be more than 0.01 times the intermediate-frequency bandwidth in order to have a stable reading. A narrow-band noise envelope follows Rayleigh distribution, for which the ratio between *rms* and mean is $\sqrt{4/\pi}$. For this reason the *rms* of noise is 1.05 dB higher than the measured average. When using a logarithmic display the signal gets compressed by a logarithmic amplifier before the averaging. But the average of the logarithm of a signal is not the same as the average of the signal. Therefore we must add 1.45 dB to the result. The measured noise level is therefore to be corrected by 1.05 + 1.45 = 2.5 dB, in order to obtain P_{No}. Some analyzers can also perform this correction automatically.

The unit of the noise power measured by the spectrum analyzer is dBm. The measurement result depends on the resolution bandwidth. If we want to know the noise-power density (dBm / Hz) the measured power in watts is divided by the noise bandwidth, after which it is changed to dBm.

AM-modulation measurement

Fig. 85 shows an amplitude-modulated signal in time and frequency domain. The carrier frequency is f_c, and the frequency of the modulating sinusoidal wave is f_m. (More complex modulating signals can be thought to consist of several sinusoidal signals.) The modulation is described by the modulation depth m. In time domain m is obtained from maximum and minimum voltage:

$$m = \frac{V_{max} - V_{min}}{V_{max} + V_{min}}$$

Modulation depth is thus between 0–1. Often m is given as percentage of modulation which is between 0–100 %.

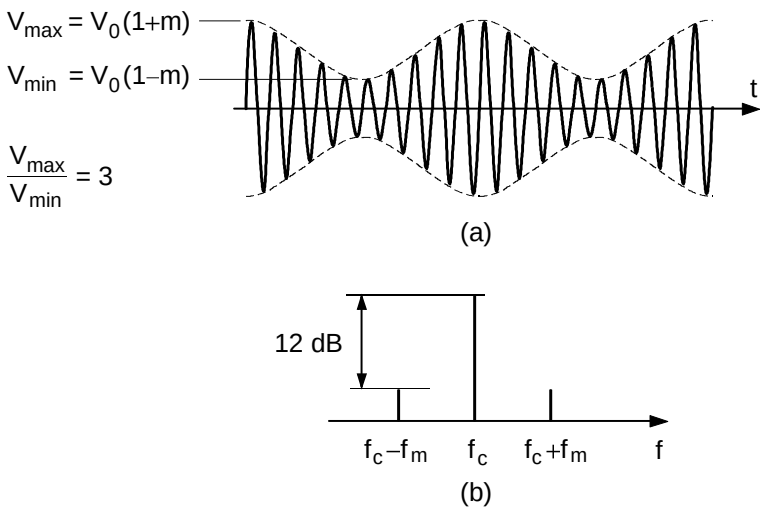

Figure 85. Amplitude-modulated signal in (a) time and (b) frequency domain, when m = 0,5.

The modulation depth of a microwave signal can be determined with an oscilloscope if the signal is first mixed to a low intermediate frequency. With a spectrum analyzer the voltage maximum and minimum values can be read from the envelope curve, which is obtained by setting the sweep width to zero and the amplitude scale to linear mode. If the modulation is small, the accuracy of this method is poor.

In the frequency domain sidebands appear on both sides of the carrier wave at a distance f_m. Modulation depth is obtained from the voltage ratio of carrier wave and sidebands:

$$m = \frac{2V_s}{V_c}$$

When m = 1, the sidebands are thus 6 dB below the carrier wave. With the typical 70-dB dynamic range of a spectrum analyzer we can measure AM modulation with a percentage of modulation of as low as 0.06% (m = 0.0006). If the modulating signal is composed of several components, they can easily be distinguished with the spectrum analyzer. If the amplitudes of the sideband are different, the signal is also frequency modulated.

FM-modulation measurement

The amplitude of a frequency-modulated signal is constant and the frequency varies in the pace of the modulating signal. If the modulating signal is sinusoidal, the signal frequency is

$$f(t) = f_c + \Delta f \sin(2\pi f_m t)$$

where f_c is the carrier-wave frequency, f_m the modulating signal frequency, and Δf the deviation, i.e. the maximum frequency offset from the carrier. In the frequency domain the signal is composed of the carrier wave and an infinite number of sidebands, with intervals f_m. The most significant sidebands are those close to the carrier.

The amplitudes of carrier and sidebands depend on the ratio of deviation and modulation frequency, i.e. the modulation index $m = \Delta f / f_m$. The amplitudes V_{sp} relative to the unmodulated carrier wave $V_{c(m=0)}$ can be obtained from the first-order Bessel functions J_p:

$$\frac{V_{sp}}{V_{c(m=0)}} = J_p(\Delta f / f_m)$$

where p is the sequential number of the sidebands, including the carrier (p = 0). In Fig. 86 the Bessel functions are shown as a function of the modulation index, and Fig. 87 is an example

of a measured spectrum, when m = 5.52. In this case, the carrier wave amplitude, i.e. J_0 is almost zero. The resolution bandwidth must be smaller than f_m in order to be able to distinguish the sidebands on the display of the analyzer.

If m is small (smaller than 1), the spectrum contains only one major sideband on both sides of the carrier and

$$\frac{V_{s1}}{V_c} \approx \frac{1}{2} m$$

In this case, an FM signal spectrum cannot be distinguished from an AM signal spectrum. There are various phase differences between carrier wave and FM modulation sidebands, but the spectrum analyzer displays only the amplitudes.

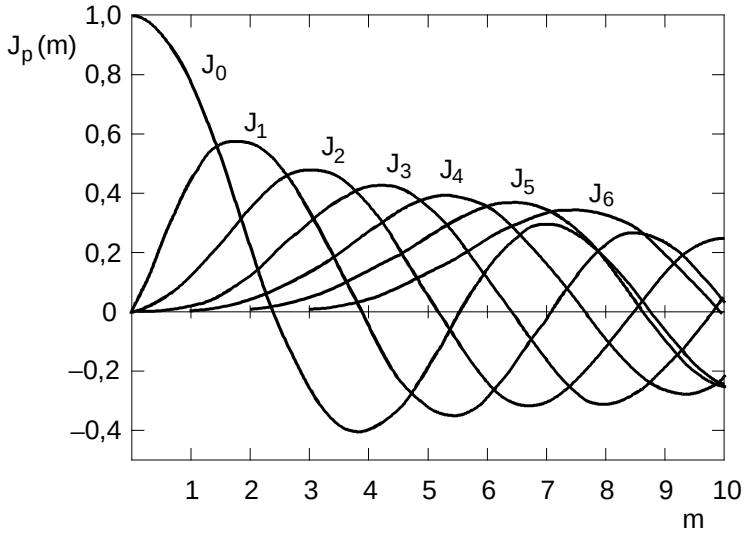

Figure 86. Bessel functions.

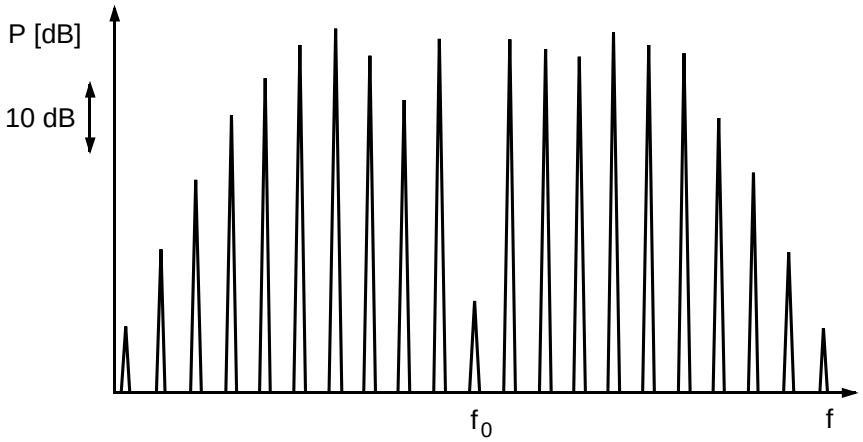

Figure 87. Spectrum of a frequency modulated signal, for m = 5,52.

A spectrum analyzer can be used e.g. to tune the FM transmitter so that the deviation Δf_1 has accurately the desired value. First we select an appropriate modulation frequency f_{m1} so that J_0 has the Nth null at the modulation index $\Delta f_1/f_{m1}$. Then we gradually increase the transmitter deviation, until we see on the analyzer display the amplitude of the carrier wave cross zero for the Nth time.

If the signal modulation can be switched off, the deviation can be calculated from the Bessel function after measuring the ratio of the carrier-wave amplitudes of the modulated signal and the unmodulated signal. If the signal modulation cannot be removed, deviation is obtained by comparing the measured sideband amplitude ratios to the according Bessel function values.

The deviation can also be measured by setting the resolution bandwidth wider than Δf, whereupon the analyzer display will look similar to Fig. 88 when we select a suitable sweep band and speed. The vertical scale is set quite sensitive so that the envelope edges are steep.

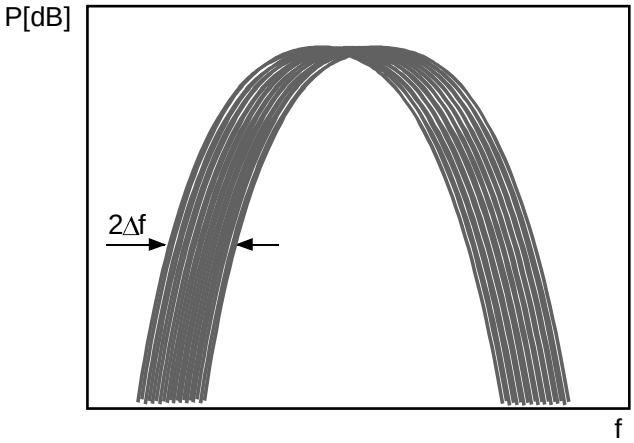

Figure 88. Measurement of the deviation of a frequency-modulated signal.

Pulsed-signal measurement

Pulsed signals are used especially in radar technology. Important characteristics of pulsed signals are carrier frequency f_c, pulse width τ, and repetition rate f_p. Peak power and average power can be measured with a spectrum analyzer. Analyzers are also used in the calibration and testing of pulsed radars.

In Fig. 89 a signal modulated with a rectangular pulse is presented in time and frequency domain. The spectrum consists of an infinite number of frequencies $f_c \pm nf_p$ (n = 0, 1, 2, ...), with intervals equal to the pulse repetition rate. The spectral envelope can be written in the form $\sin(\pi f\tau)/(\pi f\tau)$, of which the analyzer displays the absolute value. The carrier frequency is found from the central spectral line, and the pulse width is found with help of the nulls in the envelope. For an ideal pulse-modulated signal the first sidelobe maximum of the envelope is 13.3 dB below the level of the carrier wave. If the pulsed signal is also frequency modulated, the envelope zeros disappear and sidelobes increase. If the pulse shape is unknown, it can possibly be determined by comparing the measured spectrum with spectra of known pulse shapes.

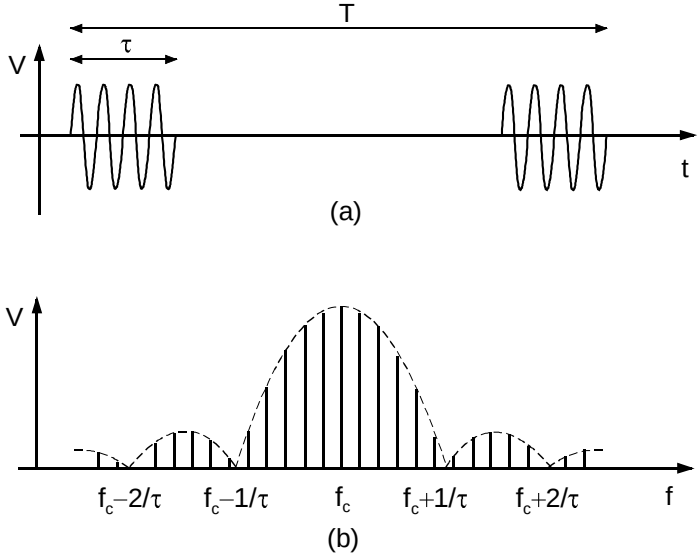

Figure 89. A pulsed signal in (a) time and (b) frequency domain.

In order to distinguish the spectral lines, the resolution bandwidth B of the spectrum analyzer must be smaller than f_p. Often one wants to display the envelope of the spectrum, then B must be set larger than f_p. Details of the spectrum can be distinguished well, if resolution bandwidth multiplied with pulse width $B\tau$ is about 0.1. If $B\tau$ is larger than 0.3, the spectrum shape is distorted. If $B\tau$ is on the other hand smaller than 0.1 no additional spectral information is obtained, but sensitivity of the measurement is reduced. Therefore the optimum $B\tau$ is around 0.1.

Harmonic distortion measurements

We speak of distortion when a device causes the waveform of a signal to deviate from the original waveform. The distorted waveform of a sinusoidal signal contains harmonic components of the fundamental frequency, i.e. harmonic distortion is generated. Harmonic distortion is generated especially when the signal power is so high that the non-linearity of a circuit affects the signal. It is important to minimize distortion of amplifiers, mixers and many other circuits, since this ensures good performance of the transmitters and receivers.

In Fig. 90 input and output signals of a device is shown in time and frequency domains. The total distortion of the device

$$THD = \frac{\sqrt{V_2^2 + V_3^2 + ... + V_n^2}}{V_1}$$

where V_1 is the voltage amplitude of the fundamental frequency at the output, and V_n (n = 2, 3, ...) is the nth harmonic frequency components. Often the distortion is expressed as a percentage, i.e. the above *THD* multiplied by 100%.

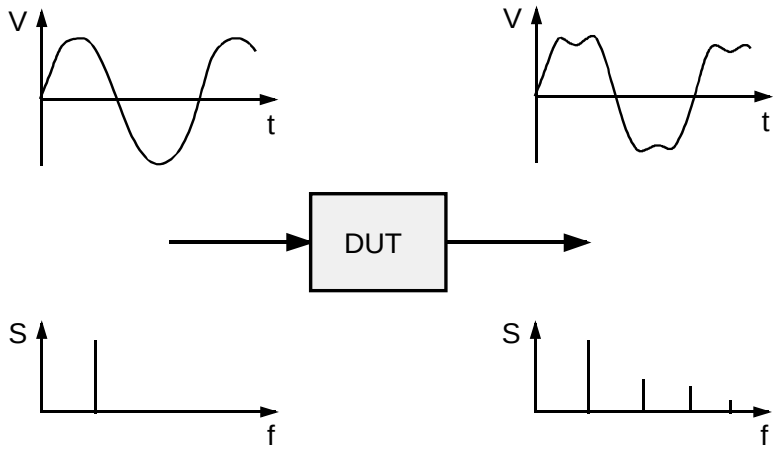

Figure 90. Sinusoidal input signal and distorted output signal in time and frequency domain.

When measuring distortion it must be taken care that the spectrum analyzer itself does not cause distortion. The power level at the input of the spectrum analyzer should be sufficiently small. A measurement result can be verified by adding a 10-dB attenuator at the analyzer input: the measured distortion should not change.

Vector signal analyzers

Analyzing complicated digitally modulated signals is achieved best with a vector signal analyzer [46] – with a standard spectrum analyzer many characteristics of such signals cannot be determined. For example, spread spectrum signals are close to the noise level and even similar to noise, so that a spectrum analyzer cannot detect such a signal.

Fig. 91 shows the simplified block diagram of a vector signal analyzer. Following the RF and IF parts the signal waveform is digitized with a high-speed AD converter. The sampling rate of the converter is usually a few tens of megasamples per second. The signal is then processed digitally (DSP). The signal is IQ-mixed and filtered, demodulated and presented e.g. in a constellation diagram, or the frequency spectrum is calculated with help of the fast Fourier transform (FFT).

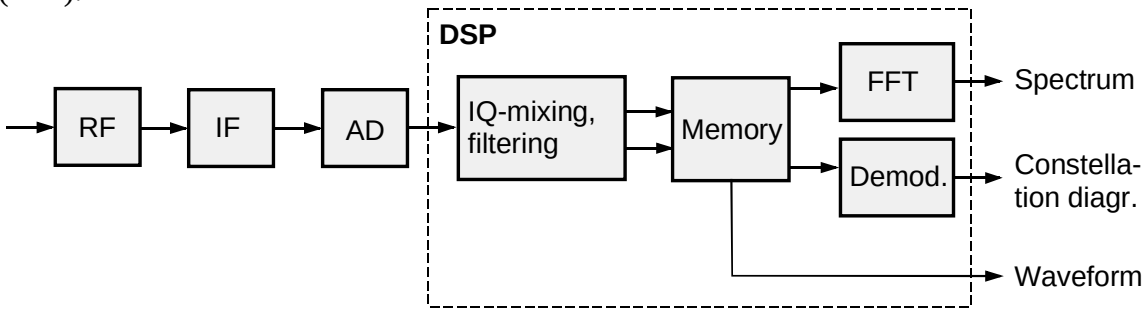

Figure 91. The block diagram of a vector signal analyzer.

Unlike a normal spectrum analyzer the vector signal analyzer does not lose the phase information, so it can measure vector quantities. It also has advantages of a real-time analyzer: A swept spectrum analyzer measures at a given moment only a narrow band, but a vector signal analyzer measures all frequency components simultaneously. We can examine also quickly varying signals, and the measurement time is short also with narrow resolution bandwidths due to the parallel post-processing.

With a vector signal analyzer e.g. FSK, PSK and QAM signals and their variations can be examined, as well as signals of different communication standards. It can also be used to measure analogue modulated signals. Many characteristics of signals can be obtained, such as demodulated I- and Q-signals at baseband, IQ-offset and IQ-imbalance, vector diagram, constellation diagram, eye diagram, errors in spectrum, frequency, amplitude or phase, and EVM.

The EVM (Error Vector Magnitude) is an important measure for the performance of digital transmitters and receivers. PSK and QAM signals can be presented is an IQ diagram as a constellation diagram. In practice the constellation points are not exactly at the ideal location because of noise, nonlinearity, or other reasons. Fig. 92 illustrates the measured phase error, amplitude error, and error vector of a constellation point. EVM is length of the error vector between measured and ideal constellation point. EVM is often given (in decibels) as the ratio of error vector *rms* and ideal signal *rms*. The analyzer calculates the EVM by comparing the modulated signal to an ideal signal. To generate the ideal reference signal the analyzer must know modulation method, symbol rate, as well as transmit and receive filter responses.

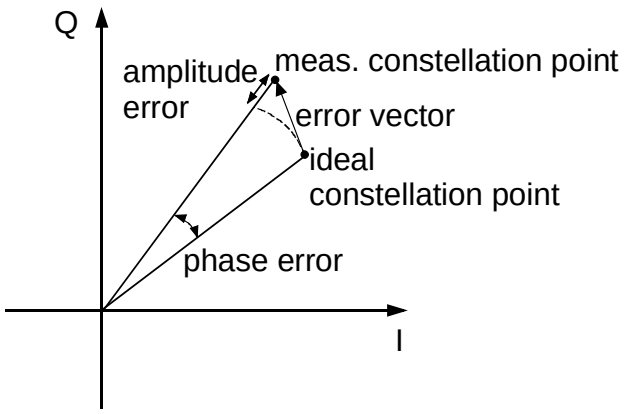

Figure 92. Constellation point, phase error, amplitude error and error vector.

With a vector signal analyzer we can study signal distortion caused by power amplifiers: the output of a vector signal generator is fed to the power amplifier and the output signal is compared to the ideal, undistorted signal. This way e.g. 1-dB compression point and AM/PM distortion can be measured.

[40] Engelson M.: *Modern Spectrum Analyzer Measurements*. Portland 1991, JMS.
[41] Tiuri M. E., Räisänen A. V.: Radio-Telescope Receivers. Chapter 7 in *Radio Astronomy* (Kraus J. D.). Powell, Ohio 1986, Cygnus-Quasar Books.
[42] Malkamäki L. J.: Actively Stabilized Acousto-Optical Spectrum Analyzer; its Design, Construction and Use for Interstellar Line Studies at Metsähovi Radio Research Station. Väitöskirja. Helsinki 1990, Yliopistopaino.
[43] Bracewell R. N.: *The Fourier Transform and Its Applications*. Singapore 1986, McGraw-Hill Book Company.
[44] Engelson M.: *Modern Spectrum Analyzer Theory and Applications*. Dedham, Massachusetts 1984, Artech House.
[45] Moulthrop A. A., Muha M. S.: Accurate measurement of signals close to the noise floor on a spectrum analyzer. *IEEE Transactions on Microwave Theory and Techniques,* 39(1991)11, s. 1882–1885.
[46] Agilent Technologies Company: *Vector Signal Analysis Basics, Application Note* 150-15. Palo Alto, CA, July 2004.

Noise measurements

The noise generated in amplifiers, mixers, and entire radio circuits is an important factor in data communications, radar technology and radio astronomy, where the noise of the receiver circuit often determines the weakest signal level that can be detected. In telecommunication systems, noise affects the signal-to-noise ratio *SNR* or the bit error rate *BER* of digital systems. Even a small reduction in receiver noise can mean significant savings, since the required transmit power is reduced or the required antenna size is reduced. With a decrease of receiver noise temperature by 50% the radar sensing distance increases by about 20%, and reduces the measurement time needed to achieve a given sensitivity in radio astronomy to one fourth.

Noise-related formulas

Noise generated in microwave and millimeter wave circuits consists mainly of thermal and shot noise. When characterizing the noise properties of a device we do not identify all the internal noise sources, but give complete device parameters, such as the noise factor and noise temperature [47-49].

The noise factor describes how the *SNR* deteriorates in a device because of its own noise:

$$F = \frac{P_{Si} / P_{Ni}}{P_{So} / P_{No}}$$

where P_{Si}/P_{Ni} is the *SNR* at the input of the device and P_{So}/P_{No} the *SNR* at the output. In this definition the input noise source is chosen to be a matched load at standard temperature T_0 = 290 K, which provides over the measurement band B the noise power $P_{Ni} = kT_0B$, where k is the Boltzmann constant ($k = 1{,}38\times10^{-23}$ J/K). The noise figure is the noise factor in decibels

$$F(\text{dB}) = 10\log\frac{P_{Si} / P_{Ni}}{P_{So} / P_{No}}\,.$$

Fig. 93 shows spectra of input and output signals of an amplifier, when a termination at standard temperature T_0 is connected to its input. The gain of the amplifier is 20 dB and the noise figure 5 dB. Both signal and noise are amplified by 20 dB. In addition, the amplifier generates its own noise, so that the noise level increases by 5 dB. Therefore, the *SNR* decreases from 30 dB to 25 dB, i.e. the difference equals noise figure F(dB) = 5 dB. If the noise level at the input is smaller than the noise generated at T_0, the *SNR* is reduced even more than F(dB). For example, in satellite receivers this is often the case, as the noise temperature of the sky is quite low.

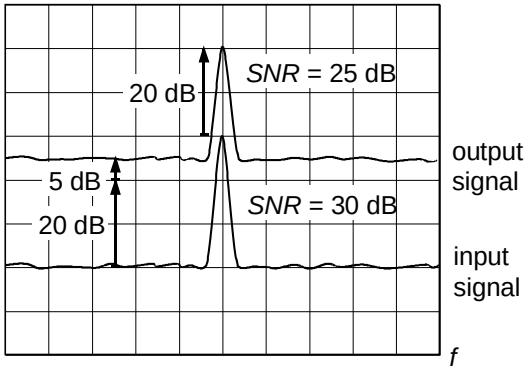

Figure 93. Input and output signals of an amplifier. Vertical scale is 10 dB/division.

Usually, when measuring the noise factor both the noise source connected to the input of the device and the noise meter at the output are matched to the reference impedance (usually 50 Ω).

The noise factor of a device depends, however, also on the input matching level. Output matching influences the gain, but not the noise factor. For linear two-ports, such as amplifiers or transistors operating at low signal levels, the noise factor is given as

$$F = F_{min} + \frac{R_n}{G_S}\left|Y_S - Y_{opt}\right|^2$$

where F_{min} is the minimum noise factor, R_n the equivalent noise resistance, $Y_S = G_S + jB_S$ is the admittance of the generator, and $Y_{opt} = G_{opt} + jB_{opt}$ the generator admittance, where the noise factor is at its minimum, i.e. $F = F_{min}$. R_n describes the speed with which the noise factor grows when the admittance of the generator moves away from the optimum. The reflection coefficients of the admittances Y_S and Y_{opt} relative to reference impedance Z_{ref} are ρ_S and ρ_{opt}. The noise factor expressed as a function of the reflection coefficients is

$$F = F_{min} + \frac{4R_n}{Z_{ref}} \frac{\left|\rho_S - \rho_{opt}\right|^2}{(1-\left|\rho_S\right|^2)\left|1+\rho_{opt}\right|^2}.$$

The noise factor of a two-port is therefore fully described at a certain frequency by the four scalar noise parameters F_{min}, R_n, $|\rho_{opt}|$ ja $\angle\rho_{opt.}$

The noise temperature of a device T_e is the temperature at which a matched termination connected to the input should be, to cause a noise power at the output of the device equal to the noise power of the device itself. The definition of noise temperature can also be interpreted so that any noise from the device is thought to be transformed into noise coming from the input termination, after which the device can be considered nominally noiseless. Noise factor and noise temperature are connected as

$$F = 1 + \frac{T_e}{T_0}.$$

Whether we use noise factor or noise temperature depends on the application, the actual noise level, and one's own habit. The formulas and calculations related to noise measurements are usually most convenient to represent as noise temperature.

The equivalent noise temperature at the input of a resistive attenuator L can be calculated from the physical temperature T, namely $T_e = (L-1)T$. If $T = T_0$, the noise factor of the attenuator is equal to the attenuation, i.e. $F = L$.

When we consider a circuit chain like in the block diagram in Fig. 94, and for each block we know the noise temperature T_{ei} and the available power gain G_i or attenuation $L_i = 1 / G_i$ (i = 1, 2 ,..., n), the equivalent noise temperature at the input of the whole chain (Friis formula) is

$$T_e = T_{e1} + \frac{T_{e2}}{G_1} + \frac{T_{e3}}{G_1 G_2} + \ldots + \frac{T_{en}}{G_1 G_2 \ldots G_{n-1}}.$$

The noise power at the output and within the frequency band B is $P_{No} = (T_{in} + T_e)kBG_1G_2\ldots G_n$, where T_{in} is the noise temperature that corresponds to the noise at the input of the first block. Using the chain formula for the noise temperature is simple if all blocks are matched to the reference impedance. In general, the values for T_{ei} and G_i of block i depend on input matching and have to be determined relative to the output impedance Y_S of the preceding block i-1.

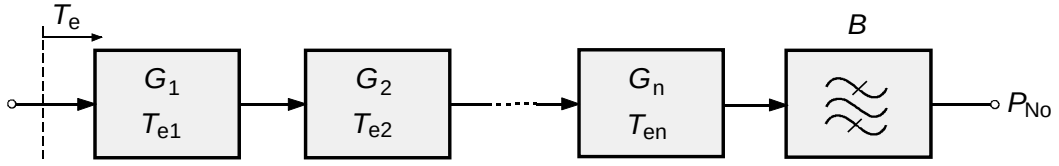

Figure 94. Cumulative noise of a chain of circuit blocks.

The noise bandwidth of a device is defined as

$$B_n = \frac{1}{G_{t,max}} \int_0^\infty G_t(f)df$$

where $G_{t,max}$ is the maximum value of the transducer power gain. A graphical interpretation of the noise bandwidth is shown in Fig. 95. For example, an ideal filter with a gain of $G_{t,max}$ within the band B_n and $G_t = 0$ outside the band, acts like a real filter when measuring white, i.e. frequency-independent noise. When calculating the output noise power P_{No} in case of Fig. 94 we need to take the noise bandwidth B_n of the filter for the bandwidth B (assuming that the response of the rest of the system is frequency independent throughout the passband of the filter). Generally, the noise bandwidth B_n of a filter is approximately equal to its –3-dB bandwidth.

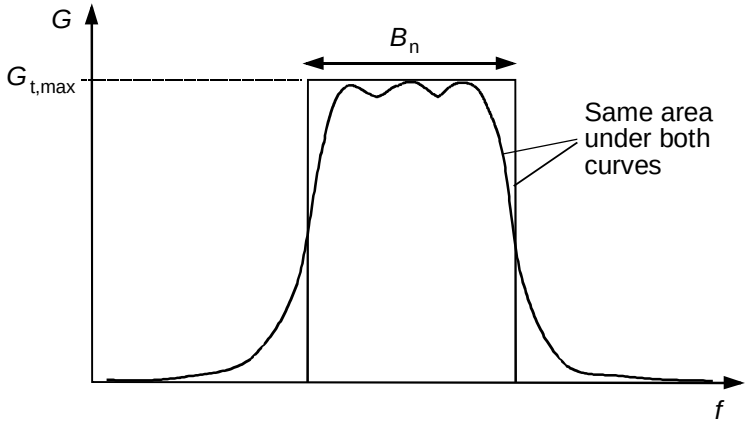

Figure 95. Definition of noise bandwidth graphically interpreted.

