RF measurements I: signal receiving techniques

F. Caspers
CERN, Geneva, Switzerland

Abstract
For the characterization of components, systems and signals in the RF and microwave range, several dedicated instruments are in use. In this paper the fundamentals of the RF-signal sampling technique, which has found widespread applications in 'digital' oscilloscopes and sampling scopes, are discussed. The key element in these front-ends is the Schottky diode which can be used either as an RF mixer or as a single sampler. The spectrum analyser has become an absolutely indispensable tool for RF signal analysis. Here the front-end is the RF mixer as the RF section of modern spectrum analysers has a rather complex architecture. The reasons for this complexity and certain working principles as well as limitations are discussed. In addition, an overview of the development of scalar and vector signal analysers is given. For the determination of the noise temperature of a one-port and the noise figure of a two-port, basic concepts and relations are shown. A brief discussion of commonly used noise measurement techniques concludes the paper.

1 Introduction

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

2 Basic elements and concepts

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

2.1 Decibel

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

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

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

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

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

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

In Eqs. (1) and (2) P_0 and V_0 are the reference power and voltage, respectively. A given value in dB is the same for power ratios as for voltage ratios. Please note that there are no 'power dB' or 'voltage dB' as dB values always express a ratio.

Conversely, the absolute power and voltage can be obtained from dB values by

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

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

Logarithms are useful as the unit of measurement because

1. signal power tends to span several orders of magnitude and
2. signal attenuation losses and gains can be expressed in terms of subtraction and addition.

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

Frequently dB values are expressed using a special reference level and not SI units. Strictly speaking, the reference value should be included in parentheses when giving a dB value, e.g., +3 dB (1 W) indicates 3 dB at $P_0 = 1$ watt, thus 2 W. However, it is more common to add some typical reference values as letters after the unit, for instance, dBm defines dB using a reference level of $P_0 = 1$ mW. Thus, 0 dBm correspond to -30 dBW, where dBW indicates a reference level of $P_0 = 1$ W. Often a reference impedance of $50\,\Omega$ is assumed. Other common units are

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

2.2 The RF diode

One of the most important elements inside all sophisticated measurement devices is the fast RF diode or Schottky diode. The basic metal-semiconductor junction has an intrinsically very fast switching time of well below a picosecond, provided that the geometric size and hence the junction capacitance of the diode is small enough. However, this unavoidable and voltage dependent junction capacity will lead to limitations of the maximum operating frequency.

The equivalent circuit of such a diode is depicted in Fig. 1 and an example of a commonly used Schottky diode can be seen in Fig. 2. One of the most important properties of any diode is its character-
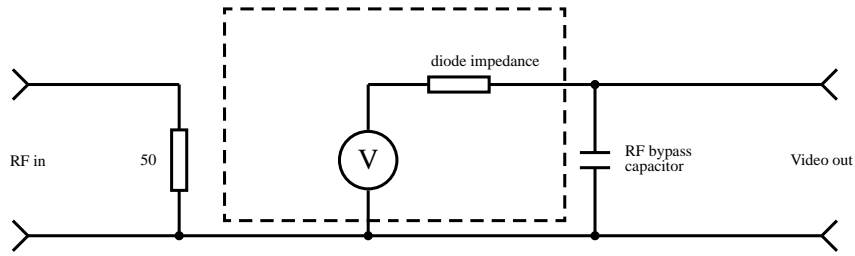
Fig. 1: The equivalent circuit of a diode

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

istic which is the relation of current as a function of voltage. This relation is described by the Richardson equation [2]:

$$I = A A_{\mathrm{RC}} T^2 \exp\left(-\frac{q\phi_{\mathrm{B}}}{kT}\right) \left[\exp\left(\frac{qV_{\mathrm{J}}}{kT}\right) - \mathrm{M}\right] \ , \tag{5}$$

where A is the area in cm^2, A_{RC} the modified Richardson constant, k Boltzmann's constant, T the absolute temperature, ϕ_{B} the barrier height in volts, V_{J} the external Voltage across the depletion layer, M the avalanche multiplication factor and I the diode current.

This relation is depicted graphically for two diodes in Fig. 3. As can be seen, the diode is not an ideal commutator (Fig. 4) for small signals. Note that it is not possible to apply big signals, since this kind of diode would burn out. However, there exist rather large power versions of Schottky diodes which can stand more than 9 kV and several 10 A but they are not suitable in microwave applications due to their large junction capacity.