When connecting e.g. an amplifier in a chain, we can see from the Friis formula that the noise performance of the amplifier depends not only on the noise temperature, but also on the gain. For the comparison of amplifiers one also uses the noise measure

$$M = \frac{F-1}{1-1/G}.$$

So, when connecting two amplifiers in a row, it is worthwhile putting first the amplifier with the lower noise measure M, and not always the one with the lowest noise factor F.

The definition of noise temperature or factor of mixers is complicated by the fact that the mixer translates power to the intermediate frequency not only from the signal frequency but also from the image frequency and from the harmonic sidebands, and it depends on the application, which of these are the wanted bands. Let us have a look at a mixer whose conversion loss is L_1 for the upper sideband $f_1 = f_{LO} + f_{IF}$, L_{-1} for the lower sideband $f_{-1} = f_{LO} - f_{IF}$ and $L_{\pm n} = \infty$ for the harmonic sidebands $f_{\pm n} = nf_{LO} \pm f_{IF}$ (n = 2, 3, ...). If the wanted signal is only in one sideband the mixer operation is best described with single-sideband (SSB) noise parameters. For example, the noise temperature of the upper sideband T_{USSB} means, that when defining the equivalent noise temperature a load with this temperature is connected to the upper side band and the lower sideband is terminated with a load at 0 K, as illustrated in Fig. 96(a). For applications, where both sidebands are wanted signals, such as in remote sensing, the mixer operation is described with double-sideband DSB-parameters. T_{DSB} is obtained when terminations at this temperature are present at both sidebands, each causing as much noise power at the output as the mixer itself, as shown in Fig. 96(b). The noise temperature and noise factor for the upper sideband can be obtained from the DSB-parameters:

$$T_{USSB} = T_{DSB}\left(1+\frac{L_1}{L_{-1}}\right),$$

$$F_{USSB} = F_{DSB}\left(1+\frac{L_1}{L_{-1}}\right).$$

For the lower sideband T_{LSSB} and F_{LSSB} can be obtained by exchanging L_1 and L_{-1} in those equations. The noise from the second sideband can however cause problems when using these equations for sensitive mixers [50].

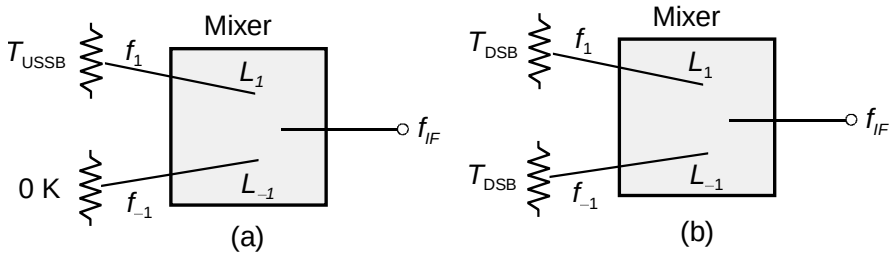

Figure 96. (a) SSB- and (b) DSB-noise temperatures of a mixer. Upper and lower sidebands are physically within the same port.

Noise sources

In microwave measurements we almost always need some kind of signal source. Usually this is a coherent signal generator. In noise measurements on the other hand a broadband frequency-independent white noise is needed. Typical noise sources are matched loads (hot and cold load), gas-discharge tubes and avalanche diodes.

Hot and cold load

A matched termination or load is the most accurate noise source. We can achieve lower than ±0.1 dB uncertainty when measuring noise parameters. The noise temperature of a matched load is directly proportional to its physical temperature, and it operates over a wide frequency range.

Matched loads are used for the calibration of other noise sources and for measuring small noise temperatures (≤ hundreds of Kelvin). Measurements of high noise temperatures are inaccurate, because the signal from the noise source is easily covered by the device's own noise. Measurements with cold and hot loads are often slow and tedious, and they are difficult to automate.

The available noise power of a matched load is $P_N = kT_B B$. The connection between the brightness temperature T_B and the physical temperature T is

$$T_B = \frac{hf/k}{e^{hf/kT}-1}$$

where h is Planck's constant (6.63×10^{-34} Js). When $hf << kT$, the brightness temperature is approximately equal to the physical temperature ($T_B \approx T$). The difference between T_B and T becomes significant at high frequencies and at low temperatures.

Constant-temperature loads are e.g. dipped in liquid nitrogen (77.3 K) or helium (4.2 K), or placed in an oven and temperature-stabilized with boiling water (373.2 K), or kept at room temperature (about 290-295 K). In noise measurements normally two loads at different temperatures

are used, which is why this method is called hot and cold load. The hot load may be quite cold. Fig. 97 shows a noise source with a hot load in boiling water and cold load in liquid nitrogen.

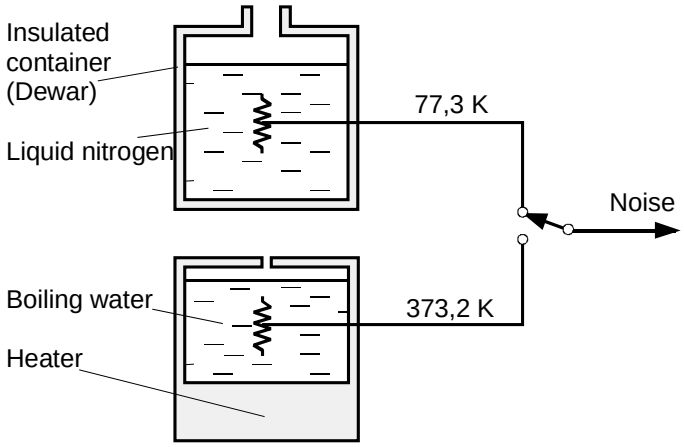

Figure 97. Noise source based on cold and hot load.

Matched loads can be of coaxial and waveguide type. Reflections from the termination must be small, since otherwise part of the noise of the device propagating towards the noise source is reflected back into the device, hereby affecting measurement accuracy.

When measuring the noise of a receiver one can place RF absorbers in front of the antenna. To reduce reflections their surface is of pyramidal shape. The absorber arrangement should be so wide that they cover the whole antenna beam; hence no thermal radiation from the environment reaches the antenna. In Fig. 98 a cold liquid-nitrogen load is placed in front of a horn antenna [51]. The non-porous polyurethane (dielectric constant close to that of air) cap forces the liquid nitrogen into the shape of the porous absorber structure. Otherwise, the interface between liquid nitrogen and air would cause significant reflections. A thin metal foil between container and antenna acts as a shield to prevent radiation from the environment to reach the antenna.

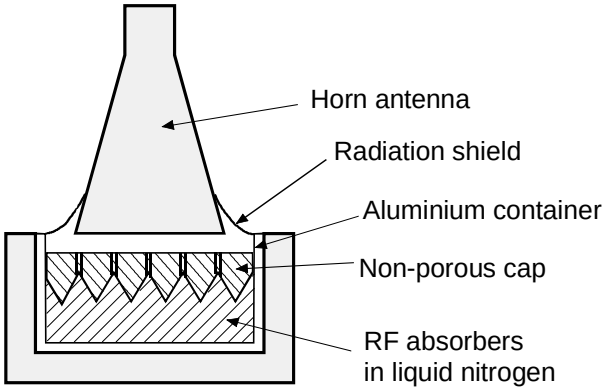

Figure 98. Antenna placed in front of a liquid-nitrogen load.

A cold or hot load cannot always be connected directly to the measured device, but there may be a piece of transmission line between the output connector of the load and the connector or flange of the device, as e.g. shown in Fig. 99. Due to the attenuation in the waveguide, the noise temperature must be corrected. If the load is at temperature T, the waveguide has temperature T_l and attenuation L, the noise temperature at the location l of the flange is

$$T' = \frac{T}{L} + \left(1 - \frac{1}{L}\right) T_l .$$

In practice, the situation is more complicated, as the temperature does not change abruptly, but it has a distribution along the waveguide. The attenuation α of the transmission line depends on temperature. If the temperature distribution $T_l(z)$ and also the distribution of $\alpha(z)$ are known

$$T' = \frac{T}{L} + \frac{2}{L}\int_0^l \alpha(z) T_l(z) L_z dz\,.$$

where L is the total attenuation and L_z is the attenuation between $0 - z$ [52].

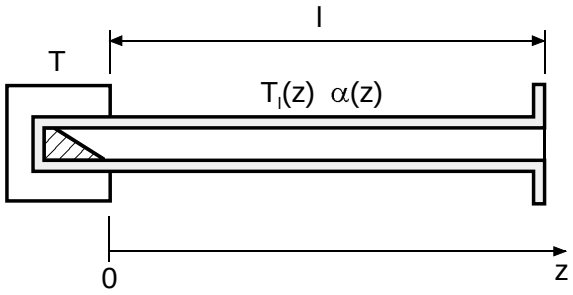

Figure 99. Cold or hot load, with a transmission line between termination and output port.

Gas Discharge Tube

Gas discharge tubes are based on the random acceleration and deceleration of electron in the discharge of ionized argon or neon gas. They are fairly broadband noise sources and their noise temperature T_N is quite high, about 10 000–20 000 K. As an alternative measure for noise sources one commonly uses the excess-noise ratio *ENR*, defined as

$$ENR = 10\log\left(\frac{T_N}{T_0} - 1\right).$$

The *ENR* of gas-discharge tubes is about 15–18 dB. When the tube is switched off, it acts as a matched cold load, i.e. the noise temperature is equal to the ambient temperature. Note that in the above equation T_0 is not the ambient temperature, but standard 290 K!

One can find coaxial and waveguide-type gas discharge tubes. Waveguide gas-discharge tubes are available up to over 200 GHz, limited mainly by physical dimensions. The gas discharge tube is positioned at a 10º–15º angle, at which the plasma is well matched to the waveguide impedance (see Fig. 100). To generate ionization several thousand volts are needed and one has to be careful not to damage the device being measured. A weakness of gas discharge tubes is the variation in matching when switching it on and off. In practice gas discharge tubes are rarely used anymore.

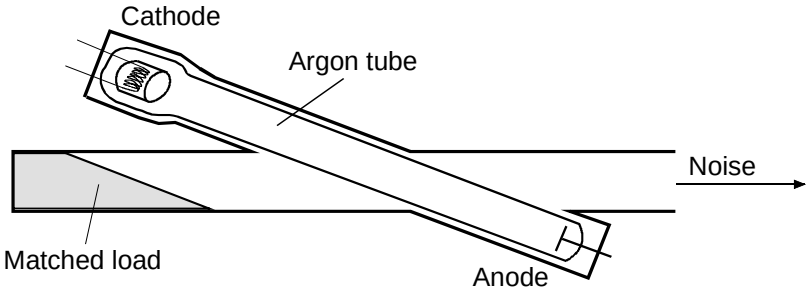

Figure 100. Waveguide gas discharge tube.

Noise diodes

An avalanche diode is a semiconductor diode with such a high reverse bias voltage that an avalanche discharge occurs. Often a DC voltage of 28 V is used. The discharge generates a broadband noise.

A typical *ENR* is 5–15 dB. When the bias voltage is removed, the diode acts as a cold load with a noise temperature equal to its physical temperature. Avalanche diodes are well suited for automated measurement system, in which the diode is switched on and off, e.g. at a rate of 500 Hz. In order to ensure good matching over a wide frequency band, e.g. a 6-dB attenuator is connected between the diode and the output connector (see Fig. 101).

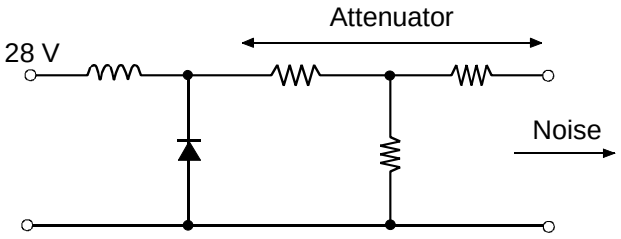

Figure 101. Avalanche diode as a noise source.

Avalanche diodes are compact and reliable. With a single coaxial diode one can cover the frequency range 10 MHz – 26 GHz, and waveguide diodes are available up to 50 GHz. The *ENR* varies as a function frequency and between individual diodes, so each must be calibrated at a sufficient number of frequencies. The uncertainty of the *ENR* is about ± 0.1 dB. The *ENR* values can be stored on an EEPROM accompanying the diode. In the noise source one can integrate a temperature sensor, which then provides the noise temperature of the cold load.

Noise-measurement methods

Methods used for noise measurement are the signal-generator, the noise-generator, and the reference method. In the reference method the noise parameter of a device is compared with a known noise-power standard. This method is suitable e.g. for quality control in mass production. The measurement of the noise factor with a spectrum analyzer was already discussed earlier.

Signal-generator method

In the signal-generator method one needs a calibrated signal generator (i.e. the power level is known accurately) and a power meter. First, the frequency response of the device is measured, so the noise bandwidth B_n can be calculated. Then, the output power of the generator is set to zero, so the noise power at the input of the device is that of a matched load. The power measured at the power meter is $P_{No} = (T_0 + T_e)kG_{t,max}B_n$. The ambient temperature is assumed to be T_0. Finally, at the center frequency of the device (where the gain is $G_{t,max}$) we feed from the signal generator P_{Si}, and now the output power of the device is $P_{(N+S)o} = (T_0 + T_e)kG_{t,max}B_n + P_{Si}G_{t,max}$.

From those two equations we can solve for the noise factor of the device

$$F = \frac{P_{Si}}{(P_{(N+S)o} / P_{No} - 1)kT_0B_n}.$$

The measurement can also be done so that P_{Si} is increased until the power meter shows a 3-dB increase of power (i.e. $P_{(N+S)o}$ has doubled). At that moment, the noise factor is simply

$$F = \frac{P_{Si}}{kT_0 B_n}.$$

The advantage of this method is that it neither needs a calibrated noise source nor the absolute value of the power-meter readings (of course except for P_{Si}). Instead, only the ratios of power levels are measured. Therefore, an accurate measurement can be performed with an adjustable attenuator, in which case a diode detector is sufficient as power meter. However, the determination of the noise bandwidth is tedious and inaccurate. This method is suitable for measuring very noisy devices. These include power amplifiers, whose noise factor F can be 30 dB.

Noise-generator method

The most accurate and widely used method for noise-temperature measurements is the noise generator method, or *Y*-factor method. The noise source connected to the input of the measured device alternately provides two well-known noise-power levels, which correspond to a hot-load noise temperature T_H and a cold-load noise temperature T_C, respectively. One can also measure a semiconductor circuit on-wafer, in which case noise source and power meter are connected to the circuit through measurement probes [53, 54]. The power spectrum of the noise generator should be flat over the current measurement bandwidth. The output noise-power levels are

$$P_H = (T_H + T_e) k G_{t,max} B_n,$$

$$P_C = (T_C + T_e) k G_{t,max} B_n.$$

From these we can solve

$$T_e = \frac{T_H - Y T_C}{Y - 1}, \qquad (12.24)$$

where Y is the ratio of output powers P_H/P_C. This method does not require knowledge of the noise bandwidth.

The method is graphically illustrated in Fig. 102. A straight line is drawn through the two measurement points, and the intersection of the line with the P-axis gives the noise power generated by the device itself, and the intersection with the T-axis gives the according $-T_e$.

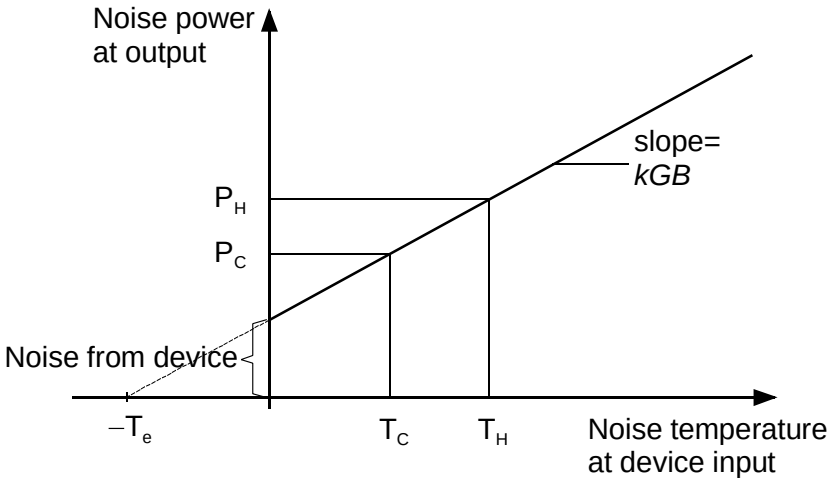

Figure 102. The *Y*-factor method graphically interpreted.

The *Y*-factor can be measured with a power meter, with a spectrum analyzer, or by adjusting a precision attenuator as shown in Fig. 103 so that the power levels at the detector will be identical, in which case the difference in attenuation yields the *Y*-factor. A bandpass filter is used to adjust the instantaneous measurement bandwidth, because otherwise the average noise temperature over a wide frequency band would be the measured. Since noise power levels are small, usually amplifiers are connected between measured device and power meter or detector. The measurement will result in the noise temperature of the whole system (of measured device + meter), so measurements must be corrected as described in the next section.

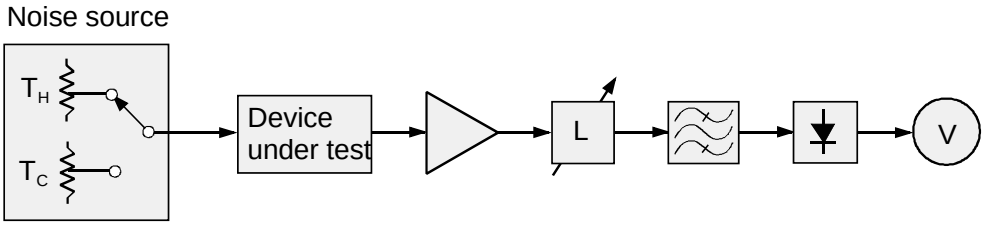

Figure 103. Noise measurement with the *Y*-factor method.

When measuring mixers the *Y*-factor method gives the double-side-band DSB noise temperature, because the noise signal is present in both sidebands. To determine upper and lower sideband SSB noise temperatures, also ratio of the conversion loss L_1/L_{-1} needs to be measured. This is done by feeding the same coherent signal alternately at both sidebands and measuring the power ratio at the intermediate frequency. SSB noise temperature T_{USSB} is then calculated as discussed earlier. In mixer measurement it must be taken into account that the LO affects the measurement result, since the oscillator noise is also mixed to the intermediate frequency.

Noise Meters

Noise can be measured manually by connecting to the device input a hot and a cold load and calculating the noise temperature from the *Y*-factor. It is easier and faster to use a noise meter, whose display directly shows noise figure or temperature. Modern noise meters are automated, carrying out the measurements over a wide frequency band quickly and accurately. Noise factor tuning in real time is thus possible.

Fig. 104 shows the block diagram of an automated noise measurement system. The basic frequency range is 10 MHz–1,5 GHz, which can be extended by inserting a mixer between the measured device and the noise meter. It should be kept in mind that without any filtering, noise from both sidebands is mixed to the intermediate frequency. A microprocessor controls the noise source, frequency synthesizer, the LO of the meter and the attenuators. As a noise source acts an avalanche diode. The frequency synthesizer acts as LO for the external mixer. With the tunable LO the frequency range of the meter is scanned so that the intermediate frequency is 2050 MHz. This is mixed to 20 MHz where the noise signal is detected. In order to avoid compression of the amplifiers, the power level is automatically adjusted with help of the attenuators. The voltage of the detector is AD-converted to digital format. The noise source *ENR* values are in memory. These values are interpolated, if the measurement frequency is between calibration points. The ambient temperature changes are taken into account in the cold temperatures of the noise source. The noise temperature T_m' is calculated from the *Y*-factor. Accuracy is improved when taking the average of several measurements.

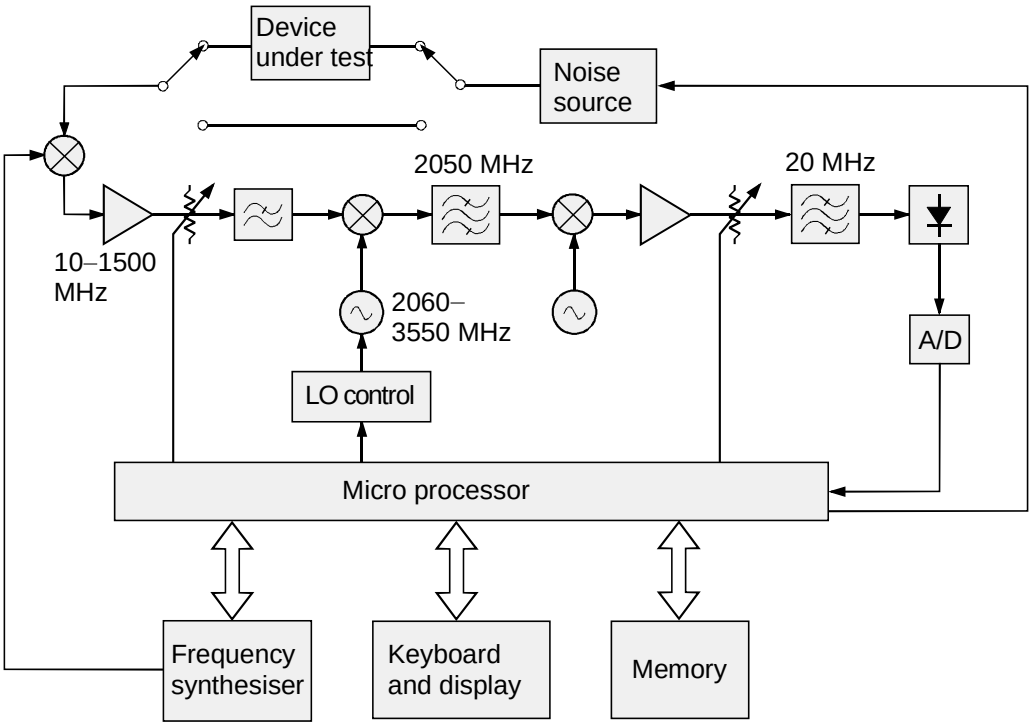

Figure 104. Automated noise meter.

The measured noise temperature T'_m is not the noise temperature T_e of the device being measured, but of the entire system, i.e. of the device being measured and the measuring system. T'_m and T_e differ significantly if the device being measured has low gain. For this reason, also a through measurement is carried out, i.e. the noise source is connected directly to the mixer. This gives both the noise temperature T_m of the measurement system and the gain G of the device. This way the corrected noise temperature is calculated

$$T_e = T'_m - \frac{T_m}{G}.$$

The meter displays the corrected noise measure in the desired format (T_e, F or Y), and the gain.

Noise-parameter measurements

The noise parameters of a two-port are obtained by measuring the noise factor with four well-known input admittances, and by solving the obtained set of equations for F_{min}, R_n, $|\rho_{opt}|$ ja $\angle\rho_{opt}$ [55]. Measuring the noise parameters is slower than a simple noise-factor measurement, but it provides important information for the design of e.g. an amplifier. In the development of integrated circuits it is useful to measure the noise parameters directly on the wafer [56-58].

The known input admittances or reflection coefficients are implemented with an adjustable impedance-tuning circuit (Fig. 105). The reflection coefficient of the generator ρ_S that is visible through the tuning circuit is measured with the vector analyzer in each case. The effect of the tuning circuit on the noise temperature of the source noise must be taken into account. Due to inaccuracies measurements should be done with more than four input admittances, and searching for the best-fit noise parameters. The impedance-tuning circuit can also be used to directly search for $\underline{F_{min}}$ and ρ_{opt}. However, this is a tedious task and does not give the noise resistance R_n.

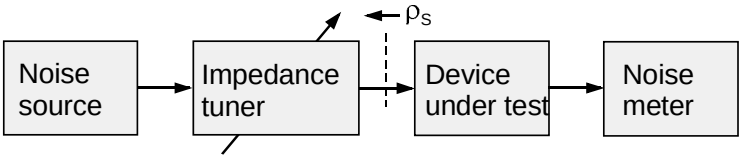

Figure 105. Noise-parameter measurement.

Error sources in noise measurements

In noise measurements the most significant source of error is often the uncertainty of noise temperatures T_H ja T_C. The calibration uncertainty of the *ENR* values of noise diodes and gas discharge tubes is about ± 0.1 dB. *ENR* also depends on ambient temperature and operating voltage: For noise diodes *ENR* variations are about 0.01 dB/ºC and 0.1 dB for a 1% variation in operating voltage.

Noise temperatures T_H ja T_C must also be suitable for the measured noise temperature T_e. If T_e is large, also T_H should be large, and T_e is small, T_C should be small. If the *Y*-factor is close to unity, a small error in the measurement of power ratio causes a large error in the noise temperature. On the other hand, if *Y* is large, the non-linearity of the power measurement becomes significant.

Another major source of error is mismatch. Multiple reflections due to impedance mismatch between noise source and measured device are a difficult problem: some of the noise from the noise source does not reach the device under test, some of the noise generated by the device itself is reflected back from the noise source, and the device characteristics (gain and noise factor) depend on the input matching. The *VSWR* of the noise source also depends on whether it is switched on or not. This is problematic if the device under test is very sensitive to input matching. Also mismatch between measured device and meter is a cause for error. One way to reduce the effect of mismatch is the use of isolators. The attenuation of the isolator must be known so that its effect can be corrected in the results. The error due to the uncertainty in the attenuation may also be larger than the benefits, e.g. when measuring low-noise devices.

The attenuation between noise source and measured device must be known so that the results can be corrected. At room temperature 0.1-dB attenuation causes an increase in noise temperature of 7 K + $0.023T_e$. If the noise source and the device are at different temperatures, in for accurate measurements the temperature and attenuation distribution of the connecting cable must be known.

If the bandwidth of the noise meter is smaller than the frequency band at which the device operates, we may get a too optimistic value for the noise temperature. In this case, the noise temperature should be measured at several different frequencies, and the average noise temperature calculated. It should be remembered that the noise signal is random, and hence the narrower the measurement bandwidth is, the longer the measurement time must be.

Other error sources in noise measurements are the measurement uncertainty of the own noise in the measurement system, errors in meter (detector and amplifier non-linearity, instability, etc.) and ambient disturbances (radio stations, fluorescent lamps, etc.) because of which accurate measurements should be performed in a shielded room.

[47] *Fundamentals of RF and microwave noise figure measurements*. Agilent Technologies, Application Note 57-1, October 2006.

[48] Mumford W. W., Scheibe E. H.: *Noise Performance Factors in Communication Systems*. Dedham, Massachusetts 1968, Horizon House - Microwave.

[49] Räisänen A., Lehto A.: *Radio Engineering for Wireless Communication and Sensor Applications*. Norwood 2003, Artech House, Inc..
[50] Räisänen A. V.: Experimental studies on cooled millimeter wave mixers. Dissertation. Helsinki 1980, Acta Polytechnica Scandinavica, Electrical Engineering Series No. 46.
[51] Hardy W. N.: Precision temperature reference for microwave radiometry. *IEEE Transactions on Microwave Theory and Techniques*, 21(1973)3, s. 149–150.
[52] Stelzried C. T.: Microwave thermal noise standards. *IEEE Transactions on Microwave Theory and Techniques*, 16(1968)9, s. 646–655.
[53] Randa J., Billinger R. L., Rice J. L.: On-wafer measurements of noise temperature. *IEEE Transactions on Instrumentation and Measurement*, 48(1999)6, s. 1259–1269.
[54] Weatherspoon M. H., Dunleavy L. P.: Vector corrected on-wafer measurements of noise temperature. *IEEE Transactions on Instrumentation and Measurement*, 54(2005)3, s. 1327–1332.
[55] Cappy A.: Noise modeling and measurement techniques. *IEEE Transactions on Microwave Theory and Techniques*, 36(1988)1, s. 1–10.
[56] Kantanen M., Lahdes M., Vähä-Heikkilä T., Tuovinen J.: A wide-band on-wafer noise parameter measurement system at 50–75 GHz. *IEEE Transactions on Microwave Theory and Techniques*, 51(2003)5, s. 1489–1495.
[57] Vähä-Heikkilä T., Lahdes M., Kantanen M., Tuovinen J.: On-wafer noise-parameter measurements at *W*-band. *IEEE Transactions on Microwave Theory and Techniques*, 51(2003)6, s. 1621–1628.
[58] Hu R., Sang T.-H.: On-wafer noise-parameter measurement using wide-band frequency-variation method. *IEEE Transactions on Microwave Theory and Techniques*, 53(2005)7, s. 2398–2402.

Measurements of active circuits

Testing of passive, linear circuits is straightforward. With a network analyzer their reflection and transmission characteristics are measured swiftly across a wide frequency range. Testing of active and nonlinear circuits is much more difficult. There are more relevant characteristics: additionally to linear quantities such as gain and reflection coefficients, a number of parameters describe the nonlinearity of active circuits (e.g. compression point, harmonic distortion, intercept points). In addition, their properties depend on power level, operating voltage, etc.

Measurements of nonlinear circuits are sometimes complicated by the fact that the output signal is measured at a different frequency than the signal fed to the circuit. This is done when measuring e.g. harmonic distortion or the efficiency of a frequency multiplier. In these measurements, a normal network analyzer cannot be used, and there are no straightforward calibration methods to reduce measurement system errors.

Active circuits are often sensitive to the input and output load impedance. To avoid that mismatch between signal generator and tested circuit affects the measurement, the input and output of the circuit should be matched with attenuators or isolators.

Since active circuits can have high gain, one must also ensure that the power level does not harm the measurement equipment. Also for this reason the use of attenuators is always recommended, one can always remove them later if after all necessary.

We now look at the measurements of non-idealities of amplifiers and oscillators: compression of the gain due to amplifier nonlinearity, harmonic distortion and intermodulation. Furthermore, measurement methods for amplitude and phase noise of oscillators are presented.

Amplifier measurements

First we look at measurement methods for the measures describing the non-linearity of amplifiers. These methods can be applied to measurement of other nonlinear circuits, such as mixers, or semiconductor-based switches, phase shifters and attenuators.

Gain compression

At low power levels amplifiers operate linearly: output power P_o is directly proportional to input power P_i, i.e. gain G is constant. When the power increases, the amplifier will at some point start to become saturated and its gain is reduced as shown in Fig. 106. Generally, manufacturers specify the 1-dB compression point of an amplifier, i.e. the output power where the gain has decreased by 1 dB from its small-signal value.

The compression point can be determined by simply measuring the gain, while the input power is adjusted in suitable increments. The measurement can be automated by combining a generator, whose output power is sweepable, with a power meter and a network analyzer.

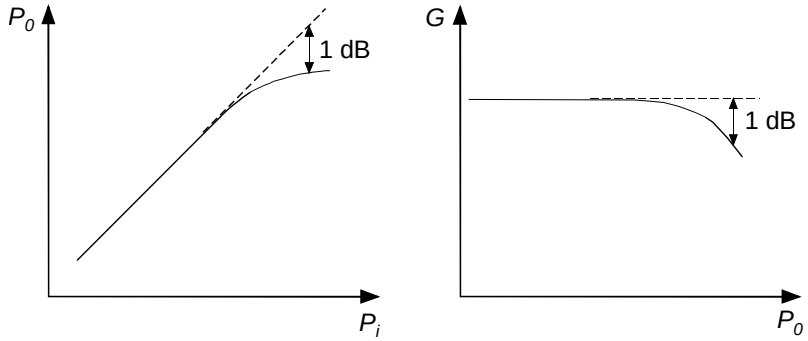

Figure 106. Gain compression represented in P_o–P_i, and G–P_o curves.

Fig. 107 shows how compression can be studied with a scalar network analyzer, as a function of frequency. The sweep generator covers the frequency band of interest. Its power is modulated between two power levels. The output level of the amplifier is kept constant during the sweep by the level control circuit. The display will show signals corresponding to the two power levels, the output power (channel B) and gain (ratio of channels A and R).

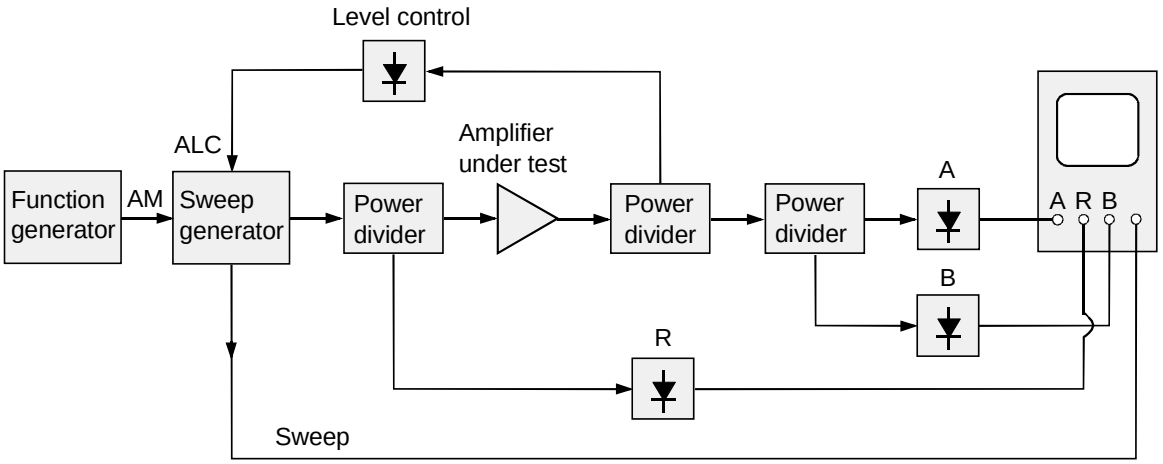

Figure 107. Block diagram of swept measurement for compression point of amplifiers.

Harmonic Distortion

When a pure sine signal is fed to an amplifier, nonlinearity causes distortion in the waveform, creating unwanted harmonic components. This harmonic distortion can be measured with a spectrum analyzer, but the measurement over a wide frequency range is rather tedious.

A swept measurement of harmonic distortion is possible with a network analyzer that operates according to Fig. 108. The LO of the analyzer is a frequency synthesizer whose frequency is multiplied in a step-recovery diode. In normal operation the IF-signal of reference channel (R) is compared to the 1 MHz reference signal in a phase locked loop. When measuring the second or third harmonic, the reference frequency is divided by two or three. For example, to measure the second harmonic, the fundamental frequency is mixed to the frequency $f_s - Nf_{LO} = 0.5$ MHz, as it does not pass through the filter it is not measured. Instead, the distortion component is mixed with double the LO frequency to the frequency of $2f_s - 2Nf_{LO} = 1$ MHz.

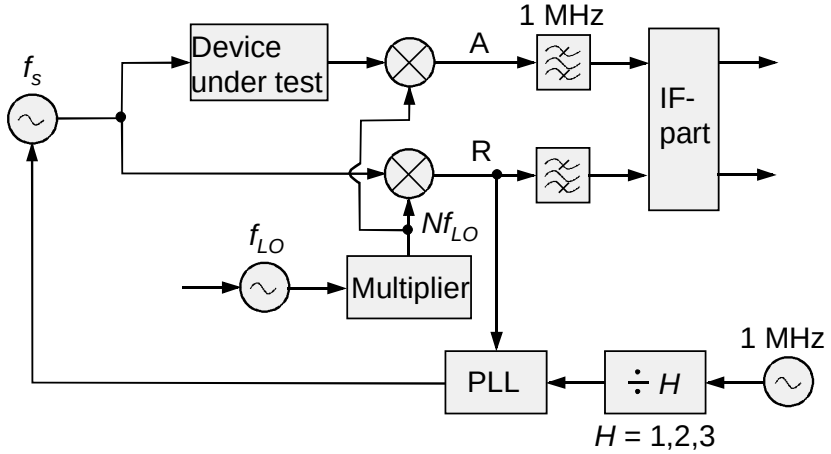

Figure 108. Network analyzer for measuring second and third harmonic distortion components.

Intermodulation

When two or more signals at frequencies f_1, f_2, f_3 etc., are fed to a nonlinear device such as an amplifier, these are mixed. At the output this causes distortion components at frequencies $\pm if_1 \pm jf_2 \pm kf_3$ (i, j and k are positive integers), which are called intermodulation products. The order of intermodulation products is $i + j + k$. For two signals with frequencies close to each other, the third-order intermodulation products at frequencies $2f_1 - f_2$ and $2f_2 - f_1$ are particularly harmful in many applications because their frequencies are so close to the frequencies f_1 and f_2, and therefore difficult to remove with filters. Also second and higher-order intermodulation products are created, but often they are either far from the signal frequencies or exhibit low power levels.

The third-order intermodulation of an amplifier can be measured with a spectrum analyzer as shown in Fig. 109. From two signal generators signals at slightly different frequencies are fed to the amplifier. The signals are combined in a power divider or a directional coupler, and their power levels at the output of the amplifier are adjusted to be equal. The harmonic components of the generators are removed by a low-pass filter. Intermodulation may also occur in other parts of the measurement setup than only in the circuit under test, e.g., in the generators, in the spectrum analyzer, and at high power levels also at the connectors. These levels must be significantly lower than the level of the intermodulation product being measured. Isolators at the outputs of the signals generators prevent power leakage from one to the other generator, where it could cause intermodulation. Also amplifiers can act as isolators, since usually an amplifier attenuates in reverse direction. The intermodulation in the spectrum analyzer is kept sufficiently low by attenuating the power level at its input.

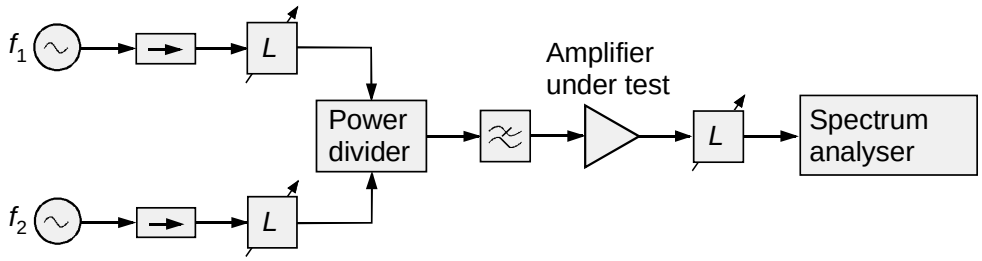

Figure 109. Measurement of the third-order intermodulation.

The display of the spectrum analyzer could look like in Fig. 110. When the level of the main signals at the output is – 5 dBm, the intermodulation products are found symmetrically on both sides with a level 50 dB lower, i.e. the third-order intermodulation distortion IMD_3 = – 50 dBc. (In practice –5 dBm input power would cause strong intermodulation in the spectrum analyzer, therefore an attenuator has to be inserted. See also Fig. 82.)

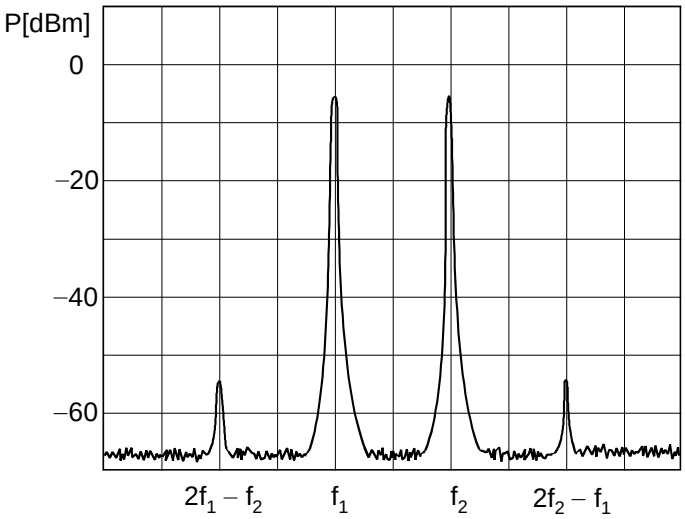

Figure 110. Intermodulation seen on the display of a spectrum analyzer.

Since the level of intermodulation depends on the power level, measurement results are normalized by determining the intercept point, which is used when comparing of amplifiers. The intercept point is the theoretical power level, at which the output power of the fundamental signal and the level of the intermodulation product would be the same (*IMD* = 0 dBc), if there was no saturation.

In Fig. 111 the definition of the third-order intercept point IP_3 is illustrated. The intermodulation product is measured at different power levels and the results are extrapolated yielding an intercept point of +20 dBm. In practice, the output of the amplifier saturates long before the intercept point. The rate at which the intermodulation products change as a function of the power level at the fundamental frequency is proportional to the order, i.e., 1 dB change in input power causes a 3 dB change in the third-order intermodulation product. If at the output power level P_o we determined the intermodulation distortion IMD_3, we get for the intercept point $IP_3 = P_o - 0{,}5 \cdot IMD_3$, where all quantities are in decibels. Accordingly, for the second-order intercept point $IP_2 = P_o - IMD_2$.

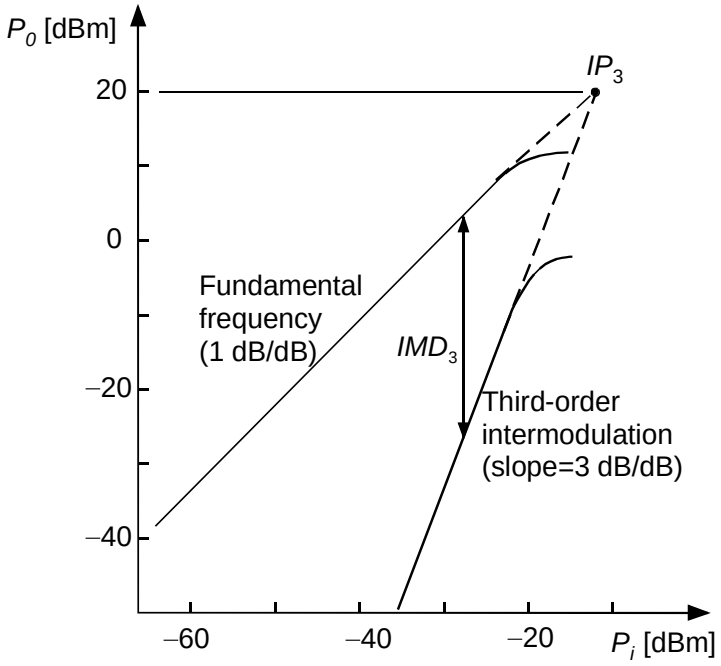

Figure 111. Determining the third-order intercept point.

Other abbreviations for IP_3 and IP_2 are *TOI* and *SOI*, respectively. If we want to emphasize that the intercept points are referred to the device output, abbreviations OIP_3 and OIP_2 are used. The intercept points can also refer to the input of the device, and then we use IIP_3 and IIP_2.

Also, passive circuits may give rise to intermodulation, which is relevant, e.g., in base stations [59], because intermodulation products of the strong transmit signals may fall into a receive frequency band. Passive intermodulation (PIM) is very difficult to quantify, because the necessary power signals are large and the generated intermodulation products may be 150 dB below the level of the desired signal.

Oscillator spectrum measurement

Oscillator noise

In the spectrum of an ideal oscillator there is just one infinitely narrow spectral line. In practice, the signal phase and amplitude of the oscillator vary due to noise. The output voltage can be written as $V(t) = [V_c + \delta V(t)]\sin[2\pi f_c t + \phi(t)]$, where f_c and V_c are the (noiseless) fundamental frequency and amplitude, $\delta V(t)$ and $\phi(t)$ are the nominal variations of the amplitude and phase. Slow fluctuations appear as frequency and amplitude drift, fast fluctuations as phase and amplitude noise, which is why the spectrum typically spread as illustrated in Fig. 112. Amplitude noise is usually much weaker than the phase noise. Phase noise is growing quickly as we get closer to the carrier-wave frequency. The spectrum of a good oscillator reaches the thermal noise level at a frequency $\pm 5\times 10^{-4} f_c$ away from the carrier wave.

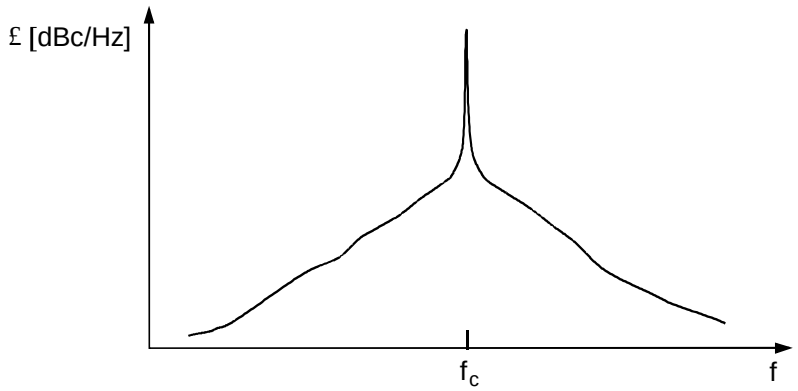

Figure 112. Oscillator spectrum.

Let us consider the situation where the signal phase is modulated with frequency f_m and a sinusoidal deviation $\pm\Delta\phi$ around the nominal value: $V(t) = V_c \sin[2\pi f_c t + \Delta\phi \sin(2\pi f_m t)]$.

Since the frequency can be obtained from the derivative of the phase, this phase modulation can also be interpreted as the frequency modulation: $f(t) = f_c + \Delta\phi f_m \cos(2\pi f_m t) = f_c + \delta f(t)$.

The deviation of this frequency modulation is $\Delta f = \Delta\phi f_m$ and the modulation index $m = \Delta f/f_m = \Delta\phi$. If the modulation index is small, sidebands appear on both sides of the carrier wave at the frequencies $f_c \pm f_m$, with an amplitude ratio $m/2 = \Delta\phi/2$ compared to the amplitude of the carrier wave. Other side bands are negligible. If e.g. $\Delta\phi = 0.002$ radians, the sideband level is –60 dBc. The spectrum in Fig. 112 can be considered the sum of many components at various modulation frequencies f_m.

In reality, $\phi(t)$ does not change in a sinusoidal manner, but in a random manner. So it has components at all modulation frequencies, i.e. a continuous spectrum, which modulates the carrier wave. The spectral density of the pahse fluctuations $S_{\Delta\phi}(f_m)$ is the square of the *rms* value of

the fluctuations in a 1-Hz band, i.e. the quadratic time-average $\overline{\phi_{1Hz}(t)^2}$ (see Fig. 113). In the measurement band B_n the *rms* value of the phase fluctuations is obtained by integrating the spectral density $S_{\Delta\phi}(f_m)$ over the whole frequency band. If the spectral density is constant over the band, $\overline{\phi_B(t)^2} = S_{\Delta\phi}(f_m)B_n$. Similarly, the spectral density of the frequency variations $S_{\delta f}(f_m)$ is the square of the *rms* value of the frequency fluctuations in a 1-Hz band $\overline{\delta f_{1Hz}(t)^2}$. These quantities are of the type rad^2/Hz and Hz^2/Hz.

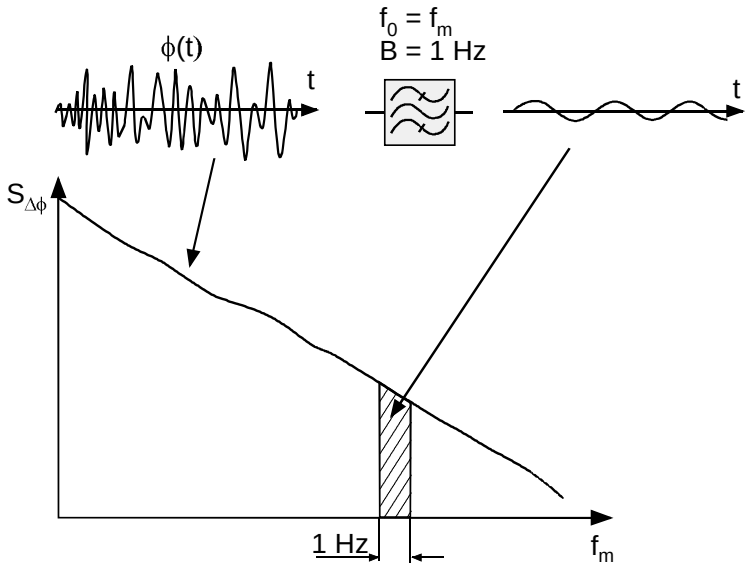

Figure 113. The spectrum of the phase fluctuations $S_{\Delta\phi}(f_m)$.

Phase noise (and amplitude noise) is usually expressed in single-side band power density or as a spectrum £(f_m), i.e. as a noise power in a 1-Hz band relative to the power of the carrier wave as a function of the modulation frequency f_m. The unit is therefore 1/Hz, or in decibels dBc/Hz.

If the nominal phase deviation of the carrier wave is small (less than 0.2 radians) we can state a relation between phase and frequency noise [60]:

$$S_{\Delta\phi}(f_m) = \frac{1}{f_m^2} S_{\delta f}(f_m) = 2£(f_m).$$

Phase noise is particularly harmful for communication and radar systems that are based on frequency or phase modulation. For example, phase noise weakens the measurement accuracy for the speed in a Doppler radar.

Phase noise measurements

Oscillator phase noise can be simply and quickly measured by connecting the oscillator directly to a spectrum analyzer. This gives single-side-band spectrum £(f_m) of the carrier wave. Two corrections need to be made to the results: 1) during any noise measurement with a spectrum analyzer the video filter must be switched on and as described earlier 2.5 dB must be added to the reading; 2) the result is divided by the analyzer noise bandwidth B_n in order to obtain the power density per Hertz. The sensitivity of the direct spectrum analyzer measurement is limeted by the amplitude noise, which cannot be distinguished from phase noise. Also the spectrum analyzer creates its own phase noise. The measured signal must be stable so that the carrier wave frequency does not drift significantly during a sweep.

More accurate than the spectrum analyzer method are the double-oscillator and the delay-line measurement methods. These are based on measuring the spectral density of the phase fluctua-

tions. To measure slow noise components very close to the carrier wave, the frequency counter method is well suitable.

Dual-oscillator measurements

In the dual-oscillator or phase-detection method, the signal phases of the oscillator under test and a reference oscillator are compared with help of a double-balanced mixer as shown in Fig. 114. The output signal of the mixer is low-pass filtered and amplified before reaching the spectrum analyzer. If the frequency is very high, the oscillator frequency needs to first be mixed to a lower frequency. In this case, the LO phase noise is added to the measured phase noise.

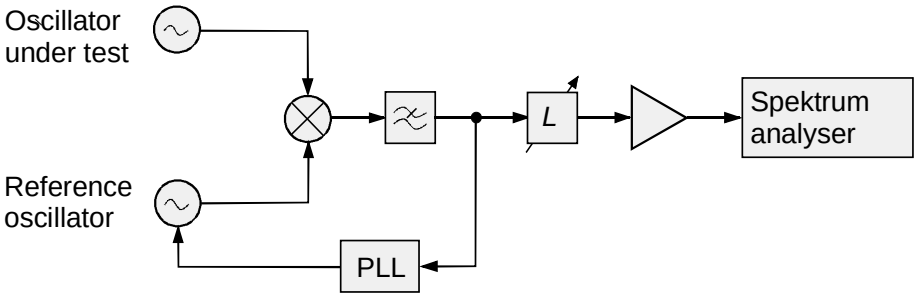

Figure 114. Dual-oscillator measurement.

If at the input of the double-balanced mixer we have the signals from both the oscillator under test $v(t) = V_c \sin[2\pi f_c t + \Delta\phi \sin(2\pi f_m t)]$ and the reference oscillator $v_r(t) = V_r \sin(2\pi f_c t + \pi/2)$, i.e. the phase difference is in average 90º, the output voltage of the mixer is directly proportional to $\phi(t)$:

$$V_o(t) = KV_cV_r\Delta\phi\sin(2\pi f_m t) = KV_cV_r\phi(t)\,,$$

provided that $\phi(t)$ is smaller than 0.2 radians; K is a proportionality factor. Since a small $\phi(t)$ is directly proportional to the voltage that modulates the carrier wave, the power reaching the spectrum analyzer is directly proportional to the modulation power. In practice, the phase variation is of random nature and the measurement bandwidth is B_n. In this case, the *rms* value of the output voltage can be obtained by replacing $\phi(t)$ with the *rms* values integrated over the measurement bandwidth.

In order to keep the oscillator frequencies equal and their phase difference at about 90º, the reference oscillator is phase-locked to the measured oscillator. All those noise components that are far from the carrier wave, i.e. outside the bandwidth of the phase-locked-loop, are attenuated. So, the loop bandwidth should be reduced when we measure only near the carrier wave. The oscillator instability limits the bandwidth reduction. The phase lock does not affect fast phase changes that lie outside the loop bandwidth. Oscillators can be locked to each other in also with help of a common reference oscillator. In this case, a phase shifter needs to be added to one branch. The phase difference of 90º between the signals is ensured by adjusting the phase shifter so that the mixer output DC voltage is zero.

The measurement is calibrated by detuning one oscillator slightly to f_c'. The amplitude of the frequency-difference signal ($f_c' - f_c$) is KV_cV_r. This corresponds to the level of the carrier wave of the oscillator under test. Inserting attenuators ensures that amplifier and spectrum analyzer will not saturate during the calibration.

During the actual measurement the frequencies of both oscillators are equal. All previously discussed corrections must also be remembered here. Since the lower and upper sideband voltages sum up at the mixer output, 6 dB is subtracted from the measured level to obtain the single-sideband spectrum. Since the measurement is the sum of the phase noise of both oscillators, the

reference oscillator should exhibit clearly lower noise than the measured oscillator. If the oscillators are identical, though, the measured result is simply reduced by 3 dB.

The dual-oscillator method is suitable for measuring phase-locked oscillators but not for measuring free-floating oscillators, as the phase difference of the two signals at the mixer input varies too much.

This method can also be used to measure for example the phase noise of amplifiers (residual phase noise). In this case, the oscillator signal is split into two branches, the first of which is connected to the amplifier under test and the second branch to a phase shifter, with which the phase difference of the two signals at the mixer input is adjusted to exactly 90°.

Delay-line measurement

In the delay-line method the signal is mixed with a delayed copy of itself as shown in Fig. 115, so that a reference oscillator is not required. The signal of the measured oscillator is split into two branches. The first branch is a delay line and the second one contains a phase shifter with which the phase difference of the two signals at the input of the double-balanced mixer is adjusted to exactly 90º. The delay line can be a coaxial cable or a narrow-band filter. The output signal of the mixer is treated the same way as in dual-oscillator measurements.

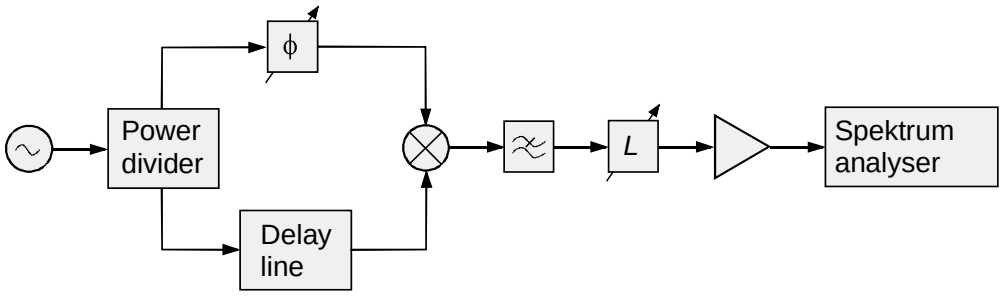

Figure 115. Delay-line measurement setup

If the nominal value of the frequency deviation $\delta f(t)$ changes sinusoidally with the frequency f_m and the length of the delay line is τ_d, the phase difference of the signals at the mixer input is $\phi'(t) = 2\pi\tau_d \delta f(t)$, provided that $f_m << 1/\tau_d$ (Fig. 116). So, by measuring the spectral density of the phase variations at the input of the mixer $S'_{\Delta\phi}(f_m)$, we get the spectral density of the oscillator-frequency fluctuations

$$S_{\delta f}(f_m) = \frac{S'_{\Delta\phi}(f_m)}{(2\pi\tau_d)^2},$$

from which we can, as done earlier, calculate spectral density of the oscillator-phase fluctuations. If f_m is large compared to $1/\tau_d$, the phase difference $\phi'(t)$ is proportional to $\sin(\pi f_m \tau_d)$. The mixer output is therefore zero at the frequencies $f_m = n/\tau_d$.

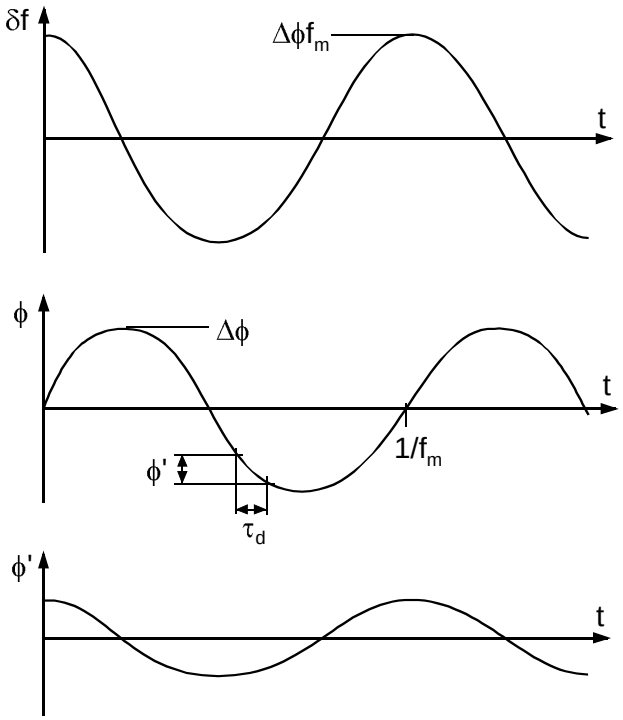

Figure 116. Nominal values of frequency deviation $\delta f(t)$ and phase deviation $\phi(t)$ of the oscillator. Phase difference $\phi'(t)$ of the signals at the mixer input in the delay line measurement.

The delay line is calibrated with a modulated signal with known deviation. For example, a signal with m = 2,405 is easy to produce, since here the carrier wave vanishes for the first time. The effective frequency deviation of this signal is $\delta f_{cal} = 2{,}405\, f_m / \sqrt{2}$. The corresponding *rms* value of the mixer output voltage V_{cal} is measured. During the calibration the input power of the analyzer is attenuated. When measuring the phase noise over the band B_n we get the effective voltage V_{rms}, and the according frequency deviation is

$$\delta f_{rms} = \delta f_{cal} \frac{V_{rms}}{V_{cal}} .$$

The spectral density of the frequency variations can now be calculated. If the measured oscillator cannot be modulated, one can use another signal generator for the calibration, as long as frequency and power are the same as during the measurement, and the phase difference is kept at 90°.

Frequency-counter measurement

In Fig. 117 the block diagram of a simple method based on a frequency counter. The intermediate frequency varies as much as the frequency of the oscillator under test, as long as the LO of the mixer is stable. With a counter that has a good resolution the average frequency is repeatedly determined from the measurement period τ. From the result the Allan variance [61] is calculated. When the Allan variance is known as a function of τ, $S_{\delta f}(f_m)$ can be determined near the carrier wave better than with any of the methods described above.

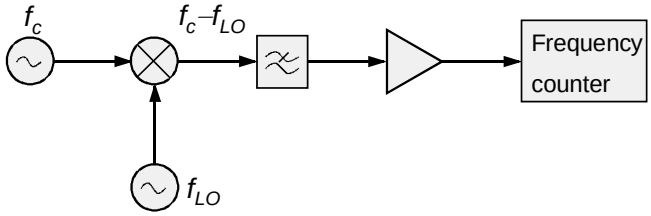

Figure 117. Frequency-counter measurement setup.

Amplitude-noise measurements

The amplitude noise of an oscillator can be measured as shown in Fig. 118. The detector diode acts as a linear envelope detector, and its output gives only the amplitude fluctuations and not the phase fluctuations. The output voltage is fed through a video amplifier to a spectrum analyzer.

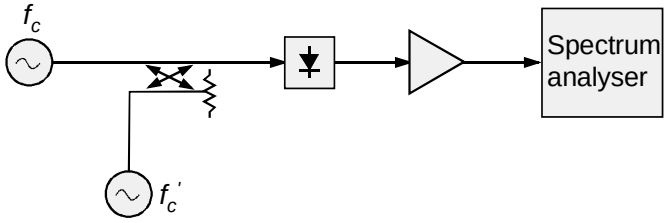

Figure 118. Amplitude-noise measurement setup.

The measurement is calibrated by feeding through a directional coupler to the detector an exactly known, about 50 – 90 dB weaker signal at frequency f_c', which differs slightly from oscillator frequency f_c. In the detector the calibration signal is mixed down to frequency $|f_c' - f_c|$. The single-side-band amplitude noise of the calibration signal is 6 dB lower than the power of the calibration signal at the output of the detector.

[59] Hienonen S., Golikov V., Vainikainen P., Räisänen A. V.: Near-field scanner for the detection of passive intermodulation sources in base station antennas. *IEEE Transactions on Electromagnetic Compatibility*, 46(2004)4, s. 661–667.

[60] Lance A. L., Seal W. D., Labaar F.: Phase noise and AM noise measurements in the frequency domain. Luku 7 kirjassa *Infrared and Millimeter Waves*, Volume 11, Millimeter Components and Techniques, Part III (Button K. J.). Orlando 1984, Academic Press, s. 239–289.

[61] Allan D. W.: Statistics of atomic frequency standards. *Proceedings of the IEEE*, 54(1966)2, s. 221–230.

Time-domain measurements

Microwave measurements are usually done in the frequency domain: scattering parameters are measured with a network analyzer as a function of frequency; modulated signals are investigated with a spectrum analyzer. The signal waveform versus time can be measured with an oscilloscope, but normal analog oscilloscopes can cover only lower microwave frequencies: the maximum bandwidth is about 1 GHz.

Sampling oscilloscopes have significantly improved the possibilities of time-domain measurements. Their bandwidth can be tens of gigahertz. Sampling oscilloscopes are suitable not only for measuring signal waveforms but also for reflection and transmission measurements in time domain. With help of time-domain network analyzers we can through reflection and transmission measurements determine the frequency response of a circuit. Time-domain measurements can also be simulated with a vector network analyzer.

Applications of time-domain measurements are, e.g., finding cable faults, measuring the characteristics and model parameters of microwave circuits, elimination of the effect of reflections on measurement results, etc.

Sampling oscilloscope

Digital sampling oscilloscopes have widely replaced analog oscilloscopes. They are suitable for measurement of single phenomena and periodic signals. Measured wave forms can be stored in memory and signal processing is much easier compared to analog oscilloscopes. From the measured samples the spectrum can be calculated with help of the fast Fourier transform.

With the oscilloscope samples of the signal voltage are taken, by e.g. closing a high-speed switch for a moment, so that a capacitor will be charged proportionally to the measured voltage. The time period during which the sample is taken must be short compared to the cycle length of the measured signal. With a Schottky-diode switch rise time of about 10 ps can be achieved, which corresponds to a bandwidth of 30 GHz. Switches based on the Josephson-junction technology are even faster, but they require cooling, making them difficult to use. The sample taken with the switch is converted into digital form with an AD-converter. At best, the sampling rate of AD-converters is 8 gigasamples per second with 8-bit resolution and a bandwidth of 1.5 GHz. The vertical selectivity can be improved by averaging.

The bandwidth of the oscilloscope should be at least 3 times the maximum frequency component f_{max} of the signal (the Nyquist theorem states that the bandwidth must be $\geq 2\ f_{max}$). The maximum frequency component is approximately $0{,}35/t_{r,sign}$, where $t_{r,sign}$ is the shortest rise time of the signal. The rise time measured with the oscilloscope is $t_r = \sqrt{t_{r,osc}^2 + t_{r,sign}^2}$, where t_r is the rise time of the oscilloscope. For example, the rise time of a 250-MHz oscilloscope is 1.4 ns. It can be used to measure the signal rise time with 5% accuracy, if $t_{r,sign}$ = 4.2 ns. The higher the bandwidth ratio between oscilloscope and signal, the smaller the error.

The frequency range of a digital oscilloscope can be extended considerably by using a special sampling technique. This can be used for periodic signals. With the so-called sequential sampling the waveform of the measured signal is reconstructed from evenly distributed sample points as illustrated in Fig. 119. The measured signal provides a trigger, after which the sample is taken with a certain delay. This delay increases linearly so that the sample point moves slowly relative to the measured periodic signal. Triggering and sampling do not have to happen within each period as shown in Fig. 119, the sampling moment can also be completely random relative to the measured signal [62]. In this case, also the delay between the trigger moment given by the measured signal and the actual sampling moment is measured and stored along with the measured voltage in the memory. From the stored measurement points the waveform is reconstructed.

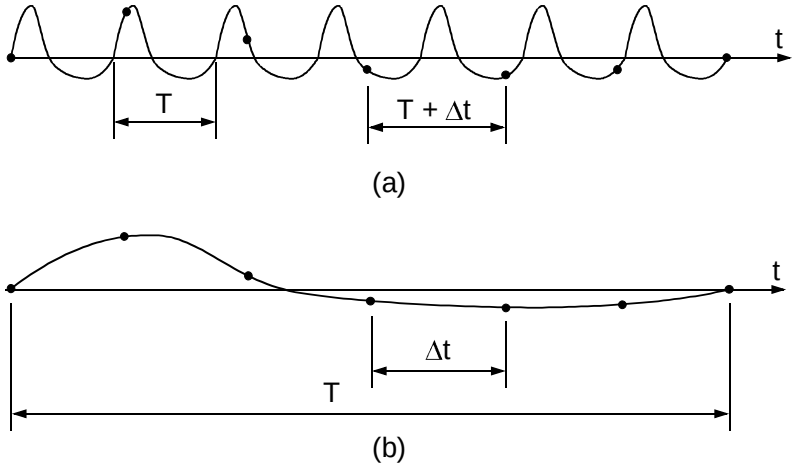

Figure 119. Sequential sampling principle: (a) the measured signal, (b) signal waveform from the collected samples.

Time-domain reflectometry measurements

Reflection measurements with a network analyzer or a slotted line reveal the existence of impedance discontinuities in a transmission line or a circuit. Such a measurement does, however, not tell the location of such a reflection and is unable to distinguish whether there are several discontinuities. The measured reflection coefficient can also be zero at one frequency, if the reflected waves cancel out each other. So, to determine circuit properties, the measurements often need to be performed at many frequencies.

Time-domain reflectrometry, or TDR measurements do not exhibit these disadvantages. For TDR measurements a step generator and an oscilloscope (Fig. 120) are needed. The step generator can be based on a tunnel or a step-recovery diode. The voltage step is propagating along the transmission line or circuit under test, and with an oscilloscope connected close to the generator the sum of forward and reflected voltage waves is measured. The measured waveform gives circuit impedance versus distance. With a pulse generator, i.e. by repetitively generating steps, and the above-described sampling technique, good time and spatial resolutions can be achieved.

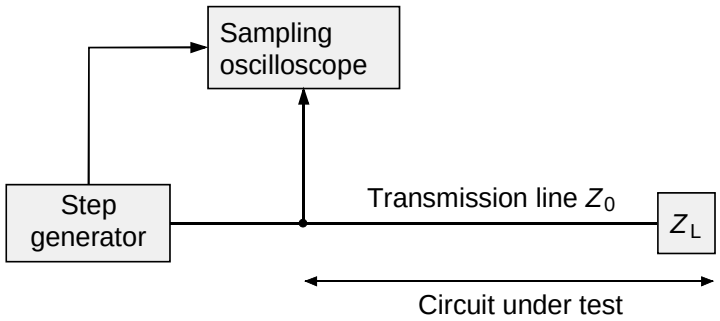

Figure 120. Time-domain reflectometry (TDR) measurement setup.

Feeding a voltage step into a coaxial line is easy, but the circuit under test can be, e.g., a digital integrated circuit or a hybrid circuit [63]. Pulses can be used instead of steps. In waveguides a high-frequency pulse must substitute for the voltage step.

The measured waveform reveals the location and the type of discontinuities, i.e. whether a discontinuity is resistive, inductive or capacitive. We can also conclude whether losses along the line are in a series or shunt circuit. During the development of a circuit the optimization is much

easier on the basis of such information than using only the reflection coefficient. First the discontinuity closest to the measurement point is fixed, the improved performance is verified with another measurement, then the next discontinuity can be corrected, etc.

Some examples are shown in Fig. 121: a transmission line with characteristic impedance Z_0 is terminated with an open termination, a short circuit, load $2Z_0$ and load $Z_0/2$. The voltage step propagating towards the load reaches the oscilloscope at $t = 0$, where the voltage rises to V_i. When the wave reaches the load Z_L, the reflection coefficient is

$$\rho = \frac{V_r}{V_i} = \frac{Z_L - Z_0}{Z_L + Z_0},$$

from which the reflected voltage V_r can be solved. If the load is an open termination, $V_r = V_i$. Within the time t_d the reflected voltage reaches the oscilloscope. The voltages are summed up, and the oscilloscope measures the voltage $V_i + V_r = 2V_i$. If the load is a short circuit, $V_r = -V_i$ and the sum is zero. The location d of the discontinuity is obtained from the delay time t_d:

$$d = v_p \frac{t_d}{2},$$

where v_p is the wave propagation speed in the transmission line.

In case of Fig. 121 the loads were frequency independent, so that the reflection coefficient is the same for all frequency components of the voltage step and the shape of the step is not changed by the reflection from the load. In Fig. 122 we see various waveforms caused by more complex loads, which can be calculated with help of the Laplace transform. The individual component values of the load circuit are derived from the time constants, the amplitude steps and the asymptotic values of the measured waveforms.

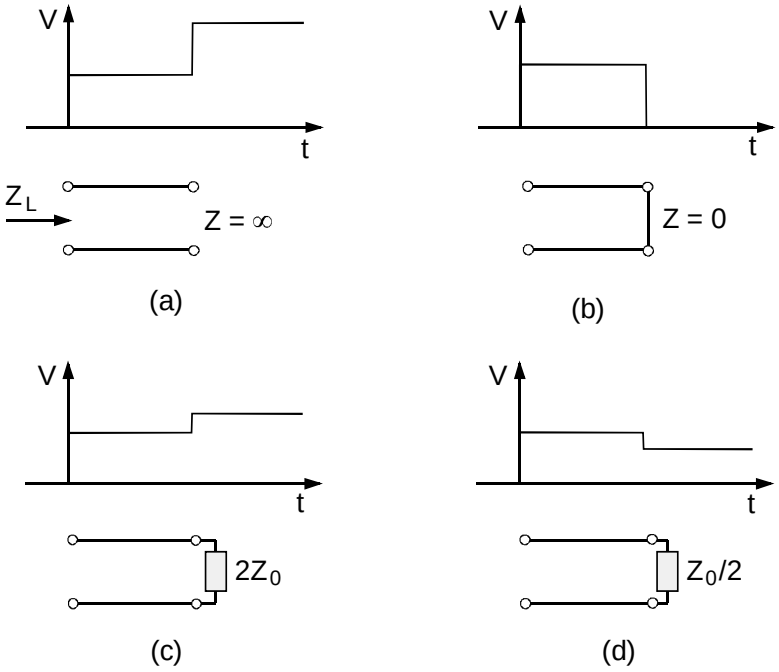

Figure 121. Oscilloscope display in a TDR measurement, when the load is (a) open termination, (b) short-circuit, (c) $2Z_0$, and (d) $Z_0/2$. The characteristic impedance of the line is Z_0.

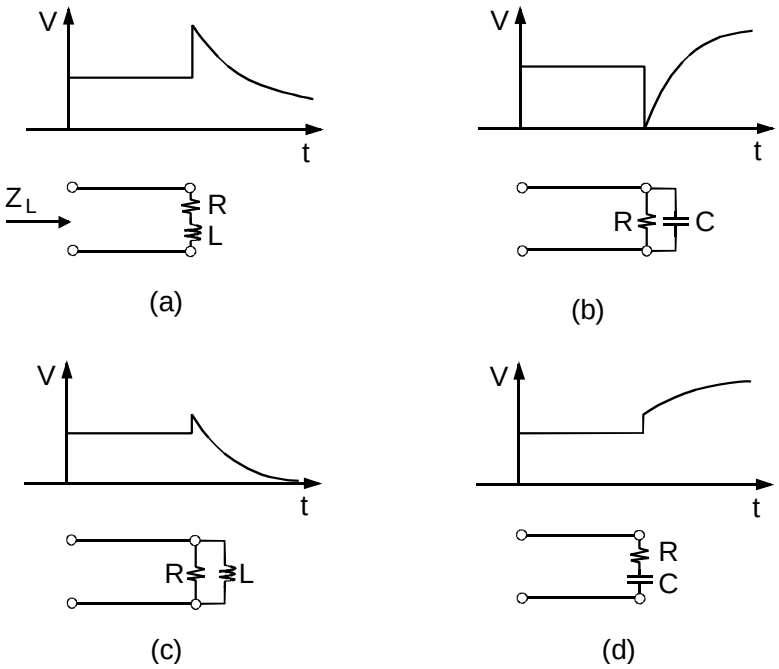

Figure 122. Oscilloscope display in a TDR measurement with more complex loads, such as: (a) *RL*-series circuit, (b) *RC*-parallel circuit, (c) *RL*-parallel circuit, and (d) *RC*-series circuit.

If there are several discontinuities, they may be distinguishable. Fig. 123 shows an example: two transmission lines are connected with a connector, which causes a small series inductance, and the termination of the second line is not matched. The discontinuities can be analyzed separately.

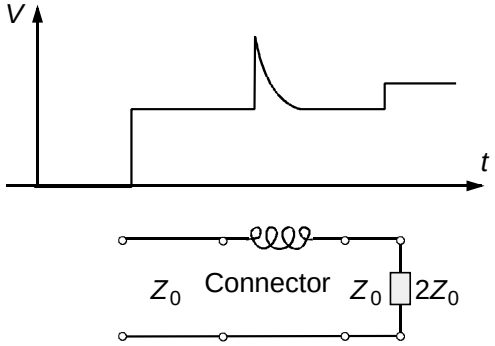

Figure 123. Oscilloscope display in a TDR measurement with a series inductance in the transmission line and a mismatched termination.

The resolution distance of two discontinuities d_{min} depends on the rise time t_r of the system:

$$d_{min} = v_p \frac{t_r}{4}$$

The rise time t_r depends on the rise times of step generator and sampling circuit, and is defined by the time in which the step increase from 10% to 90% of the final value. In the above figures the situation was ideal, i.e. $t_r = 0$ s. In practice, rise times are in the range of tens of picoseconds, so d_{min} is a few millimeters. Rise time is also limits the ability to measure small reactances.

If the reflections are large and there are several, the measurement accuracy will suffer. Multiple reflections between the discontinuities complicate the situation considerably and make the interpretation of results difficult. In practice, however, reflections are often small, in which case the effect of multiple reflections is negligible. The step generator must be well matched to the

transmission line. Otherwise, the part of V_r is reflected by a generator and summed to the measured voltage.

Time-domain transmissometry, or TDT measurements can be done the same way as reflectometry measurements. With TDT measurements the transit time delay and other features of a circuit can be determined.

Time-domain network analyzer

The representations of a signal in frequency and time domains contain the same information. From any time-domain curve $F(t)$ we calculate the frequency spectrum $G(f)$ with help of the of the Fourier transform $\boldsymbol{F}$

$$G(f) = \boldsymbol{F}[F(t)] = \int_{-\infty}^{\infty} F(t) e^{-j2\pi ft} dt ,$$

or, accordingly, from the spectrum the waveform is calculated through the inverse Fourier transform $\boldsymbol{F}^{-1}$

$$F(t) = \boldsymbol{F}^{-1}[G(f)] = \int_{-\infty}^{\infty} G(f) e^{j2\pi ft} df .$$

The time-domain network analyzer (TDNA) shown in Fig. 124 is based on TDR and TDT measurements and the fast Fourier transform FFT calculated from the samples. A short pulse with a wide spectrum is repetitively fed to the circuit under test. The sampling oscilloscope is used to measure the input pulse waveform $v_i(t)$ and the waveform of the transmitted pulse $v_t(t)$. The spectra of the pulses are calculated from the waveforms, and the frequency response of the transmission coefficient is obtained

$$S_{21}(f) = \frac{\boldsymbol{FFT}[v_t(t)]}{\boldsymbol{FFT}[v_i(t)]}.$$

Similarly, from the waveform $v_r(t)$ of the reflected pulse frequency response of reflection coefficient is calculated

$$S_{11}(f) = \frac{\boldsymbol{FFT}[v_r(t)]}{\boldsymbol{FFT}[v_i(t)]}.$$

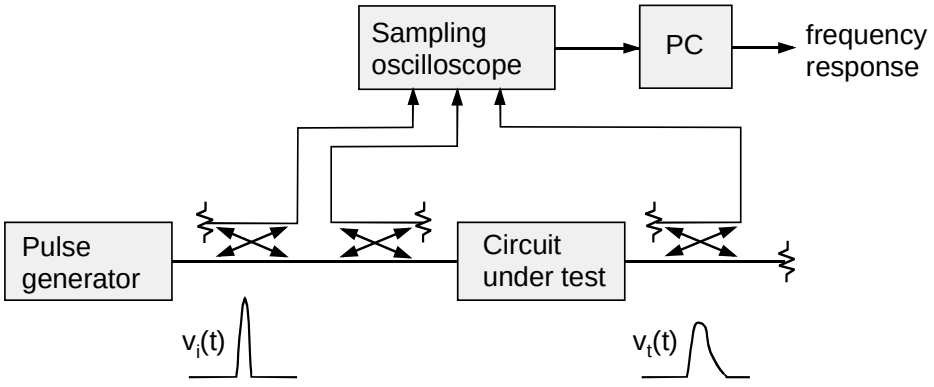

Figure 124. Time-domain transmission measurement with a network analyzer.

Instead of pulses also voltage steps with fast rise-times can be used. It is essential that the waveform contains a fast change, so that the spectrum of $v_i(t)$ is broad. Using pulses with rise or fall times in the range of picoseconds, it has been possible to measure the frequency responses of millimeter wave circuits [64, 65]. To ensure that $v_i(t)$ contains enough energy over a wide frequency band, the voltage of the pulse has to quite large. Therefore, the method is best suited for

testing linear circuits. As in the case of vector network analyzers, the calibration is important to correct systematic errors also when using time-domain network analyzers [66].

Vector network analyzers in time-domain measurements

The use of vector network analyzers for simulating time-domain measurements is based on the direct link between frequency and time domain. From the frequency responses of reflection and transmission, $S_{11}(f)$ and $S_{21}(f)$, the time-domain responses $S_{11}(t)$ and $S_{21}(t)$ are calculated for a desired stimulus. Modern circuit analyzers often include software that calculates these conversions automatically [67]. Compared to actual time-domain measurements we also get the benefits of vector network analyzers such as a good dynamic range due to the narrow instantaneous measurement bandwidth, and a reduction of systematic errors due to the error model obtained from the calibration.

The spectrum $g(f)$ of an arbitrary input excitation $f(t)$ is calculated with the Fourier transform $\boldsymbol{F}$ discussed earlier. It is composed of components, which are the harmonics of the repetition frequency of the stimulus. If the circuit is linear, each component can be considered being independently reflected or transmitted and changing according to the frequency response of the circuit [68]. (In the case of time-domain network analyzers this really happens, but now the excitation is calculational.) In other words the spectrum $G(f)$ equals $g(f)S_{11}(f)$ or $g(f)S_{21}(f)$. The time-domain response $F(t)$ is obtained by linearly summing up the components, which corresponds to the integration defined above in the Fourier transform $\boldsymbol{F}$. Pulses or steps are used as excitations. The spectrum of a pulse is constant, i.e. it contains all frequencies equally. Therefore, the pulse responses of the reflection and the transmission $S_{11}(f)$ and $S_{21}(f)$ can be directly calculated with the inverse Fourier transform. Fig. 125 illustrates how the impulse response of the circuit is calculated. It should be kept in mind that the reflection response $S_{11}(f)$ of the circuit presented in the figure is complex.

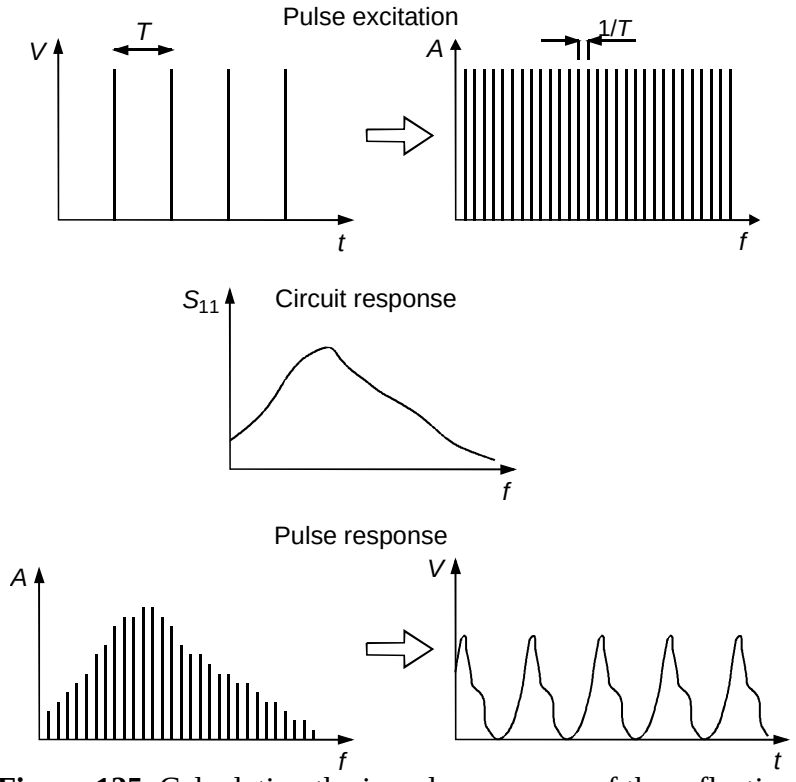

Figure 125. Calculating the impulse response of the reflection.

Network-analyzer measurements are made over a finite frequency range: the measurement can be a low- or band-pass measurement. Low-pass measurements are suitable for lowpass-type of devices, and extends nearly to the frequency zero, for which the response can be extrapolated. The calculational excitation can be a pulse or step. The results correspond to the TDR method described above and contain information about the impedance. Band-pass measurements are suitable for bandpass-type device, and do not extend to zero frequency. The excitation can only be a pulse. The reflection measurement reveals the location of discontinuities, and the average absolute value within the measurement band. When measuring the transmission we obtain the electrical lengths of different routes and average attenuations.

Let us consider the effect of the measurement bandwidth. For example, the pulse response of the reflection of an ideal short-circuit is a pulse. Since the measurement is made over a finite bandwidth $f_s = f_2 - f_1$, instead of the ideal pulse we have a pulse with a shape $|\sin x/x|$ (in time domain), where $x = \pi t f_s$. The interpretation of the time-domain results is now complicated by some kind of "side lobes". For this reason, the frequency-domain data is filtered prior to conversion. The amplitude is multiplied with a windowing function. The side lobes in time-domain weaken, but the response is also flattening, yielding lower resolution. If the windowing function is Gaussian, no side lobes appear in time-domain. In Fig. 126 the windowing effect on the pulse and step responses of a short circuit is presented.

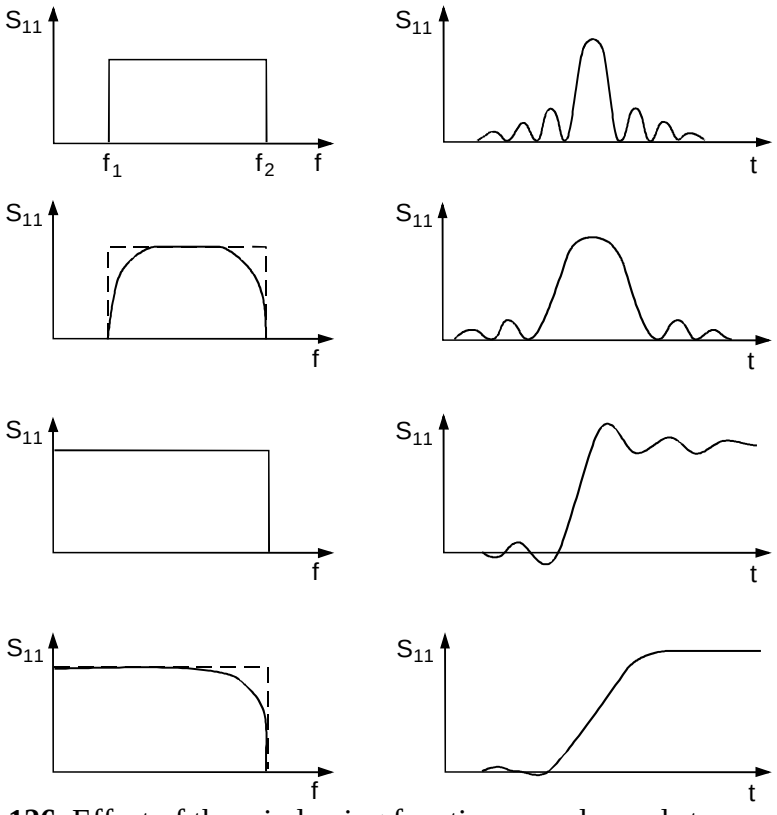

Figure 126. Effect of the windowing function on pulse and step responses.

The pulse width (between the –6-dB points) is approximately $0{,}6w_l/f_s$ is the case of low-pass measurements and twice that in band-pass measurements. The w_l coefficient is 1 – 2.4, depending on the windowing function. The step rise time is $0{,}45w_s/f_s$, where w_s = 1–3,3. For a good resolution a broad frequency range f_s is needed. This method is thus best suited for testing broadband circuits.

Since a network analyzer measures at discrete frequencies with spacing Δf instead of over a continuous spectrum, the time-domain response is periodic at intervals $1/\Delta f$. This can cause con-

fusion when interpreting the results. We can increase the time range free of recurring waveforms by reducing Δf. Accordingly, the number of measurement points must be increased so that f_s and at the same time the resolution does not decrease.

The time-domain representation can be used not only for locating discontinuities but also to study frequency characteristics of different parts in the circuit, or to eliminate undesired effects of discontinuities from the results. This is realized by the so-called time-gating filter, with which the time-domain response is multiplied. The time-gate is placed around the desired measured response, hereby eliminating unnecessary responses. After this we switch back to the frequency domain. Different shapes of time-gating filters are available.

Fig. 127 illustrates the use of a time-gating filter. A 30-cm long air-insulated coax cable with a matched load is measured over a frequency band, and $S_{11}(f)$ is calculated from the time-domain pulse response. In the time-domain graph we see the reflections from connector (between the analyzer and the cable) and termination separately. When the time-gating filter is placed around the connector, this will show the frequency response of the –55-dB mismatch. In Fig. 128 the same cable is terminated with a short circuit. Now we see the strong reflection with a level of 0 dB. Again the time-gate is placed around the connector, and the frequency response is almost the same as in the previous figure. In other words, a very small reflection can still be detected and analyzed even in the presence of very strong reflections.

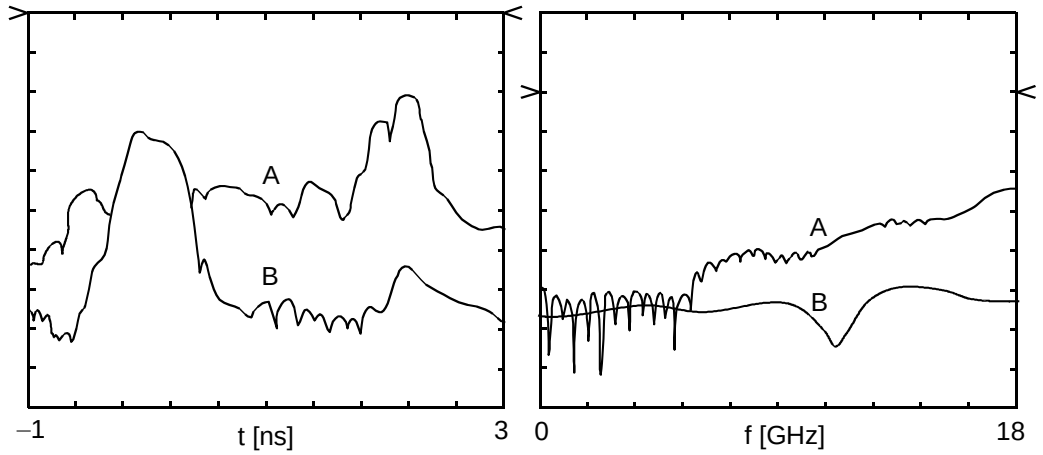

Figure 127. Use of time gating: cable with matched termination. Time and frequency responses before time gating (A) and after (B). Vertical scale: 20 dB/div., 0-dB marks (> <).

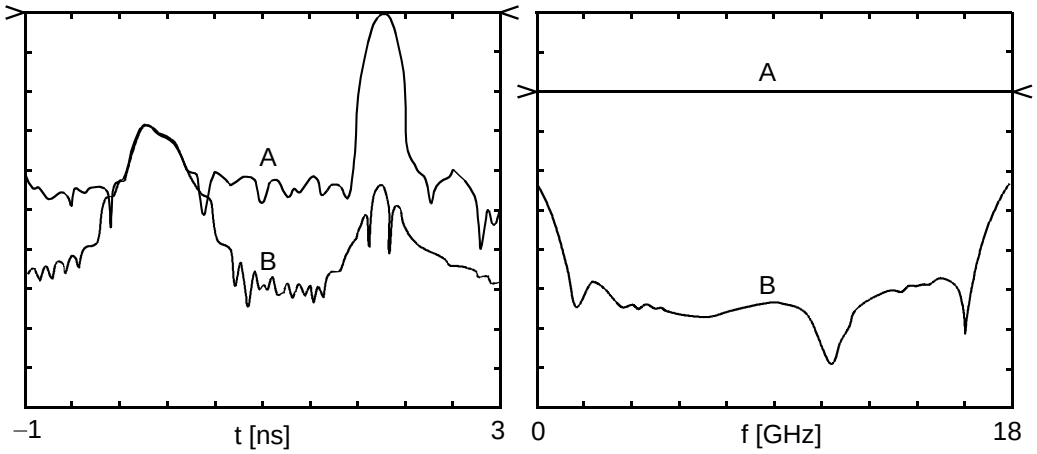

Figure 128. Use of time gating: same cable terminated with a short-circuit.

[62] Rush K., Oldfield D. J.: A data acquisition system for a 1-GHz digitizing oscilloscope. *Hewlett-Packard Journal*, April 1986, s. 4–11.
[63] Frieden B.: Advanced TDR techniques. RF & Microwave Measurement Symposium and Exhibition, 1990, Hewlett-Packard, s. 139–154.
[64] Konishi Y., Kamegawa M., Case M., Yu R., Allen S. T., Rodwell M. J. W.: A broadband free-space millimeter-wave vector transmission measurement system. *IEEE Transactions on Microwave Theory and Techniques*, 42(1994)7, s. 1131–1139.
[65] Yu R. Y., Reddy M., Pusl J., Allen S. T., Case M., Rodwell M. J. W.: Millimeter-wave on-wafer waveform and network measurements using active probes. *IEEE Transactions on Microwave Theory and Techniques*, 43(1995)4, s. 721–729.
[66] Ferrari P., Angénieux G.: Calibration of a time-domain network analyzer: a new approach. *IEEE Transactions on Instrumentation and Measurement*, 49(2000)1, s. 178–187.
[67] Student guide for basic network measurements using the HP 8510B network analyzer system. HP Course number HP 8510B + 24D, 1988.
[68] Hines M. E., Stinehelfer H. E.: Time-domain oscillographic microwave network analysis using frequency-domain data. *IEEE Transactions on Microwave Theory and Techniques*, 22(1974)3, s. 276–282.

Antenna measurements

Many characteristics of an antenna, such as its pattern, gain and polarization, can be calculated fairly accurately with reasonable effort. The performance of the antenna should, however, eventually be verified by measurements to find potential design or manufacturing errors. Often measurements are the only practical way to determine the characteristics of an antenna. They are also needed when adjusting the main components of an antenna on site.

Different antennas and their applications cover a very wide range, which is why antenna-measurement technology is fairly diverse [69 – 72]. The intended use of the antenna determines which of its characteristics need to be known, at what frequency band measurements must be made, and what are the accuracy requirements. Measurements may be of a routine nature, or prove to be the so difficult they are almost impossible. There is a variety of alternative measurement sites and methods, whose advantages depend e.g. on antenna size and frequency.

Near and far field, reciprocity, coordinate system

The radiated field of an antenna is characterized by a space and time-dependent amplitude, phase and direction. In the vicinity of the antenna at distances $r \leq \lambda/(2\pi)$ is the reactive near-field region, where the reactive field component is dominant. In the reactive field energy is transferred back and forth between electric and magnetic fields, i.e. it is not radiated away from the antenna.

Beyond the reactive near-field begins the radiating near-field region area in which the reactive component decreases rapidly with distance ($1/r^2$) until it is negligible. The radiating characteristics in this region depend on the distance, because the distances (in terms of wavelength) between different parts of the antenna and the observation point change at different rates when moving away from the antenna.

Generally, antennas are designed to operate across long distances, and therefore we want to measure the characteristics of the radiating field in the far-field region, where the field decreases inversely proportional to the distance ($1/r$) and resembles a plane wave at the location of the receiving antenna. In this case, pattern, gain and polarization do not depend on the distance. Figure 129 illustrates these different regions.

When measuring antennas we often take advantage of a property called reciprocity. Reciprocity means that the antenna characteristics are the same whether it is transmitting or receiving. Thus, an antenna can be tested as a receiving antenna, even if it is used as a transmit antenna, or vice versa. In the following it is assumed, unless otherwise stated, that the measured antenna acts as the receiving antenna, since this viewpoint is usually more intuitive.

In Fig. 130 the standard spherical coordinate system used in antenna measurements is presented. The measured antenna (antenna under test, AUT) is fixed at the origin. The angle θ is zero at the positive *z*-axis (north pole) and increases (southwards) until it reaches 180 degrees at the negative *z*-axis. The ϕ angle is obtained by projecting the vector indicating the direction onto the *xy*-plane, and ϕ increases (when looking from the direction of the positive *z*-axis) counterclockwise starting from the positive *x*-axis at 0 degrees moving up to 360 degrees. In this coordinate system we chose the orientation of the antenna usually based on its mechanical structure or the maximum in the directional antenna pattern and its polarization. If the *xy*-plane lies in the horizontal plane, the angle ϕ is called the azimuth angle and $90° – \theta$ the elevation angle.

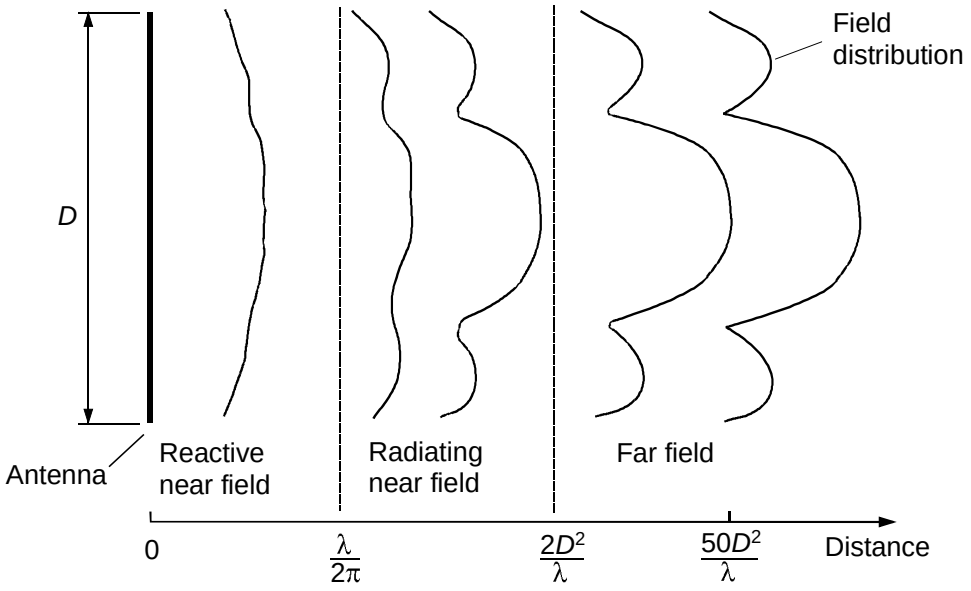

Figure 129. The near and far fields of an antenna.

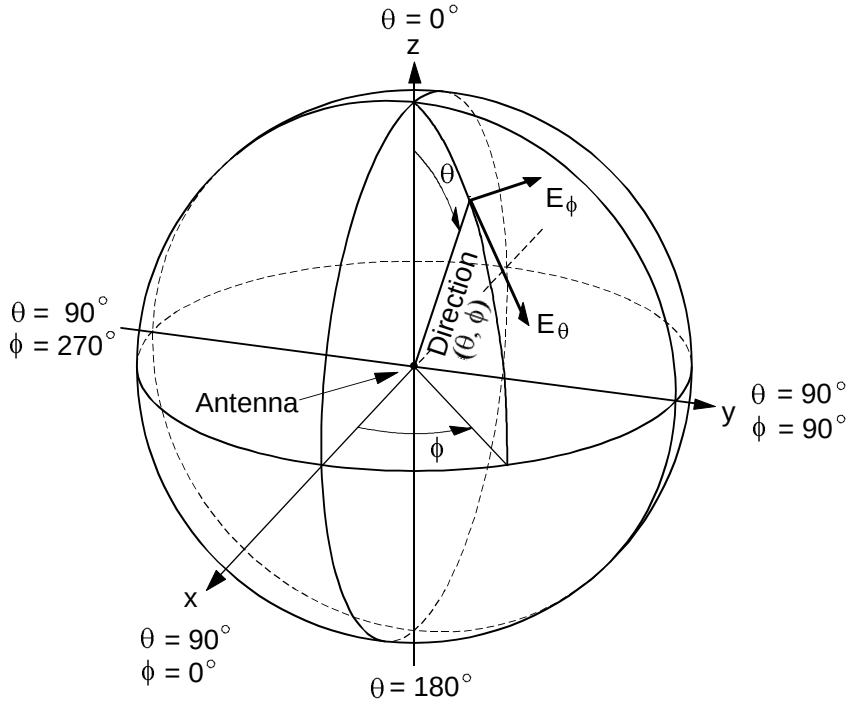

Figure 130. Standard spherical coordinate system used in antenna measurements.

Measured quantities

The radiation pattern of the antenna describes the electromagnetic field radiated by the antenna. Often, the radiation pattern means the directional pattern that describes the angular dependence of the radiated power density $P(\theta,\phi)$, or field strength $E(\theta,\phi)$. Normally we used the normalized directional pattern $P_n(\theta,\phi)$ or $E_n(\theta,\phi)$, where the power density or field strength in the direction of the maximum is defined as 0 dB. From the directional pattern e.g. the direction of the mainlobe, its half-power beam width, sidelobe levels and pattern-null directions, as well as the front-to-back ratio of the antenna can be investigated.

The directivity D is obtained from the directional pattern by integrating:

$$D = \frac{4\pi}{\iint_{4\pi} P_n(\theta,\phi)\, d\Omega},$$

where $d\Omega$ is the unit solid angle and $\iint_{4\pi}$ means integration over the entire solid angle 4π.

The antenna gain G ratio of the radiated power density in the mainlobe direction to the power density that would be obtained if a lossless isotropic antenna was radiating the same power that is coupled to the antenna input port. The gain and directivity of a lossless antenna are equal. Part of the power is, however, dissipated in conductive or dielectric losses. Due to these losses the gain is lower than the directivity:

$$G = \eta_r D,$$

where η_r is the radiation efficiency. Directivity and gain can also be presented as a function of the direction, $D(\theta,\phi)$ and $G(\theta,\phi)$. Those are obtained by multiplying the above defined D and G with the normalized directional pattern $P_n(\theta, \phi)$.

Antenna polarization describes the behavior of the radiated electric-field vector $\vec{E}$. $\vec{E}$ can be divided into two perpendicular components, which vary in a sinusoidal manner as a function of time, with a phase difference δ:

$$\vec{E}(t) = E_\theta \cos\omega t\, \vec{u}_\theta + E_\phi \cos(\omega t + \delta)\, \vec{u}_\phi,$$

where $\vec{u}_\theta$ and $\vec{u}_\phi$ are unit vectors. During each period the tip of the field vector describes an ellipse in the plane perpendicular to the direction of propagation (Fig. 131). The elliptical polarization is described by the axis ratio E_{max}/E_{min} and the handedness and tilt angle τ of the ellipse. Viewed in the direction of propagation the rotation is either clockwise (right-handed polarization) or counterclockwise (left-handed polarization). Linear and circular polarizations are special cases of elliptic polarization. Polarization is a function of the direction (θ, ϕ).

The antenna is normally intended to operate in a specific polarization, the co-polarization. The orthogonal polarization is called cross-polarization. If the co-polarization is, e.g., vertical linear polarization, the cross-polarization is horizontal linear polarization, whereas for right-handed polarization the cross-polarization is left-handed polarization. If the received wave is of the same polarization as the antenna polarization, we speak of matched polarizations. Otherwise, polarization mismatch is generated, which is described by polarization efficiency η_p.

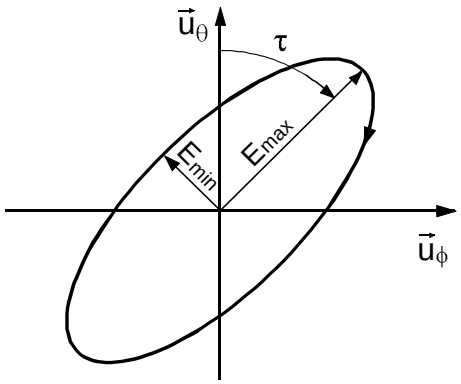

Figure 131. Polarisation ellipse.

The above quantities $P_n(\theta,\phi)$, $E_n(\theta,\phi)$, $D(\theta,\phi)$ and $G(\theta,\phi)$ can similarly be divided into co- and cross-polarization components. An ideal antenna does not create cross-polarization. The amount of cross- polarization of a real antenna depends on the direction (θ, ϕ), and has its mini-

mum usually in the direction of the mainlobe. Cross-polarization is relevant e.g. in applications where two different signals are transmitted at the same frequency at orthogonal polarizations.

The phase pattern $\psi(\theta, \phi)$ describes the deviation of the constant-phase front of the antenna from the ideal phase distribution of a spherical wave. The phase pattern is determined in relation to an appropriate reference point from which the virtual spherical wave is originating. The phase ψ is positive if the antenna phase is ahead of the spherical-wave phase, and negative if it is lagging behind the reference. The phase center of the antenna is that reference point, for which the phase difference between the constant-phase pattern in the mainlobe region of the antenna and the ideal spherical wave is as small as possible. The phase center is important to know, e.g., when positioning the feed antenna in the focal point of the reflector antenna.

The goal is to transfer the signal from the transmission line completely through the antenna to free space, or vice versa. In practice, the antenna impedance Z is somewhat different from the impedance of transmitter or receiver, so that part of the power is reflected at the mismatch. The reactive part of the impedance is created when energy is stored in the reactive near field of the antenna. The resistive part consists of the resistance caused by losses, and radiation resistance R_r. The input impedance of the antenna is also affected by reflections from the surrounding objects. For example, the mutual effect of near-by antenna-array elements changes the impedance of each antenna element compared to a situation where the elements are alone.

The antenna bandwidth can be defined as that frequency band, for which impedance matching, gain, beam width, cross-polarization or any other characteristic is within specified bounds.

Which antenna parameters are then measured depends on the application of the antenna. For some applications it is sufficient to measure the directional pattern in the mainlobe region, or even only the gain. An antenna for deployment on a communication satellite in a geostationary orbit will have to cover a specific geographical area beyond which it gives the lowest possible radiation exposure. In this case, we want to measure the directional pattern with all the sidelobes accurately. The cross-polarization of an antenna is important when two signals are transmitted at the same frequency with orthogonal polarization. The phase pattern of the feed horn of a reflector antenna may be needed for the design of the reflector surface or for calculating the far-field radiation pattern.

Error sources in antenna measurements

Typically, antennas are designed for long distances and unobstructed space, so that at the location of the receiving antenna we have a purely plane wave. In practice, this situation can be reached only approximately in antenna measurements due to the short measuring distance and effects of surrounding objects. Inherent error sources of antenna measurements are phase and amplitude variations of the field at the location of the measured antenna, reflections between transmit and receive antennas, coupling to the reactive field, reflections from surrounding objects, uncertainty in the alignment, atmospheric effects, non-idealities in the measuring system, and external interference signals.

If the measurement distance is short, the incoming spherical wave at the antenna under test differs significantly from a plane wave i.e. phase errors are generated. The distance of different parts of the receiving antenna (planar aperture) from the transmitting antenna is therefore varying. The distance where the phase error across the antenna surface is $\pi/8$, or 22.5° (path difference $\lambda/16$), is considered to be the boundary between near field and far field (see Fig. 132). If the antenna diameter is D, this limit is at the distance

$$r = \frac{2D^2}{\lambda}.$$

The gain decreases as a consequence of the phase error. At a distance of $2D^2/\lambda$ the gain is reduced by about 0.06 dB. Due to the phase error also sidelobes increase and pattern nulls are filled. Changes in the directional pattern are significant especially close to the mainlobe.

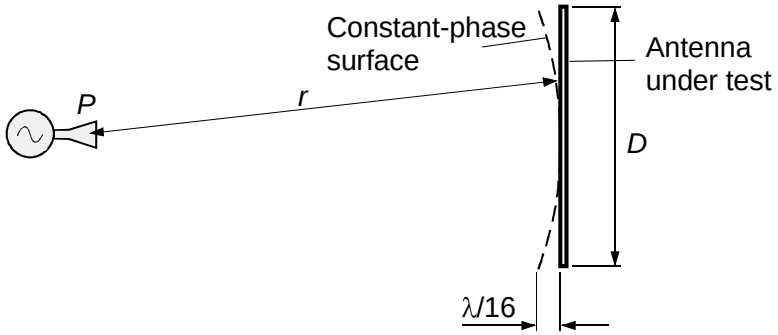

Figure 132. Definition of near- and far-field boundary.

Measurements should be carried out at a distance $r \geq 2D^2/\lambda$. Very accurate measurements may require a much longer distance, e.g. up to $r = 50D^2/\lambda$. When measuring reflector antennas the phase error caused by too short a measurement distance r can be reduced by changing the focus of the antenna from infinity to distance r by moving the feed location. The correction can also be done mathematically: from the measured far field we calculate the aperture field with help of the inverse Fourier transform, the phase correction is applied to the aperture field, and a Fourier transform yields the corrected far-field.

The amplitude of the incoming field should be constant in the region of the antenna under test. However, the directional antenna pattern of the transmit antenna produces a slightly lower power density at the edge of the receiving antenna than in its center (Fig. 133). In addition, the edge is slightly further away from the transmitting antenna. Normally a 0.25-dB decrease of the field amplitude from the center to the edge is acceptable. The error caused by the amplitude fluctuations depends on the aperture illumination function of the receiving antenna. For a typical aperture illumination, a 0.25 dB lower amplitude at the edge causes a decrease of about 0.1 dB in the measured gain as well as minor changes in the sidelobes. The field amplitude can be smoother using an antenna with a wider mainlobe for transmission, but then errors due to the reflections from the environment increase. Sometimes, one should use an antenna with narrower beamwidth to reduce reflections, at the expense of larger amplitude errors.

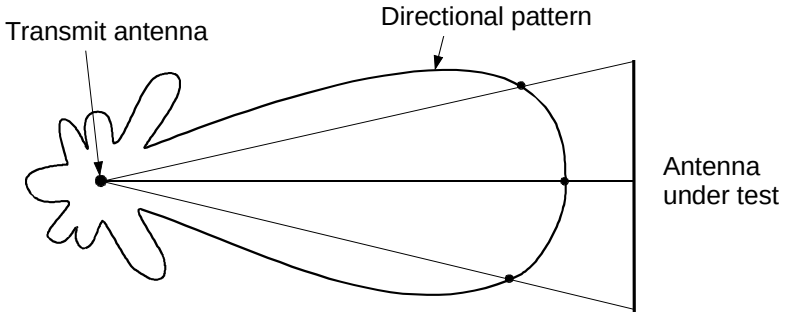

Figure 133. Amplitude error.

The field amplitude should be constant also in the radial direction, i.e. in the direction of wave propagation. In order to stay within a 1-dB variation, the measurement distance should be $r \geq 10L$, where L is the length of the antenna in the direction of the measurement (usually L varies with the rotation of the antenna).

If the measurement distance is short, the receiving antenna could couple to the reactive field of the transmitting antenna. The resulting error is negligible if the distance is more than 10 times

the wavelength. For example, at a distance of 10λ the reactive field of a short-dipole is 36 dB weaker than its radiated field.

Multiple reflections between antennas cause measurement errors. Part of the incoming field does not match the aperture field of the antenna and is therefore reflected back toward the transmitting antenna, which again reflects some part, and so on. This gives rise to a standing wave pattern between the antennas with a period of half wavelength. With increasing distance between the antennas the multiple reflections decrease rapidly.

Reflections from the environment are often the most significant source of error. Reflections may come from the ground, the walls of a measurement chamber, and antenna support structures. The reflectivity level R of a measuring range is the ratio of the field strengths of the reflected wave, E_r, and the directly transmitted wave, E_d, in decibels:

$$R = 20\log(E_r / E_d)\,.$$

In Fig. 134 the amplitude measurement error is presented when measuring at the level a in the directional pattern (in decibels), when the reflectivity level is R. Even a reflected wave with quite low power may cause a significant error, since the field vectors E_d and E_r sum up. For example, a reflected wave with a power density 20 dB (or 1%) lower than that of the direct wave, causes an error in the field strength of +0.83...–0.92 dB (or +21...–19%), depending on the phase difference. The corresponding phase error is ± 5.7°. Reflections cause errors especially in sidelobes, and the error grows with a decreasing level a of the directional pattern. Methods to reduce reflections are presented in the upcoming descriptions of the different measurement sites.

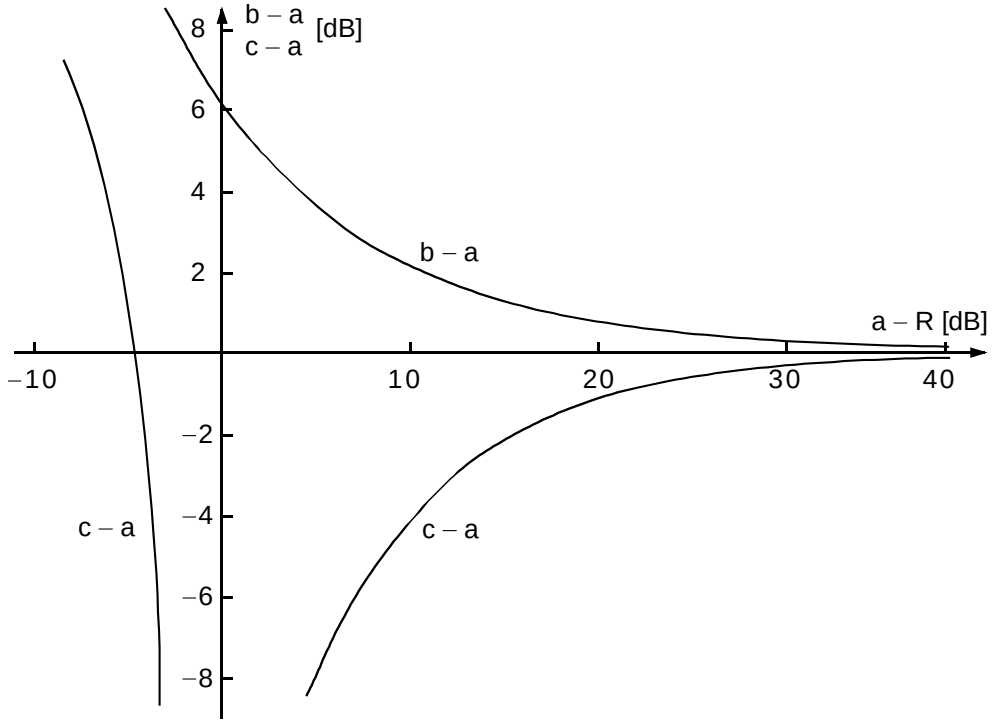

Figure 134. Measurement error limits as a function of the ratio between reflected and direct field strength; a is the level of the directional antenna pattern, R is the reflectivity level, b and c are the highest and lowest measured levels of the directional pattern.

Misalignment in the main lobe of the transmit antenna increases the amplitude variations in the measurement region. If the antenna under test has a narrow mainlobe, it must be accurately directed towards the transmitting antenna when adjusting the coordinate system. Otherwise, the directional 2-D pattern cuts do not go through the maximum of the mainlobe and the measured

gain is too small. A wrong tilt angle of the antenna under test causes polarization mismatch with the transmit antenna. The polarization efficiency η_p describes the power loss due to differences in the polarizations. For a difference $\Delta\tau$ between the angles of two linear polarizations,

$$\eta_p = \cos^2 \Delta\tau .$$

The measured cross-polarization is then

$$P_{nX}(\theta,\phi) = (1-\eta_p)P_n(\theta,\phi) = \sin^2(\Delta\tau)P_n(\theta,\phi) .$$

For example, a one-degree tilt-angle decreases the measured gain only by 0.0013 dB, but increased the cross-polarization from minus infinity to –35 dB relative to the mainlobe level.

Atmospheric attenuation is relevant in measurements at long distances and at high frequencies, especially in the millimeter-wave region near the resonances of oxygen and water vapor. Attenuation variations during the measurement can cause errors. The scintillation due to variation in the atmospheric refractive index is another source of error in the measurement of directional antennas.

Antenna measurement ranges

A variety of measurement ranges have been developed for measuring antennas, such as elevated range, ground-reflection range, anechoic chamber, compact range, and near-field measurement range. Which one is the most appropriate measurement range, depends e.g. on frequency and antenna size. If the antenna is to be used near a large structure affecting the radiation pattern, it cannot be tested in a standard measurement range, measurements should rather be carried out at the antenna site. The far-field distance of very large antennas can be so long that measurements cannot be carried out with ground-based equipment. In this case, measurements may be carried out with help of an airplane, a satellite, or a strong signal from an astronomical object.

Elevated ranges

An elevated range can be realized by placing the antennas on high hills or towers with a long enough distance to meet the requirements for the far field (Fig. 135). Usually this kind of measurement range is designed in a fairly flat landscape.

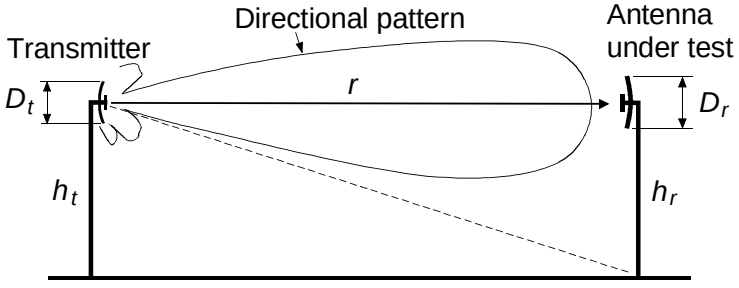

Figure 135. Elevated measurement range.

We can reduce the impact of the environmental by selecting a transmission antenna whose mainlobe does not illuminate the ground, removing any reflective objects on the ground near the line of sight, re-directing the reflected power of those objects that cannot be removed, or by using RF absorbers. The direct reflection from the surface is reduced with help of diffraction fences, which usually consist of metal nets. To reduce ground reflections one can measure half of the

radiation pattern at a time by turning the antenna in the elevation so that the mainlobe points towards the sky.

If the path lengths of direct and reflected wave are sufficiently different those reflections can be eliminated by transmitting short pulses and keeping the receiver open only when the measured signal arrives. The pulse length is typically tens of nanoseconds, and the gate circuits must be at least as fast. Using pulses reduces the dynamic range, since the average pulse power is reduced according to the pulse ratio, and pulses require a wider bandwidth, which increases the noise level. Another possibility is to sweep the transmitter frequency and calculate from the measured frequency response with the inverse Fourier transform. This gives the time-domain response, from which unwanted components can then be removed mathematically. This will be followed by a Fourier transform yielding the corrected frequency response.

In order to limit the amplitude errors to less than 0.25 dB at the surface of the antenna under test, the diameter of the transmitting antenna (assumed to be a parabolic antenna) should satisfy

$$D_t \leq 0{,}37\frac{\lambda r}{D_r},$$

where D_r is the diameter of the antenna under test. To avoid that the mainlobe of the transmitting antenna does not illuminate the surface between the antennas, it is recommended that the first null of the directional pattern is aimed at the base of the tower. This gives the second condition for the diameter of the transmitting antenna,

$$D_t \geq 1{,}5\frac{\lambda r}{h_r},$$

where h_r is the height above ground of the antenna under test. Combining these formulas provides the condition for the height:

$$h_r \geq 4D_r .$$

The slanted measurement range (Fig. 136) is a variation of the elevated range. The antenna under test is mounted on a tower and the transmit antenna is mounted close to the surface in such a manner that the first null of the directional pattern is pointing parallel to the ground.

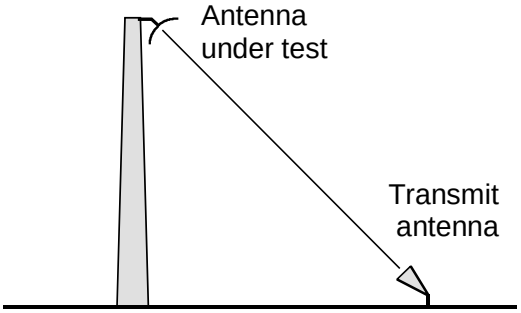

Figure 136. Slant measurement range.

Ground-reflection ranges

At lower frequencies, such as at the VHF range, eliminating reflections from surrounding objects is difficult, since a transmit antenna with sufficient directivity would be very large. In this case, it is easier to build a ground-reflection range instead of an elevated range.

In a ground-reflection range (Fig. 137) there is a smooth and conductive surface between the antennas, so that the direct and reflected signals form an interference pattern. We can consider the transmit antenna and its mirror image form an antenna array. If the reflection takes place in a

small oblique angle φ, the absolute value of reflection is almost 1. The phase of the reflection coefficient is about 180° for horizontal polarization and changes rapidly as a function of φ in the vertical polarization. The *rms* value of the surface roughness should be less than (λsinφ)/16. By placing the antennas at a suitable height, the first lobe of the interference pattern, which has at the same time the most even amplitude distribution, reaches exactly the antenna under test. This occurs when the height of the transmitting antenna is

$$h_t \approx \frac{\lambda r}{4h_r}.$$

When measuring at a different frequency this height always has to be adjusted. To keep the amplitude ripple near the measured antenna below 0.25 dB, the height h_r should be at least 3.3 times the diameter of the antenna D_r. It is generally recommended that the $h_r \geq 4D_r$.

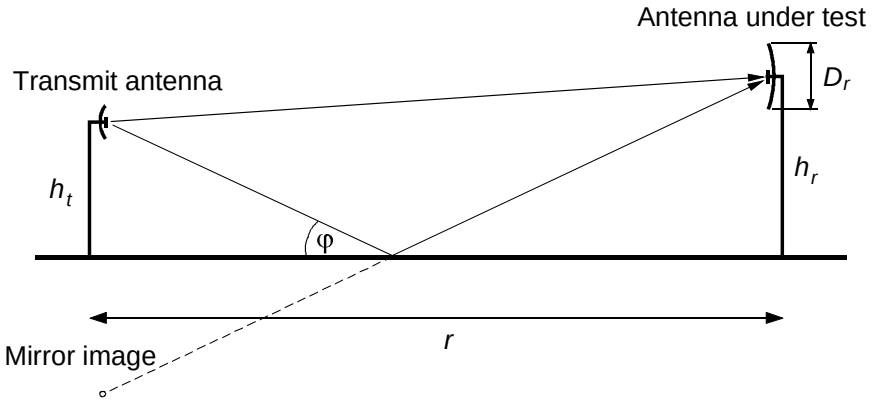

Figure 137. Ground-reflection measurement range.

Anechoic chambers

An anechoic chamber is a room completely lined with RF absorbing material, which simulates a free-space environment. The room is usually shielded against external disturbances by an additional metal layer under the absorbing material. Since it is an indoor range, measurements in an anechoic chamber can be carried out under controlled conditions regardless of the weather.

In Fig. 138 two basic anechoic chamber geometries are presented. The most critical parts of a rectangular room are the central regions of the walls, from which a reflected wave can reach the other antenna after only one reflection. These wall areas must be covered with the best quality absorbers. Also, support structures for the antennas should be protected well, so that no standing waves are created between them. Measurements are performed in the co-called quiet zone of the chamber, in which the reflections have been minimized. The length and width ratio of the room is about 2:1. In a too narrow room the incident angle on the walls is too large and with it the reflections. The transmitting antenna is selected so that the mainlobe does not illuminate the side walls. The frequency range of the chamber is mainly determined by the absorber thickness. The far-field requirement limits the size of the antenna under test. The measurement distance can be increased by opening the rear wall and combining the chamber with an elevated range.

The tapered chamber (Fig. 138(b)) has two operating modes. At low frequencies (≤ 1 GHz) the transmit antenna is located in the narrow tip of the room, in which case the reflections are generated very close to the transmitter. Hereby, the phase difference between direct and reflected wave is small and it also changes slowly within the measurement region. This will show a better performance than the rectangular chamber. However, at high frequencies the phase difference

becomes significant. Then, a more directive transmit antenna is used and moved somewhat towards the center of the room, in other words it is used like a rectangular chamber.

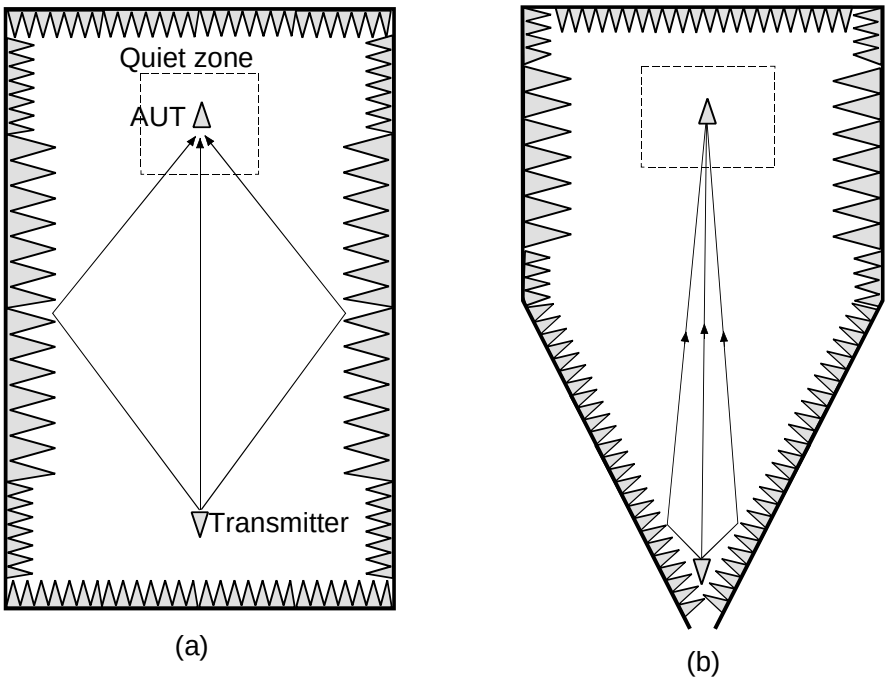

Figure 138. Anechoic chambers: (a) rectangular, and (b) tapered anechoic chamber.

Microwave absorbers are made of polyurethane foam, impregnated with carbon particles to increase the losses. They are typically in pyramid or wedge shape, so that possible reflections from the surfaces are not reflected directly back into free space. Also flat absorber is used. Their resistivity is gradually changing as a function of depth. If the waves reach the absorbers at a fairly low angle (large incident angle), wedge-shape absorbers should be used. The thickness of the absorbers determined the lowest operating frequency of the room: the thickness should be several wavelengths. The largest reflection at small incident angles is typically R = -50 dB in the microwave region. At millimeter-wave region reflections increase rapidly, so that at 200 GHz R can be -25 dB [73, 74].

Below 1 GHz polyurethane-based absorbers must be very thick. Therefore, at these frequencies lossy ferrite tiles are preferred, whose thickness is much smaller than the wavelength. The ferrite material is chosen so that its $\varepsilon_r \approx \mu_r$, so that there are no reflections at the air-to-ferrite interface.

In Finland, anechoic chambers exist e.g. at the VTT Technical Research Centre of Finland (dimensions of 12 m $\times$ 12 m $\times$ 17 m) and at the Department of Radio Science and Engineering of Aalto University (dimensions 6.3 m $\times$ 5.5 m $\times$ 8.7 m, i.e. width x height x length).

Compact antenna test ranges

The compact test range is an antenna-measurement site, which realizes a plane-wave-like field within a small space with help of e.g. reflectors. Compact test ranges are also widely used in measurements of the radar cross section.

In Fig. 139 a compact test range is shown that is based on a side-illuminated parabolic reflector arrangement, which transforms the spherical wave into a plane wave, i.e. the rays from the feed horn into parallel rays. Other arrangements use two cylindrical parabolic reflectors or a hyperbolic sub-reflector and parabolic main reflector. The plane wave can also be formed with help of a lens, an antenna array, or a hologram.

In a compact antenna test range measurements can be performed under controlled laboratory conditions in a space with the fraction of the size that would be needed for outdoor measurements. Also the transmit power needed is much smaller than in a free-space site. The measurements are performed in a quiet zone where the plane wave is good enough. The quiet zone diameter is about 1/3 of the diameter of the reflector when we consider a test range as shown in Fig. 139. Adding a hyperbolic sub-reflector yields a larger quiet zone, about 2/3 of the main reflector diameter.

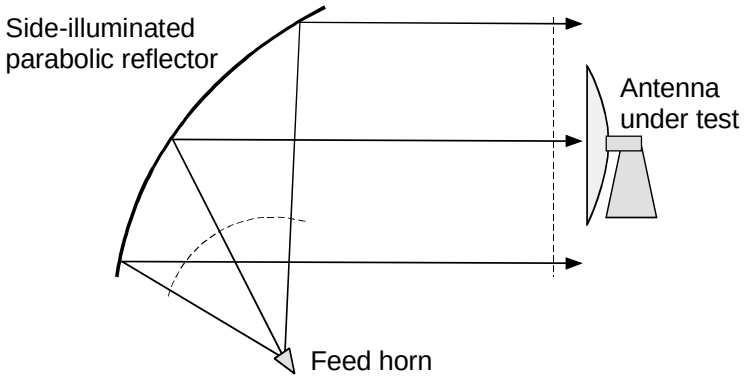

Figure 139. Compact antenna test range based on a side-illuminated parabolic reflector.

There are a number of reasons why the wave front is not ideally plane in the quiet zone: the amplitude and phase vary within and decreases towards the edge of the quiet zone. The sidelobes from the horn may reach the quiet zone directly, this is reduced by placing absorbers in between. Room reflections are reduced by placing absorbers at critical points along the walls, especially the wall behind the quiet zone is important. Multiple reflections between the antenna under test and the source antenna can be significant if the feed horn is not well matched. The effect of reflections can also be reduced by using short pulses or the Fourier-transform technique as described earlier. Often, the most important reason for unwanted radiation is the edge diffraction of the reflector. This diffraction is reduced by shaping the edges of the reflector in a curved or saw-tooth manner. Due to the directional pattern of the feed horn the field amplitude is smaller at the edges of the reflector than in its center and due to the phase pattern the phase front is not completely planar in the quiet zone. In the side-illuminated case the path differences affect the amplitude distribution in the quiet zone. In addition, a side-illuminated reflector will inevitably give rise to cross polarization. Cross polarization can be reduced by using a long a reflector with a focal length or two reflectors, which cancel each other's cross polarization components. The inaccuracy of the reflector surface limits the maximum measurement frequency. Surface roughness should be better than about $\lambda/100$. In this case, the plane wave has at most 7.2° phase error relative to the reflector.

Compact test ranges have been built up to a frequency of 200 GHz for measuring antennas up to 5 m in diameter [75]. In Fig. 140 a compact test range that fulfills these requirements is shown. The dimensions of the parabolic main reflector are 7.5 m $\times$ 6.0 m and the surface roughness is 13 μm (*rms*). The size of the hyperbolic sub-reflector is 5.6 m $\times$ 5.3 m, and its surface roughness is 18 μm. The length of the saw-teeth on the edges of the reflectors is 0.75 m.

Another way to create a plane wave is by use of a hologram, which is the interference pattern of a plane wave and a spherical wave [76]. The interference pattern can be realized with a suitable metallization of a dielectric sheet. The metallized sections correspond to zero amplitude and the clear sections to the amplitude 1. Such a hologram can be interpreted as an antenna array that is formed by a number of radiating slots. When illuminating the hologram with a spherical wave, part of the power passing through is diffracted into a plane wave in a certain direction, which depends on the manufacturing parameters of the hologram. A compact

test range for 650 GHZ has been built for measuring antennas up to 1.5 m in diameter. The diameter of the hologram is about 3 m [77].

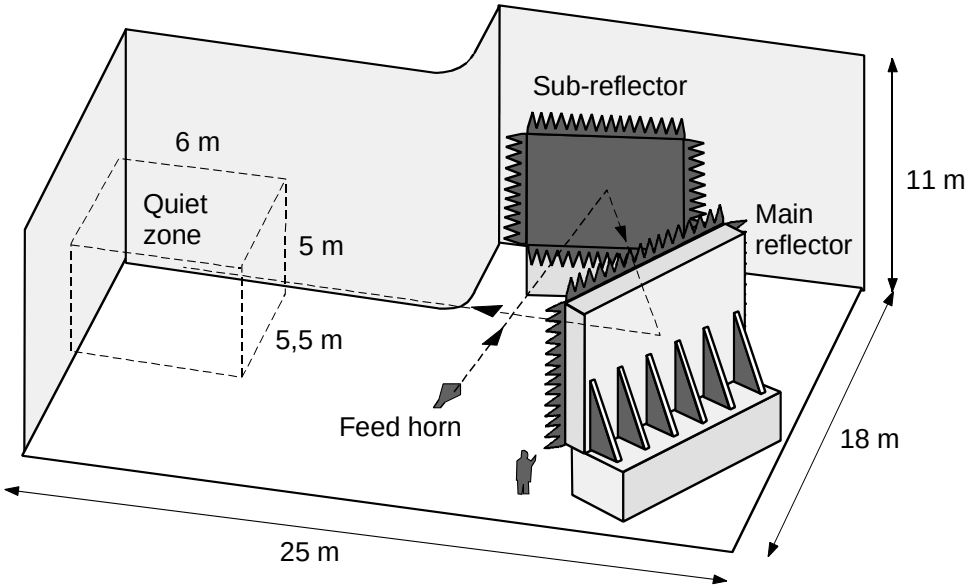

Figure 140. Compact antenna test range with two reflectors.

Near-field measurement

Often, the far-field distance of an antenna is impractically long for a measurement. In this case, one option is to first measure the near field of the antenna and then calculate the far-field radiation properties [78].

In near-field antenna measurements the amplitude and phase of the field generated by the antenna are measured in two orthogonal polarizations close to the surface of the antenna with a well-known, broad-beam probe antenna. The probe antenna should be at least a few wavelengths away from the surface, so that the reactive near-field is negligible.

Three most typical near-field scanning modes are shown in Fig. 141, cylindrical and spherical scanning around the antenna, or planar scanning only in front of the antenna. Measurements in the spherical coordinate system are suitable for omnidirectional antennas. Measurement can be done by turning the antenna under test, but the subsequent calculations are difficult. When measuring in cylindrical coordinates the antenna is turned in the azimuth direction and the probe antenna is moved vertically. The calculations are then easier. Measuring in the planar coordinate system is suitable for antennas, which radiate strongly in a single direction. In this case, the antenna under test is fixed. The calculations are relatively simple in the planar case: using the Fourier transform the so-called plane-wave spectra are calculated from the measured near fields and corrected for the effect of the probe antenna pattern; from the plane-wave spectra the far field radiation pattern can then be obtained.

Only with spherical scanning we cover the whole antenna perfectly, but in case of very directional antennas it is sufficient if the planar scanning range exceeds well beyond the edges of the aperture. The interval between measurement points should be $\lambda/2$ and the position accuracy better than $\lambda/100$. For an antenna with a rather wide mainlobe the required sampling interval can be $\lambda/4$, whereas for a highly directive antenna a sample interval of λ is sufficient.

Measurements should be performed in a room lined with RF absorbing material. Special attention should be paid to multiple reflections between the probe antenna, its positioning mechanisms and the antenna under test.

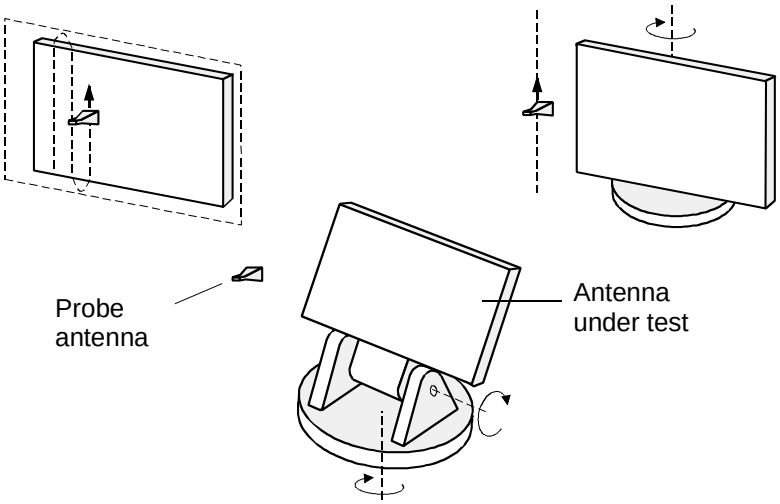

Figure 141. Planar, cylindrical, and spherical near-field scanning.

Near-field measurements at high frequencies are mechanically demanding due to the required positioning accuracy. In case of a large antenna the number of sampling points may be millions, and to perform a measurement within a reasonable time requires efficient technologies. Considering that measurements that can take hours, the electrical and mechanical stability of the equipment has to be very high. In addition, calculating the far field requires high-performance computers, because the near-field–to–far-field transformation is mathematically demanding.

On the up side, those tedious near-field measurements provide more information than traditional methods. Field amplitude and phase can be calculated at the antenna surface, in the far field, or anywhere in between, and the dynamic range is very wide, as probe antenna and antenna under test are close to each other. Near-field measurements are performed up to 500 GHz.

Testing of antenna-measurement ranges

The main non-ideality of an antenna measurement site is field fluctuations in the quiet zone. Due to the directional pattern of the transmit antenna, the field amplitude decreases towards the edges of the quiet zone, and reflections cause amplitude and phase ripples. The reflected field in the quiet zone consists of a large number of components whose resultant varies in a complicated manner as a function of location, frequency and polarization. The reflectivity level R also depends on the directional patterns of the used transmit and receive antennas. If the antennas have low directivity, the transmitting antenna illuminates strongly all surrounding structures and likewise the receiving antenna collects reflections from a wide area.

When investigating the impact of reflections, typically two methods are used: the free-space standing wave method (VSWR) and the antenna pattern comparison (APC) [79].

For the VSWR method the probe antenna is moved across the quiet zone in vertical or horizontal direction (Fig. 142). During the movement the phase difference between direct and reflected field changes, hereby creating an interference pattern that reminds of a standing-wave pattern. If the probe antenna has an as large beamwidth as possible and is directed towards the transmit antenna, approximately the total field in the quiet zone is measured. With a more directive probe antenna we can separate reflections from different directions. In this case, the probe antenna is directed to different angles, at different levels a of the directional pattern.

If the peak-to-peak value of the ripples is S(dB) = b–c, the reflectivity level R(dB) is

$$R = a + 20\log\frac{10^{S/20}-1}{10^{S/20}+1}.$$

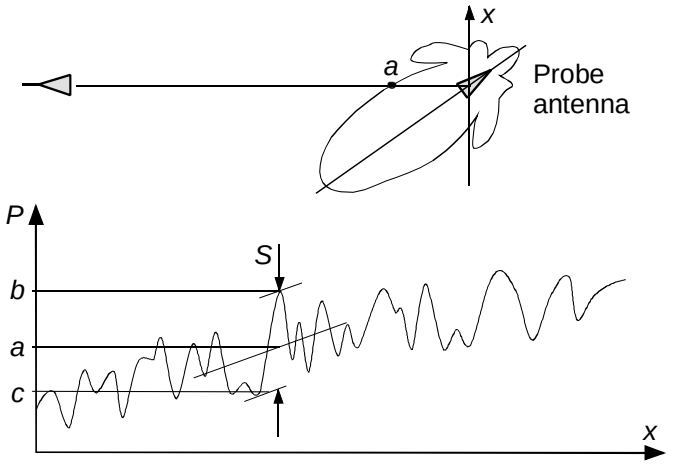

Figure 142. Measurement of the reflectivity level *R* with the VSWR-method.

This equation holds when E_d is larger than E_r. If there is one dominant reflection, the interference pattern clearly shows a period L_i. If the probe is moved in the same plane in which the direct and reflected signals propagate, the angle between these two signal paths is

$$\theta_r = \arcsin\frac{\lambda}{L_i}$$

If there are several reflections, they can be distinguished from the interference pattern with help of the Fourier analysis.

Example: A small horn antenna, which acts as the probe antenna, is turned towards the side wall so that the level of its directional pattern in the direction of the transmitting antenna is a = –20 dB. The peak-to-peak value of amplitude ripple is found to be S = 2.6 dB. Then, the reflectivity level is R = –20 dB – 17 dB = –37 dB.

In the APC method (Fig. 143) the pattern is measured in many different locations, with an interval of e.g. $\lambda/4$. Due to the random reflections the measured patterns differ. All measured directional patterns are plotted on top of each other, so that from the observed level of fluctuations and with help of Fig. 134 the reflectivity level *R* can be determined in different directions.

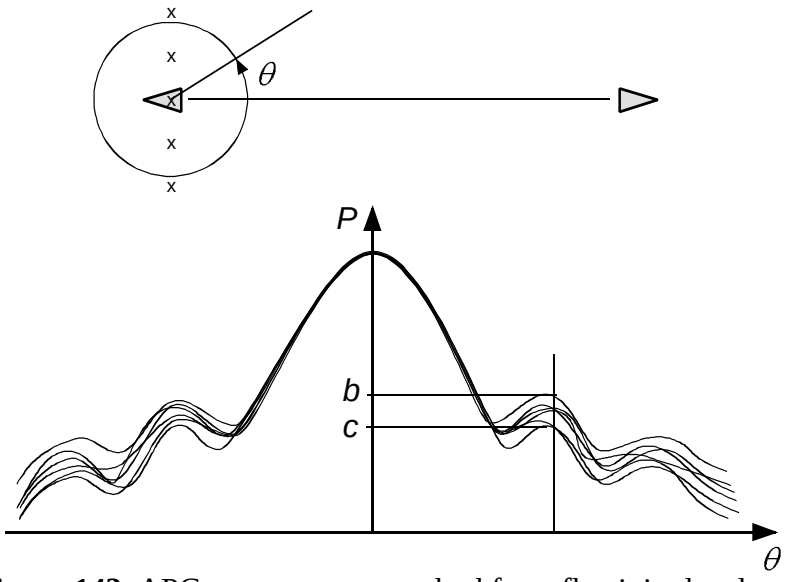

Figure 143. APC measurement method for reflectivity level.

The influence of the measurement range can be checked simply by measuring the antenna twice, but the second time turned upside down. The left and right halves of the measured directional pattern should then be exchanged, whereas the reflections due to reflections rather remain on the same side.

Antenna measurement-site instrumentation

In Fig. 144 we see the block diagram of a typical antenna measurement system. The extent of the equipment varies according to the need, from a simple system for measuring vertical and horizontal plane cuts of the directional pattern to a complex system that can automatically measure quite a complete set of antenna characteristics.

Automatizing the control of the equipment, the data collection and the post processing makes the antenna measurements more efficient. There may be so many measurement points as a function of frequency and angle and in two polarizations, that one simply needs an efficient measurement system. The computer can be programmed to analyze the results, e.g. to calculate the polarization properties and directivity, and the results can be presented in the desired format. Patterns can be presented in polar or Cartesian coordinate system. The polar plot is not very useful for directive antennas, since in a Cartesian plot more details can be presented. The measurement results can also be presented in a three-dimensional plot or as a contour map.

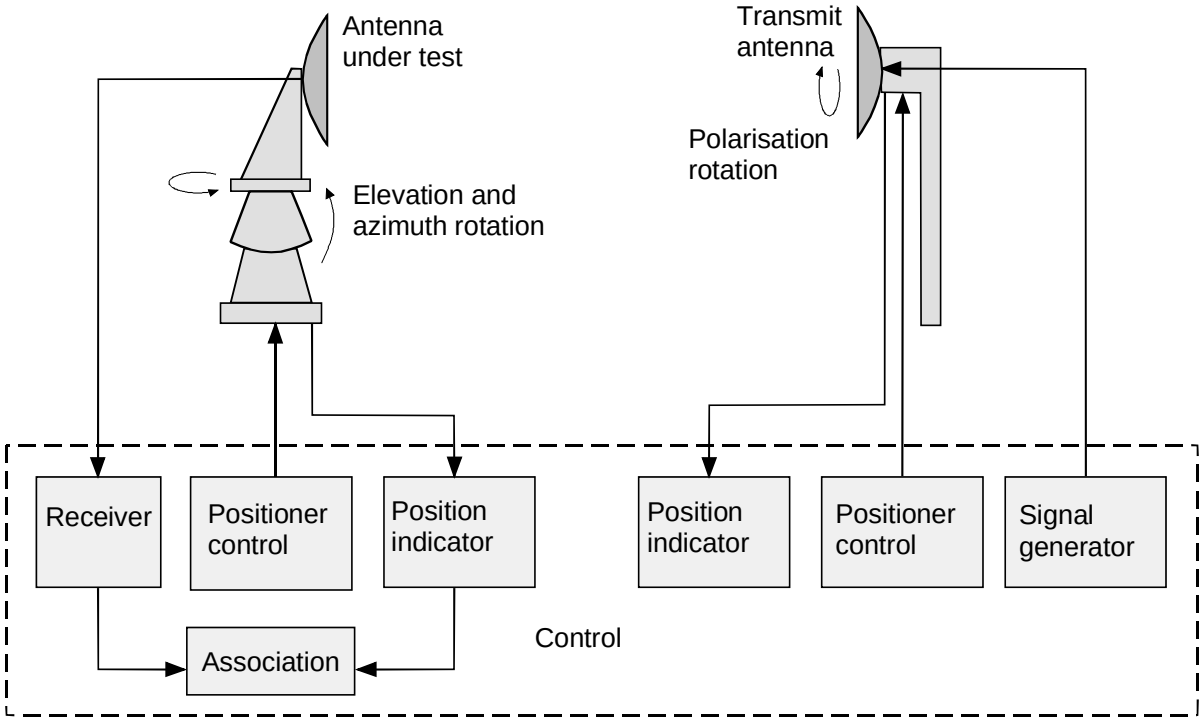

Figure 144. Block diagram of a typical measurement system forming an antenna range.

Transmitter and antennas

The output power of the signal generator needs to be sufficiently high, the frequency stable and the spectrum clean. For automatic measurements a remote-controlled frequency synthesizer is best suited, since it realizes a frequency change quickly and accurately. In this case, measurements can be made at many frequencies during a single antenna rotation.

For each frequency range transmit antennas are needed with a suitable beamwidth, low side lobes, and a pure linear polarization. Broadcast antennas are usually horn or parabolic antennas. When measuring the gain, a standard antenna with an accurately defined gain is needed.

Receivers

A high-quality receiver has a high sensitivity, operates linearly over a wide power range and is frequency selective. A diode detector or power meter can operate as a very simple receiver. However, the sensitivity is poor and the dynamic range is relatively small, e.g. the square-law operation range of a diode detector is about 40 dB. A much larger dynamic range, up more than 80 dB, is achieved by using a spectrum analyzer as a receiver. The shortcoming of these receivers is that the phase cannot be measured.

The most suitable receivers for antenna measurements are special superheterodyne receivers, or network analyzers. The block diagram of a phase-locked receiver is shown in Fig. 145. The signals from the antenna under test and the reference antenna are fed to the harmonic mixers. Their basic frequency band is 1–2 GHz and their operating band extends up to 18 GHz. The first intermediate frequency of 45 MHz is locked to a very stable crystal oscillator. The frequency changes in the transmitter are visible in a change of intermediate frequency in the reference channel, which initiates the phase-locked loop to tune the local oscillator accordingly until it is again locked. A second mixing stage uses a separate crystal oscillator whose frequency differs exactly by 1 kHz compared to the first intermediate frequency. Hereby, a 1-kHz output signal is obtained from the receiver, which can be fed directly to a plotter that is designed for this frequency and for plotting the directional pattern. Phase measurements can be made by

adding a reference channel to the receiver, and by comparing the 1-kHz output signals. The final measurement is made over a narrow frequency band, yielding high sensitivity (e.g. –130 dBm). A narrow bandwidth, however, limits the measurement speed. Measurement receivers can achieve over 80 dB dynamic range. The frequency range can be expanded by adding one more mixing stage between the antennas and the harmonic mixers.

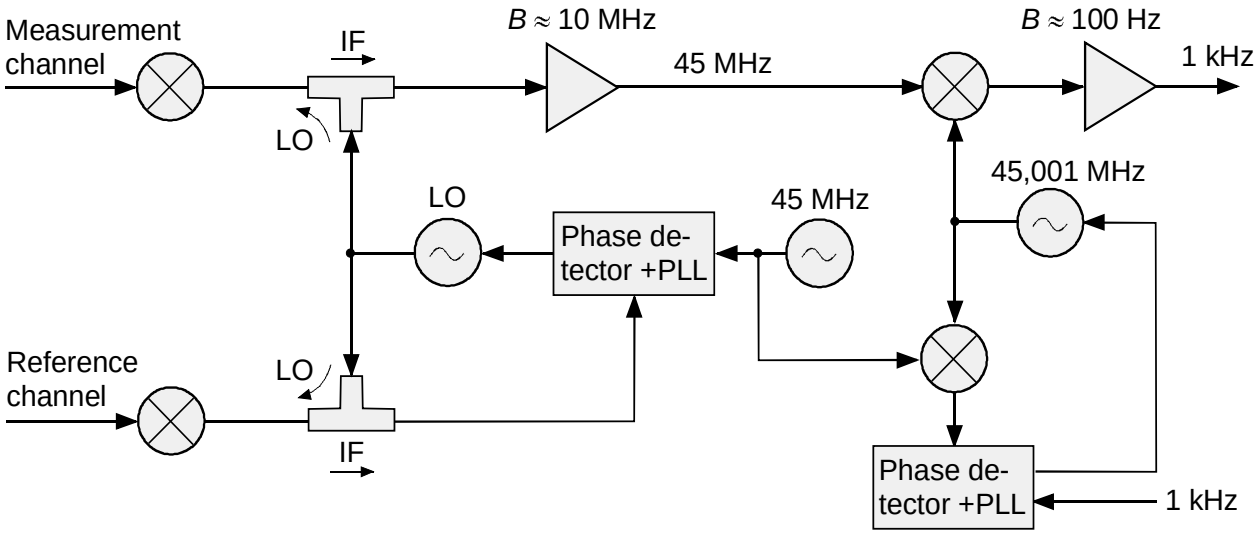

Figure 145. Phase-locked receiver for antenna measurements.

An antenna measurement system that is based on a vector network analyzer is illustrated in Fig. 146. The frequency synthesizer is used as signal generator. The signal received by the antenna under test and the reference signal received by the adjacent fixed antenna is fed to a frequency converter, which converts the signals to an intermediate frequency of 20 MHz. The intermediate-frequency signals are processed in the network analyzer, which gives the antenna pattern and the phase pattern. The dynamic range is better than 80 dB. The sensitivity can be improved by averaging up to 4 096 measurements, at the expense of the measurement speed. The frequency range of this measurement system is 45 MHz – 26.5 GHz, and it can be extended to 110 GHz by multiplying the input signal from the frequency synthesizer, and by replacing the converter with a mixer.

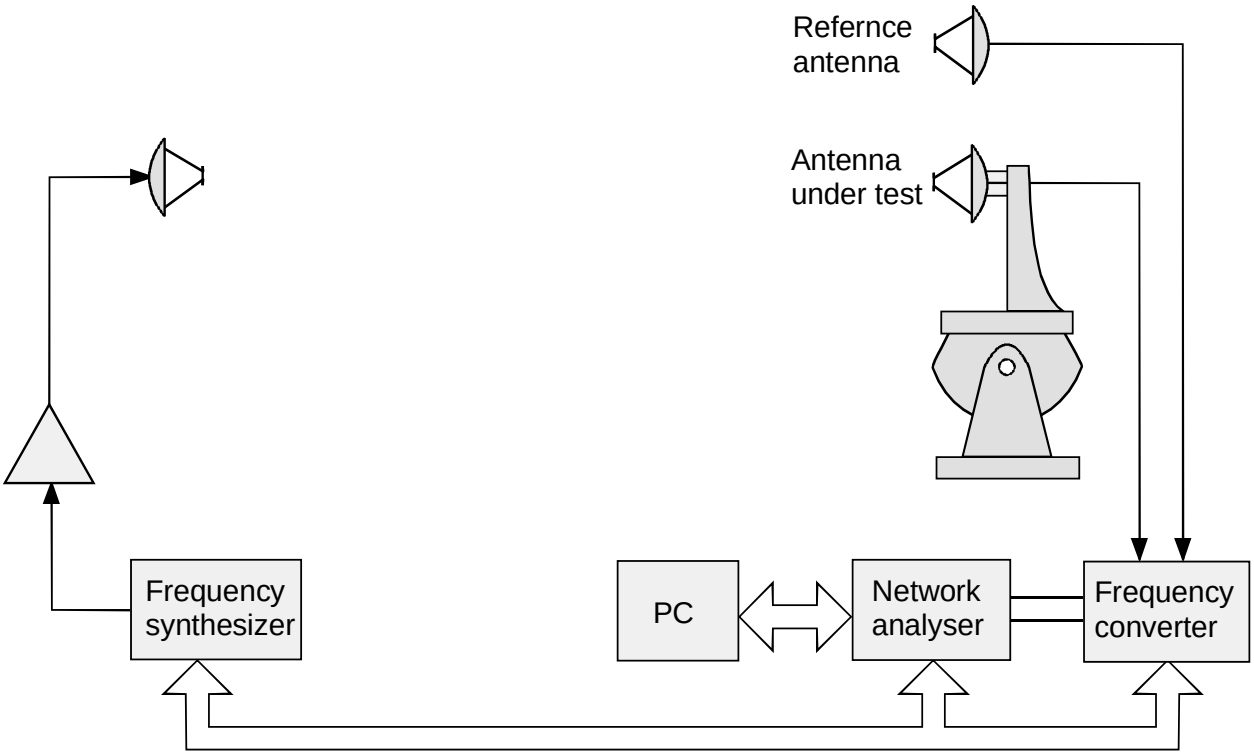

Figure 146. Antenna measurement system based on a network analyser.

Positioners

Typically, the radiation pattern is measured by turning the antenna under test while the transmitting antenna is fixed. In the positioning device illustrated in Fig. 147(a) the azimuth axis is above the elevation axis. This way we can measure in the standard spherical-coordinate system both ϕ-cuts (ϕ changes while θ is a constant) and θ-cuts (θ changes while ϕ is a constant) by turning the antenna around the according axis. Sometimes the elevation axis is above the azimuth axis as shown in Figure 147(b), but then the cuts are not according to the mentioned standard. The most versatile positioners offer both possibilities in such a way that the elevation axis is between two azimuth axes.

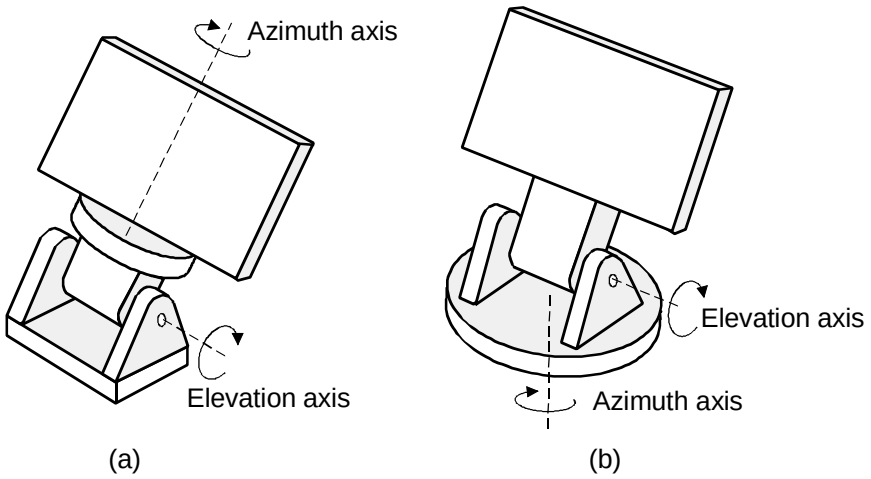

Figure 147. Antenna positioners: (a) azimuth-over-elevation, (b) elevation-over-azimuth.

Another possibility is to measure a fixed antenna under test with the transmit antenna moving around it so that the distance remains constant and the transmit antenna is always directed toward the receiving antenna. Such a measurement facility is shown in Fig. 148.

Also the transmitting antenna has to rotatable: it must point exactly towards the antenna being measured, but when measuring different polarizations it must be rotated around the axis that points in the direction of the mainlobe.

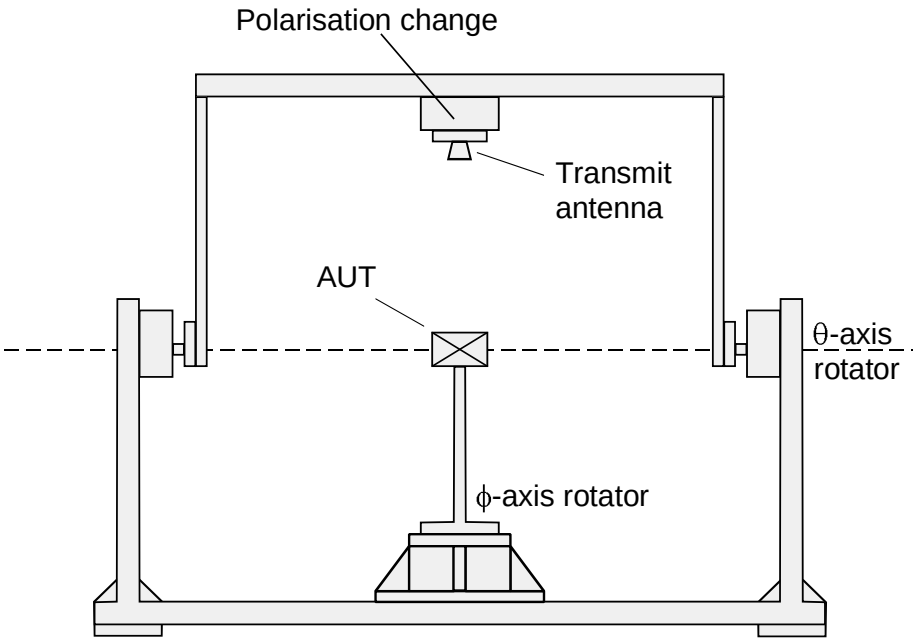

Figure 148. Positioner for moving the transmit antenna around the AUT at a constant distance.

Important mechanical properties of the positioner are weight capacity, torque and speed. In positioners generally stepping motors or DC motors are used. In a stepping motor the current angle is basically available through the number of control pulses fed to the motor. The disadvantage is that it requires a rather complex control unit. In order to turn a DC motor accurately and at a constant speed, angle sensors and a feedback system are needed. Angle sensor can e.g. be synchros and optical encoders. The typical angular accuracy of a synchro is 0.03°. The error caused by the play (free movement) of the positioner is typically 0.1°, but its effect can be eliminated by always turning the antenna in the same direction. The rotation speed of the device must be sufficiently slow so that the measurements are not distorted. The suitable rotation speed depends on receiver speed, number of frequency points, and desired angular resolution.

Directional pattern measurement

Directional patterns are often measured only in the principal planes of the antenna. These planes are perpendicular to each other, and both go through the maximum of the mainlobe. The principle planes of a linearly polarized antenna are typically the E-plane that is parallel to the electric field and the H-plane that is parallel to the magnetic field. For elliptically polarized antenna the principal planes can be defined according to the mechanical structure of the antenna. When measuring many different ϕ- and θ-cuts we can compose a three-dimensional presentation of the directional pattern.

When measuring the directional pattern $P(\theta, \phi)$ the polarization of the incoming wave must match the polarization of the antenna under test. E.g., a linearly polarized antenna is measured

using a linearly polarized transmit antenna in such a way that the field vectors of both antennas are parallel. In practice, co- and cross-polarized patterns are measured separately. When measuring the co-polarized pattern the antenna polarizations are aligned in the mainlobe direction, and for the cross-polarization the transmit antenna is then tilted by 90° (in case of linear polarization). Fine-tuning of the alignment can be done by maximizing the mainlobe power in the co-polarization measurement, and by minimizing the power in the mainlobe direction for the cross-polarization measurement. The polarization of the transmitting antenna must be as linear as possible in its mainlobe direction.

Phase pattern measurement

To measure the antenna phase pattern we needed a receiver that can measure phase differences, and accurate axial and linear positioners to move the antenna around the desired rotation axis and into the reference point.

In Fig. 149(a) the antenna under test is fixed and the phase of the AUT signal is compared to the signal phase of a receiving antenna that is moving along a circle around the AUT. In the method in 149(b) the antenna under test is rotated, and next to it is a fixed reference antenna. The phases of the signal received by the antennas are compared. The measured phase difference as a function of angle is the phase pattern $\psi(\theta, \phi)$.

In the so-called differential method, the antenna under test is transmitting and rotated as shown in Fig. 149(c). A fixed two-channel receiver measures the phase difference $\psi_d(\theta, \phi)$, from which the phase pattern $\psi(\theta, \phi)$ is calculated [80]. Measurements according to the differential method can be done with three power meters [81], or simply with a single power meter [82].

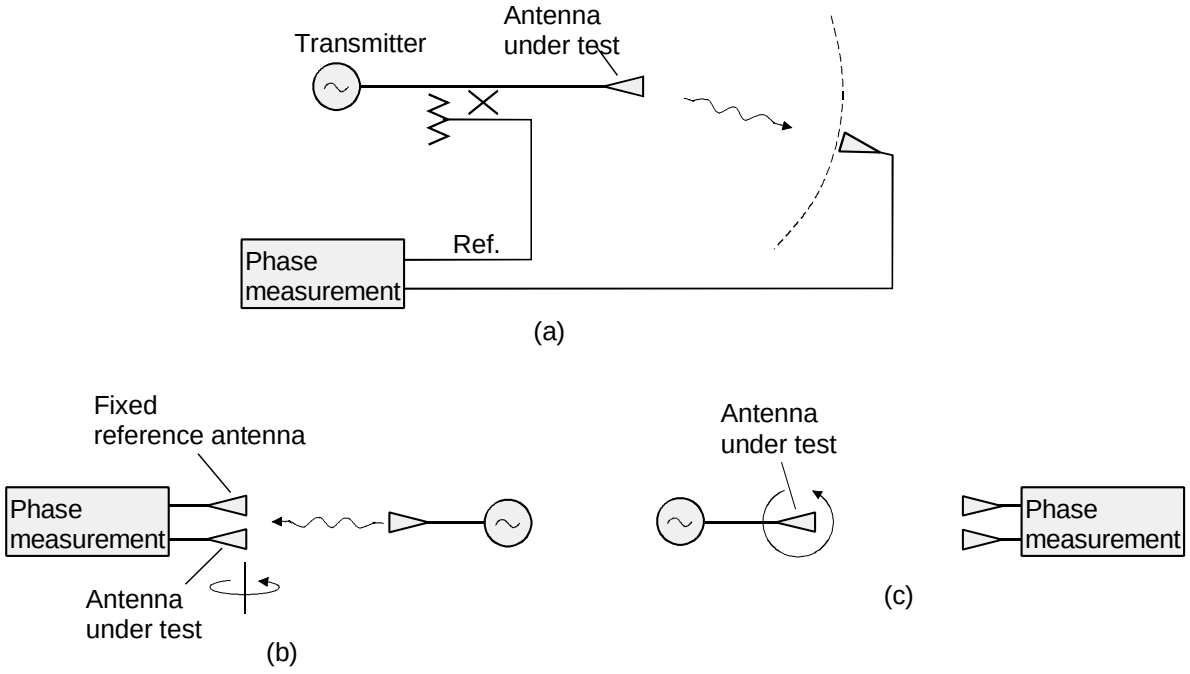

Figure 149. Phase pattern measurements (a) by comparing transmitted and received signals, (b) by using a reference antenna, (c) with a differential method.

The measured phase pattern depends on the location of the reference point. The phase pattern at some other point can be calculated geometrically. The location of the phase center is determined by moving the axis of rotation until the phase pattern is constant in the mainlobe direction, or it is calculated from the phase pattern.

Error sources in phase pattern measurement are those listed earlier and additionally differences in the electrical length of the two channels, phase errors from moving cables and rotary joints, uncertainty in the location of the reference point, inaccuracies in the positioner, and errors in the phase meter. The difference in electrical length is less relevant if the signal frequency is sufficiently stable. Phase errors from cables and rotary joints get more and more significant with increasing frequency. The differential method is especially suitable for millimeter-wave frequencies since those components are not needed. The alignment of the axis of rotating, e.g. in the aperture of a horn antenna can be done with an accuracy better than 0.1 mm with help of a video camera placed above the measurement setup [80].

Polarization measurement

The polarization of the wave radiated by the antenna can be measured in many ways. Often there is no need for a complete knowledge of this polarization, which simplifies the measurement. E.g., for an almost linearly polarized wave the handedness of rotation is not significant, just like for a nearly circularly polarized wave the tilt angle of the polarization ellipse.

The axis ratio and tilt angle of the polarization ellipse can be measured with a linearly polarized antenna. When the polarization plane is rotated, the received signal strength, i.e. the polarization pattern similar to the one shown in Fig. 150: The polarization pattern touches the ends of the major and minor axes of the polarization ellipse, the handedness cannot be determined.

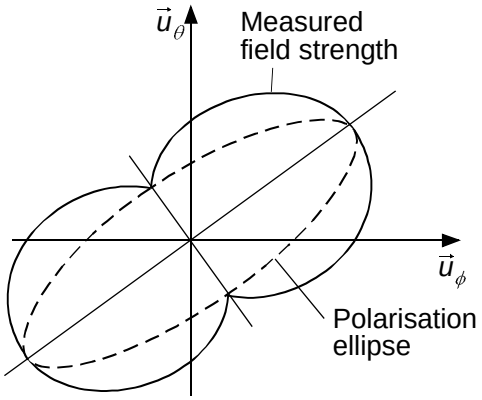

Figure 150. Polar plot of the field strength of an elliptical polarized wave, measured with a linearly polarized antenna.

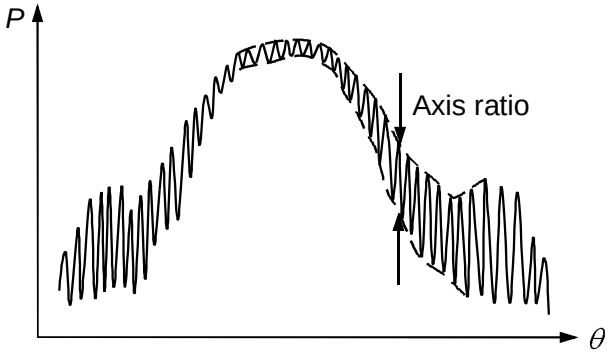

Figure 151. Directional pattern measured with a fast-rotating transmit antenna.

If the rotation speed of a linearly polarized antenna is large compared to the rotation speed of the AUT we obtain a pattern as shown in Fig. 151. From the envelope we obtain the axis ratio as a function of angle. The method is particularly suitable for testing circularly polarized antennas. The polarization can be determined in a complete manner by measuring the amplitude with four antennas, which have different and well-known polarizations. The gain of the antennas also needs to be known. The following measurement antennas are usually used for this method:
– a vertically or a horizontally linearly polarized antenna,
– a linearly polarized antenna at an angle of 45 or 135 degrees,
– a right-handed or a left-handed circularly polarized antenna.

Of these six options, four antennas are selected so that we have at least one antenna from each of the three orthogonal pairs. Often all six antennas are used in such a manner that the amplitude ratios of three antenna pairs are measured, from which the polarization is calculated.

The polarization can also be measured with a dual-channel receiver, when the channels measure two orthogonal polarizations. If the channel polarizations are vertically and horizontally polarized, the relevant factors of the electric field $\vec{E}(t) = E_\theta \cos\omega t\, \vec{u}_\theta + E_\phi \cos(\omega t + \delta)\, \vec{u}_\phi$, i.e. E_θ, E_ϕ ja δ, are obtained directly by measuring the amplitudes of the channels and phase difference. In this case the antenna under test must be transmitting, whereas in the previous methods it could also be the receiving antenna.

For all the above methods we first need antennas whose polarization is well-known. Linearly polarized antennas can usually be made with sufficient accuracy without the need to verify the polarization by measurements. Yet, an absolute polarization calibration can be done with the three-antenna method: three antennas are measured in pairs, for each pair two measurements are done by turning the receiving antenna 90°, then from the received complex amplitudes all three antenna polarizations can be calculated.

Gain measurement

To measure the (maximum) gain of an antenna one can use absolute and comparison measurements. Comparison measurements with a gain reference are more common. To calibrate the reference antenna, absolute measurements are used. The (high) gain of very large antennas can also be measured using an astronomical object as a reference.

Absolute gain measurements are based on the formula for wave propagation in free space, according to which the received power is

$$P_r = P_t G_A G_B \left(\frac{\lambda}{4\pi r} \right)^2 ,$$

where P_t is the radiated power of the transmit antenna, G_A and G_B are the gains of tranmit and receive antennas, and r is the distance between the antennas. The mainlobes of the antennas are directed exactly towards each other.

If there are two identical antennas ($G_A = G_B$) and we measure received power and transmit power, the gain can be calculated directly from the above equation. Another way is to measure three different antennas in pairs. From the measurements we get three equations from which the gains of all three antennas can be obtained.

Error sources in the absolute measurement method are polarization mismatch, reflections from surrounding objects, and a impedance mismatch between generator, antennas and power meter, resulting in a power-measurement uncertainty. The effect of a short measurement distance and amplitude ripples on the gain can be partly compensated: if the measurement distance is

$2D^2/\lambda$ and the mainlobe pattern of the transmit antenna is 0.25 dB lower at the edges of the antenna under test, about 0.15 dB has to be added to the measured gain.

For the comparison measurement (see Fig. 152) we need an antenna whose gain G_{ref} is known. From the power received by the unknown and the known antenna P_x and P_{ref} we get the gain of the unknown antenna as

$$G_x = \frac{P_x}{P_{ref}} G_{ref} .$$

The absolute power levels are thus not needed. The measurement of the power ratio with a high-precision attenuator is sufficient. In this case, the non-linearity of the receiver has no effect.

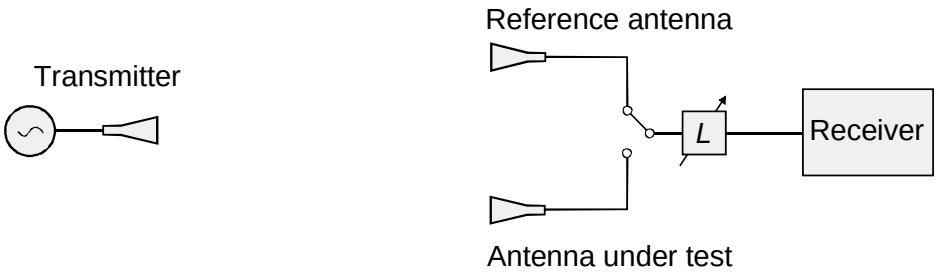

Figure 152. Comparison measurement of antenna gain using a gain reference.

As reference antennas dipole antennas are commonly used up to about 1 GHz, and pyramidal horns in the microwave and millimeter-wave region. The gain of these antennas can be calculated fairly accurately, and their construction is simple. The gain of a thin half-wave dipole is 2.15 dB. Commercial standard gain horns have gains of 14 – 25 dB and an accuracy of $\pm$ 0.25 dB. Pyramidal horns have a slightly elliptical polarization: the axis ratio is about 40 dB.

Error sources in comparison measurements are differences between the AUT and the reference antenna in terms of their polarization-efficiency and losses due to impedance mismatch. If the antennas are of a similar type, the decrease in gain due to the short measurement distance affects both antennas by almost the same amount, so the error is not significant.

The gain of elliptically or circularly polarized antennas is obtained by measuring the gains separately for two orthogonal linear polarizations, and summing up the gains. E.g., if the gain was measured in θ- and ϕ-directions, we get

$$G = G_\theta + G_\phi .$$

When determining the gain of very large antennas, the above described measurements are not possible since the far-field requirement cannot be fulfilled. These can then be measured with help of satellites or stars. There are a number of astronomical radio sources whose flux density S is known quite accurately as a function of frequency. The method requires a radiometer for measuring the antenna temperature T_a caused by the radio source. If the radio source is point-like, the gain is

$$G = \frac{8\pi k T_a}{LS\lambda^2}$$

where k is the Boltzmann constant and L is the atmospheric attenuation. For distributed radio sources the effect of its size must be taken into account with a correction factor. Because of

gravitationally induced distortions in the surface of any very large antenna, the gain should be measured as a function of elevation (the angle above the horizon).

Impedance measurement

The measurement of the input impedance of an antenna does not differ substantially from that of other circuits, for the impedance measurement we can use a slotted line, an impedance bridge, or a network analyzer. An impedance measurement is a measurement of the complex reflection coefficient, and therefore reflections from surroundings of the antenna will cause errors. Reflections caused by the antenna support structure are not error signals. The antenna under test is placed inside realistic operating conditions (including the support structure), with usually means a free-space environment. An anechoic chamber is a suitable measurement site. In impedance measurements the far-field condition does not exist, so one can place absorbers right in front of the antenna aperture. The effect of reflections can be reduced if the measurement is repeated in slightly different locations and the results are averaged. With omnidirectional antennas, reflections from all measuring equipment must be eliminated carefully.

In antenna arrays there exist interactions between the antenna elements. The impedance of an individual element can be measured when the other elements are fed according to the real operation. The active impedance depends on the phasing of the elements, which in a phase-controlled antenna array changes during operation. Another possibility is to generate an impedance matrix, from which the active impedance can be calculated. The matrix consists of the mutual impedances Z_{mn} and each element's impedance Z_{mm}. The mutual impedance between two elements is $Z_{mn} = V_m / I_n$, where I_n is the current in the reference point of element *n*, and V_m the causative voltage in the reference point of element *m*, when all others elements (except for element *n)* have an open termination in their reference points. Reference points can be selected freely, but they must be defined. The mutual impedance between two elements is obtained by measuring each element's impedance Z_{11} and Z_{22}, and the impedance Z_s, of element 1 while element 2 is short-circuited. The mutual impedance is then $Z_{12} = \sqrt{Z_{22}(Z_{11} - Z_s)}$.

[69] *IEEE Standard Test Procedures for Antennas*. IEEE Std 149-1979.

[70] Evans G. E.: *Antenna Measurement Techniques*. Boston–London 1990, Artech House.

[71] Appel-Hansen J., Gillespie E. S., Hickman T. G., Dyson J. D.: Antenna measurements. Chapter 8 of *The Handbook of Antenna Design*, vol. 1 (Rudge A. W., Milne K., Olver A. D., Knight P.). London 1982, Peter Peregrinus Ltd., s. 584–694.

[72] Kummer W. H., Gillespie E. S.: Antenna measurements – 1978. *Proceedings of the IEEE*, 66(1978)4, s. 483–507.

[73] Lehto A., Tuovinen J., Räisänen A.: Reflectivity of absorbers in 100–200 GHz range. *Electronics Letters*, 27(1991)19, s. 1699–1700.

[74] Lehto A., Tuovinen J., Räisänen A.: Reflections in anechoic chambers in 100–200 GHz range. *Electronics Letters*, 27(1991)19, s. 1708–1709.

[75] Steiner H.-J., Kämpfer N.: A new test facility for EM field measuring of large antenna systems up to 200 GHz. *Proceedings of the 19th European Microwave Conference*, 1989, Microwave Exhibitions and Publishers Ltd., s. 480–485.

[76] Ala-Laurinaho J., Hirvonen T., Piironen P., Lehto A., Tuovinen J., Räisänen A., Frisk U.: Measurement of the Odin telescope at 119 GHz with a hologram-type CATR. *IEEE Transactions on Antennas and Propagation*, 49(2001)9, s. 1264–1270.

[77] A.V. Räisänen et al, Measurement of a high-gain antenna at 650 GHz in a hologram-based CATR, proceedings of the 18th International Symposium on Space Terahertz Technology, 2007, California, US, p. 211-215.
[78] Yaghjian A. D.: An overview of near-field antenna measurements. *IEEE Transactions on Antennas and Propagation*, 34(1986)1, s. 30–45.
[79] Appel-Hansen J.: Reflectivity level of radio anechoic chambers. *IEEE Transactions on Antennas and Propagation*, 21(1973)4, s. 490–498.
[80] Lehto A., Tuovinen J., Räisänen A.: Measurement of amplitude and phase patterns of 110 GHz and 183 GHz corrugated horns. Espoo 1990, Helsinki University of Technology, Radio Laboratory, Report S 184.
[81] Lehto A., Tuovinen J., Boric O., Räisänen A.: Accurate millimeter wave antenna phase pattern measurements using the differential phase method with three power meters. *IEEE Transactions on Antennas and Propagation*, 40(1992)7, s. 851–853.
[82] Mallat J., Lehto A., Tuovinen J.: Antenna phase pattern measurements at millimeter wave frequencies using the differential phase method with only one power meter. *International Journal of Infrared and Millimeter Waves*, 15(1994)9, s. 1497–1506.

Measurements of material properties

The permittivity ε and permeability μ of a material affect the propagation of microwaves within that medium. Material parameters must be known in the design of most microwave circuits and components. The electromagnetic material parameters correlate with other physical properties of the material. Therefore, we can obtain information, for example, on moisture, density, temperature, and structure of the material.

The main features of a useful material measurement method are: the method does not destroy the material being measured, it does not require material samples in a specific, exact shape, it measures the complete sample volume, and does this in real time. Measurement methods can be divided into passive and active ones, i.e., we use with radiometric methods to measure the thermal noise emitted by the substance, or we measure the interaction of radio waves and matter, namely how the waves are reflected by the material or pass through the sample. Radar is an example of active methods. This chapter is limited to the measurement of the dielectric constant with resonator, waveguide and free-space methods. Many other methods are described in literature [83–86].

At Aalto University Department of Radio Science and Technology the measurements of material properties have been researched and developed over many decades. The results have been utilized, for example, in the development of RF and microwave sensors for automated industrial processes, and in solving difficulties with materials in the millimeter wave region [87-94].

Basics of material measurements

In a medium the wave length and attenuation as well as the wave impedance in the medium depend on the permittivity $\varepsilon = \varepsilon_r\varepsilon_0$ and permeability $\mu = \mu_r\mu_0$ of the substance. These are complex parameters, i.e. they can be divided into real and imaginary parts:

$$\varepsilon_r = \varepsilon_r' - j\varepsilon_r''$$
$$\mu_r = \mu_r' - j\mu_r''$$

If a plane wave propagates in the direction of the the z-axis, the field can be written as $E = E_0 e^{-\gamma z}$, where γ is the propagation constant $\gamma = \alpha + j\beta = j\omega\sqrt{\mu\varepsilon}$. (Here the possible conduction losses of free charges is included in the imaginary part ε_r''.) The phase constant $\beta = 2\pi/\lambda$ determines the wavelength and the attenuation coefficient α the field attenuation. The wave impedance in the medium is $\eta = \sqrt{\mu/\varepsilon}$ and it determines the reflections from the interfaces between the different media.

If the medium contains bound charges that an external electric field can move, the dielectric constant $\varepsilon_r > 1$ and the material is called dielectric. Moving of bound charges can mean the movement of electrons relative to the atom nucleus, a movement of atoms is a crystal or a molecule, or the turning of polarized molecules. (We speak of free charges when free electrons or ions move through the material.) For example, the turning of polar molecules is slowed down by friction, i.e. collision with other molecules, which results in the imaginary part of the dielectric constant. The relative permeability of a magnetic medium $\mu_r > 1$. Since most materials are not magnetic ($\mu_r = 1$), we will only handle dielectric materials in the following. But based on the duality of ε and μ the results could be straightforwardly applied also to magnetic materials.

Dielectric constant depends on many physical properties of the material and it is also a function of frequency. Often, material measurements are based on a well-known connection between ε_r and physical properties. For example, dielectric constant indicates humidity with a fairly large sensitivity, because water is a polar molecule and the ε_r of water at microwave frequencies is large, about 80. Hereby, we can examine the quality for example of paper, chipboard or food. Sometimes one needs ε_r itself because it is important in the design of microstrip circuits or quasi-optical components such as lenses. To minimize losses, it is important to know especially ε_r''.

Resonator method

A resonator partially or completely filled with a material is suitable for measuring the dielectric constant. The material sample changes the resonant frequency and loaded quality factor of the cavity. From this change ε_r' and ε_r'' of the material is calculated. The method is very accurate and also suitable for measuring low-loss materials. The disadvantage is that ε_r can be obtained only at the resonance frequency of the resonator.

The resonance condition requires that the resonator has a certain length in terms of the wavelength, for example, $\lambda/2$, or $\lambda/4$. The wavelength in a dielectric medium is

$$\lambda = \frac{\lambda_0}{\mathrm{Re}\sqrt{\varepsilon_r}}$$

where λ_0 is the wavelength in vacuum. If the resonator is completely filled, from the resonance condition follows

$$\mathrm{Re}\sqrt{\varepsilon_r} = \frac{f_{r0}}{f_r}$$

where f_{r0} is the resonance frequency of the empty resonator, and f_r the resonance frequency of the filled resonator. If $\varepsilon_r' >> \varepsilon_r''$, as is often the case,

$$\varepsilon_r' \approx \left(\frac{f_{r0}}{f_r}\right)^2$$

The quality factor Q of the resonator is defined as

$$Q = \frac{2\pi \times \text{energy stored in resonator}}{\text{energy dissipated in one period}}$$

The energy that left the resonator, i.e. the loss, can be divided into a number of factors causing losses. Since losses are inversely proportional to the quality factor, one can write

$$\frac{1}{Q_L} = \frac{1}{Q_e} + \frac{1}{Q_m} + \frac{1}{Q_r} + \frac{1}{Q_d} = \frac{1}{Q_e} + \frac{1}{Q_0}.$$

The loaded quality factor Q_L describes the total losses, the external quality factor Q_e the power dissipated in the load, Q_m the conductive losses in the resonator walls, Q_r the radiation losses, and Q_d the dielectric losses in the medium filling the resonator. The unloaded quality factor Q_0 includes all losses apart from external losses in the load.

Q_L can be obtained by a reflection or a transmission measurement. If we couple to the resonator and measure the reflected power around the resonant frequency, we could get e.g. the curve shown in Fig. 153. If the resonator has two coupling ports, the power passing through the resonator as a function of frequency would look like in Fig. 154. In both cases, Q_L can be obtained from the half-power bandwidth B_{-3dB}:

$$Q_L = \frac{f_r}{B_{-3dB}}$$

In the reflection-coefficient measurement B_{-3dB} is obtained from those frequencies where the power absorbed in the cavity $(1-|\rho|^2)P_i$ is half of the maximum value, and in the transmission-coefficient measurement from those frequencies, where the transmitted power $|\tau|^2 P_i$ is half of the maximum value.

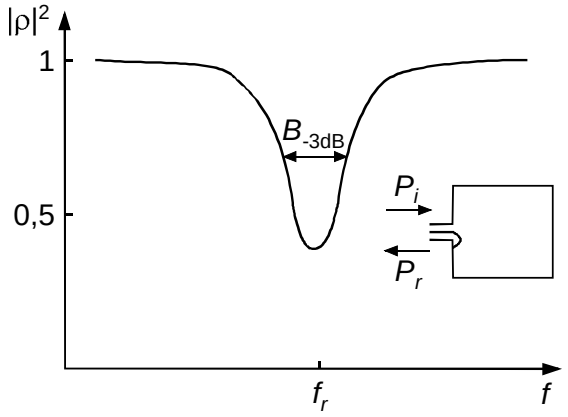

Figure 153. Reflection-coefficient measurement of a resonator.

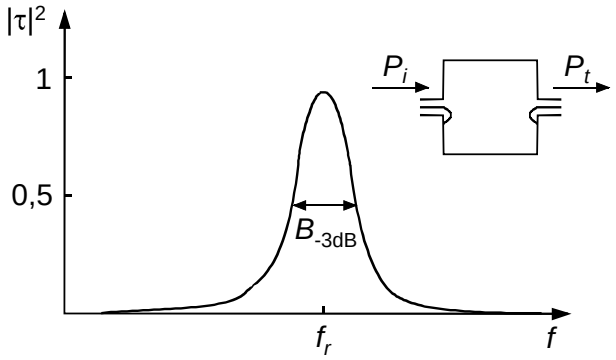

Figure 154. Transmission-coefficient measurement of a resonator.

For reflection and transmission measurements we can use a scalar or vector network analyzer. This method requires a good frequency resolution, so we need a frequency counter or a synthesized signal generator.

The unloaded quality factor can be calculated from the reflection-coefficient measurement

$$Q_0 \approx \frac{2Q_L}{1 \pm |\rho_r|},$$

where ρ_r is the voltage reflection coefficient at the resonance frequency. The plus sign applies to the so-called undercoupled case ($Q_e > Q_0$), the minus sign to the overcoupled case ($Q_e < Q_0$). For the transmission measurement, provided that all couplings are identical,

$$Q_0 \approx \frac{Q_L}{1-|\tau_r|},$$

where τ_r is the voltage transmission coefficient at the resonance frequency.

The quality factor Q_r describing the radiation losses is obtained by measuring two otherwise identical resonators made of different metals [83]. The conductivity of the metals must be different and well-known. The radiation losses of the two resonators are equal and the metal losses are proportional to $1/\sqrt{\sigma}$ (σ is the conductivity). The pair of equation obtained from two measurements can be solved for Q_r.

In case of a closed resonator $Q_r \approx \infty$. In this case, unloaded quality factor of the empty resonator is the quality factor of the metal losses Q_{m0} at the frequency f_{r0}. The filled resonator Q_m is obtained by taking into account that this figure is proportional to $\sqrt{f}$.

It can be shown that for a completely filled resonator

$$Q_d = \frac{\varepsilon_r'}{\varepsilon_r''}.$$

The imaginary parts of the dielectric constant is now

$$\varepsilon_r'' = \varepsilon_r'\left(\frac{1}{Q_0} - \frac{1}{Q_r} - \frac{1}{Q_{m0}}\sqrt{\frac{f_{r0}}{f_r}}\right).$$

It is often more practical to only partly fill the resonator. In this case, the dependence of resonator parameters on material properties is more complex. Sometimes this connection can be calculated. If the sample is small and causes a small frequency change ($<0.001 \cdot f_r$), the so-called perturbation method is used in the calculations [95]. If the calculation is too complex a calibration can be done using materials with well-known ε_r.

Various resonator structures are suitable for material measurements. A short-circuit coaxial line (Fig. 155(a)) or a closed piece of circular waveguide (Fig. 155(b)) are popular resonators. The open quasi-optical resonator in Fig. 155(c) is suitable for millimeter-wave measurements. It has a very high quality factor Q_0 ($\geq 10^5$), which enables measurements of small losses and real parts of the dielectric constant close to unity. The stripline resonator in Fig. 155(d) can be used for industrial process monitoring, or for measuring substrates for microstrip lines. A pair of wire-resonators as shown in Fig. 155(e) is used for rapid measurements of soft or liquid substances.

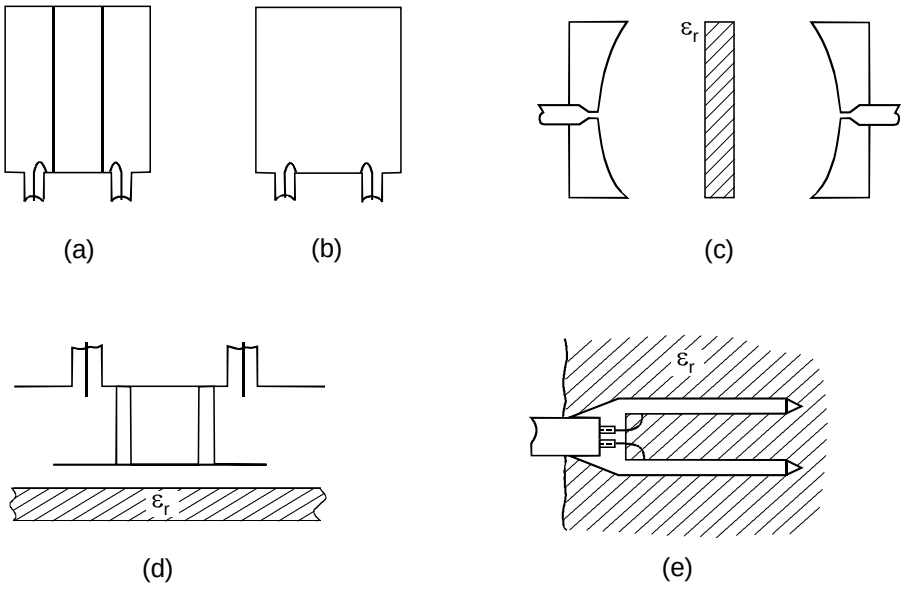

Figure 155. Examples of measurement resonators: (a) coaxial resonator, (b) cylindrical resonator, (c) quasi-optical resonator, (d) stripline resonator, (e) two-wire line resonator.

Waveguide method

The material sample is now placed in a waveguide. The dielectric constant is then calculated from the measured reflection or transmission coefficient. This method enables measuring the properties of a material over a wide frequency band.

In Fig. 156 we see a sample of length l located in a waveguide. Taking into account the multiple reflections between the two interfaces we get the voltage transmission coefficient

$$\tau = \frac{V_t}{V_i} = \frac{(1-\Gamma^2)e^{-\gamma l}}{1-\Gamma^2 e^{-2\gamma l}} .$$

The reflection coefficient Γ at the interface between air and the material is

$$\Gamma = \frac{Z_d - Z_0}{Z_d + Z_0} = \frac{\sqrt{1-(f_c/f)^2} - \sqrt{\varepsilon_r - (f_c/f)^2}}{\sqrt{1-(f_c/f)^2} + \sqrt{\varepsilon_r - (f_c/f)^2}},$$

where f_c is the waveguide cutoff frequency without the sample (in a coaxial cable $f_c = 0$). The propagation constant γ is

$$\gamma = j\frac{2\pi}{\lambda_0}\sqrt{\varepsilon_r - \left(\frac{f_c}{f}\right)^2} .$$

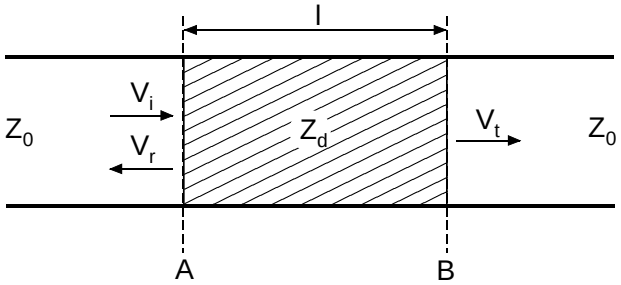

Figure 156. Material sample in a waveguide.

For the reflection coefficient we get accordingly

$$\rho = \frac{V_r}{V_i} = \frac{(1-e^{-2\gamma l})\Gamma}{1-\Gamma^2 e^{-2\gamma l}} .$$

Usually, a short circuit is placed at the interface B in Fig. 156. In this case, the previous equation changes into the form

$$\rho = \frac{\Gamma - e^{-2\gamma l}}{1-\Gamma e^{-2\gamma l}} .$$

The highest accuracy is achieved if the sample length is $\lambda/4$, where λ_d is the wavelength in the sample.

From the measured amplitude and phase of τ or ρ, ε_r' and ε_r'' are solved. By varying frequency or length of the sample, ε_r' and ε_r'' can be solved merely from the measured amplitudes.

The sample can be inserted to a short-circuited coaxial cable as part of the center conductor, as shown in Fig. 157(a). The method in which the sample is placed at the end of an open waveguide, as shown in Fig. 157(b), does not provide the most accurate ε_r measurements, but it has many advantages that make the method very useful: it is fast and simple, the end of the waveguide can be dipped into a liquid or pressed onto the flat surface of a solid medium, so a preparation of a sample is rarely necessary, and with a coaxial cable broadband measurements, e.g. from 200 MHz to 20 GHz is possible [96].

A sample in a transmission line can also be measured in the time domain as described earlier and in [97, 98]: a short pulse is sent towards the sample and with help of the waveform of the original pulse and that of the reflected or transmitted signal, the properties of the sample are calculated over a wide frequency band.

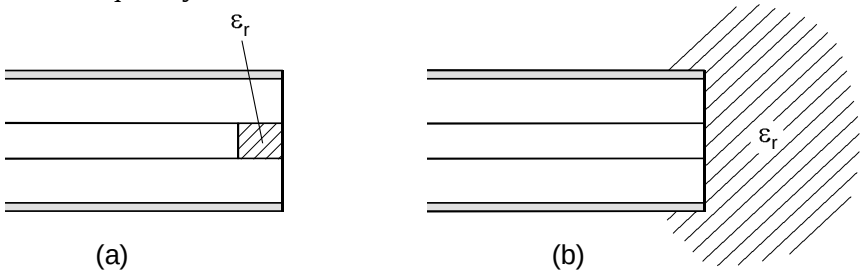

Figure 157. Waveguide method, the sample is (a) part of the center conductor of a short-circuited coaxial line, or (b) placed at the open end of a transmission line.

Free-space method

Reflection and transmission measurements can be made by measuring the power that couples from one antenna to another (Fig. 158). Samples should be large enough to cover the antenna main lobe almost completely. The method is best suited for frequencies above 5 GHz. At lower frequencies diffraction around the edges of the sample are difficult to avoid. Free-space measurements resemble the waveguide method. The above equations apply to plane waves, where $f_c = 0$.

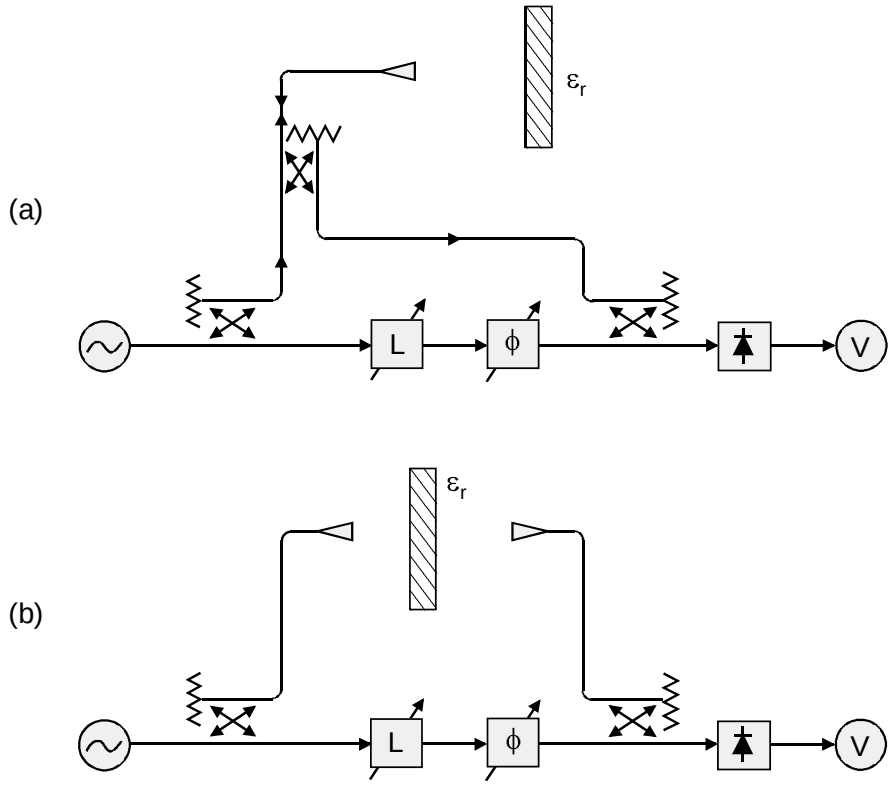

Figure 158. Free-space method: (a) reflection measurement, (b) transmission measurement.

Measurements are performed with a vector network analyzer, or as shown in Fig. 158 using the bridge setup. With the attenuator and the phase shifter the detector output is adjusted to zero, i.e. the signals have equal amplitude and opposite phase. From the change in attenuation and phase of the transmission without and with the sample ρ and τ are obtained. Reflections of the environment and multiple reflections between the antennas must be minimized.

The sample can also be measured as a function of angle as illustrated in Fig. 159, by measuring only the transmitted power [99]. In this case, the EM-field transmission factor is

$$\tau = \frac{(1-\Gamma^2)e^{-(\gamma_1-\gamma_0)l}}{1-\Gamma^2 e^{-2\gamma_1 l}}$$

where

$$\gamma_1 = j\frac{2\pi}{\lambda_0}\sqrt{\varepsilon_r - \sin^2\theta}\,,$$

$$\gamma_0 = j\frac{2\pi}{\lambda_0}\cos\theta\,.$$

The reflection coefficient Γ at the interface depends on the polarization. For orthogonal polarization (electric field vector perpendicular to the plane of the measurement, i.e. the rotation axis of the sample)

$$\Gamma_\perp = \frac{\cos\theta - \sqrt{\varepsilon_r - \sin^2\theta}}{\cos\theta + \sqrt{\varepsilon_r - \sin^2\theta}}$$

and for parallel polarization

$$\Gamma_{\parallel} = \frac{\sqrt{\varepsilon_r - \sin^2\theta} - \varepsilon_r \cos\theta}{\sqrt{\varepsilon_r - \sin^2\theta} + \varepsilon_r \cos\theta}$$

With orthogonal polarization we achieve higher accuracy, because $\Gamma_{\perp} > \Gamma_{\parallel}$. The dielectric constant is obtained by fitting the theoretical curve to the measured transmission curve $|\tau|^2$

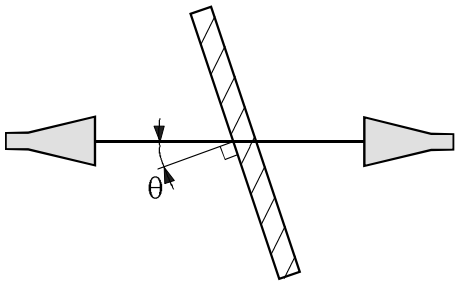

Figure 159. Free space method in which power transmission is measured as a function of the angle of incidence.

[83] Nyfors E., Vainikainen P.: *Industrial Microwave Sensors*. Norwood, Massachusetts 1989, Artech House.

[84] Afsar M. N., Birch J. R., Clarke R. N.: The measurement of the properties of materials. *Proceedings of the IEEE*, 74(1986)1, s. 183–199.

[85] Afsar M. N.: Dielectric measurements of millimeter-wave materials. *IEEE Transactions on Microwave Theory and Techniques*, 32(1984)12, s. 1598–1609.

[86] Birch J. R., Clarke R. N.: Dielectric and optical measurements from 30 to 1000 GHz. *The Radio and Electronic Engineer*, 52(1982)11/12, s. 565–584.

[87] Vainikainen P. V.: *Measurement electronics of industrial microwave resonator sensors*. Helsinki University of Technology, Radio Laboratory, Report S 194, 1991.

[88] Fischer M.: *Microwave stripline sensors for industrial measurement applications*. Helsinki University of Technology, Radio Laboratory, Report S 218, 1995.

[89] Toropainen A. P.: *Analysis of a method for measuring properties of granular materials based on cross-polarized Rayleigh scattering of microwaves*. Helsinki University of Technology, Radio Laboratory, Report S 219, 1996.

[90] Jakkula P.: *Microwave sensors for consistency measurements in the pulp and paper industry*. Helsinki University of Technology, Radio Laboratory, Report S 231, 1998.

[91] Nyfors E. G.: *Cylindrical microwave resonator sensors for measuring materials under flow*. Helsinki University of Technology, Radio Laboratory, Report S 243, 2000.

[92] Hirvonen T. M., Vainikainen P., Lozowski A., Räisänen A. V.: Measurement of dielectrics at 100 GHz with an open resonator connected to a network analyzer. *IEEE Transactions on Instrumentation and Measurement*, 45(1996)4, s. 780–786.

[93] Dudorov S. N., Lioubtchenko D. V., Mallat J. A., Räisänen A. V.: Differential open resonator method for permittivity measurements of thin dielectric film on substrate. *IEEE Transactions on Instrumentation and Measurement*, 54(2005)5, s. 1916–1920.

[94] Dudorov S. N., Lioubtchenko D. V., Mallat J. A., Räisänen A. V.: Millimeter-wave permittivity measurement of deposited dielectric films using the spherical open resonator. *IEEE Microwave and Wireless Components Letters*, 15(2005)9, s. 564–566.
[95] Harrington R.: *Time-Harmonic Electromagnetic Fields*. New York 1961, McGraw-Hill.
[96] Hewlett-Packard: HP 85070A dielectric probe kit 200 MHz to 20 GHz, convenient non-destructive dielectric measurements.
[97] Courtney C. C.: Time-domain measurements of the electromagnetic properties of materials. *IEEE Transactions on Microwave Theory and Techniques*, 46(1998)5, s. 517–522.
[98] Courtney C. C., Motil W.: One-port time-domain measurement of the approximate permittivity and permeability of materials. *IEEE Transactions on Microwave Theory and Techniques*, 47(1999)5, s. 551–555.
[99] Shimabukuro F. I., Lazar S., Chernick M. R, Dyson H. B.: A quasi-optical method for measuring the complex permittivity of materials. *IEEE Transactions on Microwave Theory and Techniques*, 32(1984)7, s. 659–665.