The Richardson equation can be roughly approximated by a simpler equation [2]:

$$I = I_{\mathrm{s}} \left[\exp\left(\frac{V_{\mathrm{J}}}{0.028}\right) - 1\right] \ . \tag{6}$$

This approximation can be used to show that the RF rectification is linked to the second derivation (curvature) of the diode characteristic.
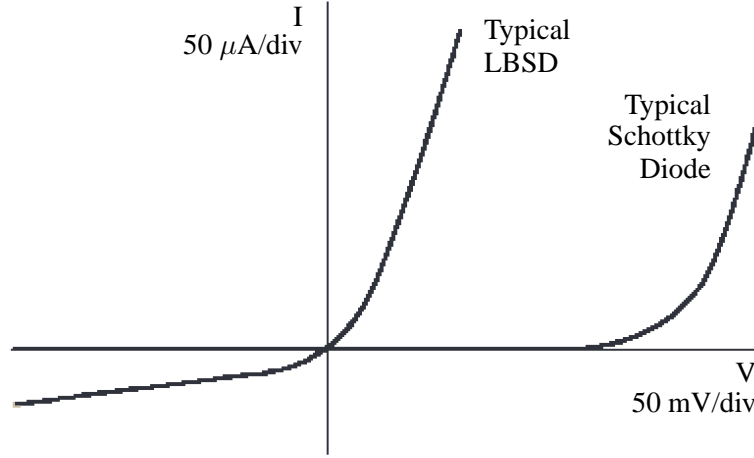
3

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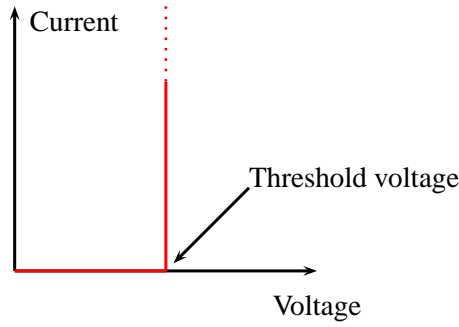

Fig. 4: The current-voltage relation of an ideal commutator with threshold voltage

If the DC current is held constant by a current regulator or a large resistor assuming external DC bias[1], then the total junction current, including RF is

$$I = I_0 = \mathrm{i}_0 \cos \omega t \tag{7}$$

and hence the current-voltage relation can be written as

$$V_\mathrm{J} = 0.028 \ln \left(\frac{I_\mathrm{S} + I_0 + \mathrm{i}\cos\omega t}{I_\mathrm{S}} \right) = 0.028 \ln \left(\frac{I_0 + I_\mathrm{S}}{I_\mathrm{S}} \right) + 0.028 \ln \left(\frac{\mathrm{i}\cos\omega t}{I_0 + I_\mathrm{S}} \right) \quad . \tag{8}$$

If the RF current I is small enough, the second term can be approximated by Taylor expansion:

$$V_\mathrm{J} \approx 0.028 \ln \left(\frac{I_0 + I_\mathrm{S}}{I_\mathrm{S}} \right) + 0.028 \left[\frac{\mathrm{i}\cos\omega t}{I_0 + I_\mathrm{S}} - \frac{\mathrm{i}^2\cos^2\omega t}{2(I_0 + I_\mathrm{S})^2} + \ldots \right] = V_\mathrm{DC} + V_\mathrm{J}\cos\omega t + \text{higher order terms} \tag{9}$$

With the identity $\cos^2 = 0.5$, the DC and the RF voltages are given by

$$V_\mathrm{J} = \frac{0.028}{I_0 + I_\mathrm{S}}\mathrm{i} = R_\mathrm{S}\mathrm{i} \quad \text{and} \quad V_\mathrm{DC} = 0.028 \ln \left(1 + \frac{I_0}{I_\mathrm{S}} \right) - \frac{0.028^2}{4(I_0 + I_\mathrm{S})^2} = V_0 - \frac{V_\mathrm{J}^2}{0.112} \quad . \tag{10}$$

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

---

[1]Most diodes do not need an external bias, since they have a DC return self-bias.
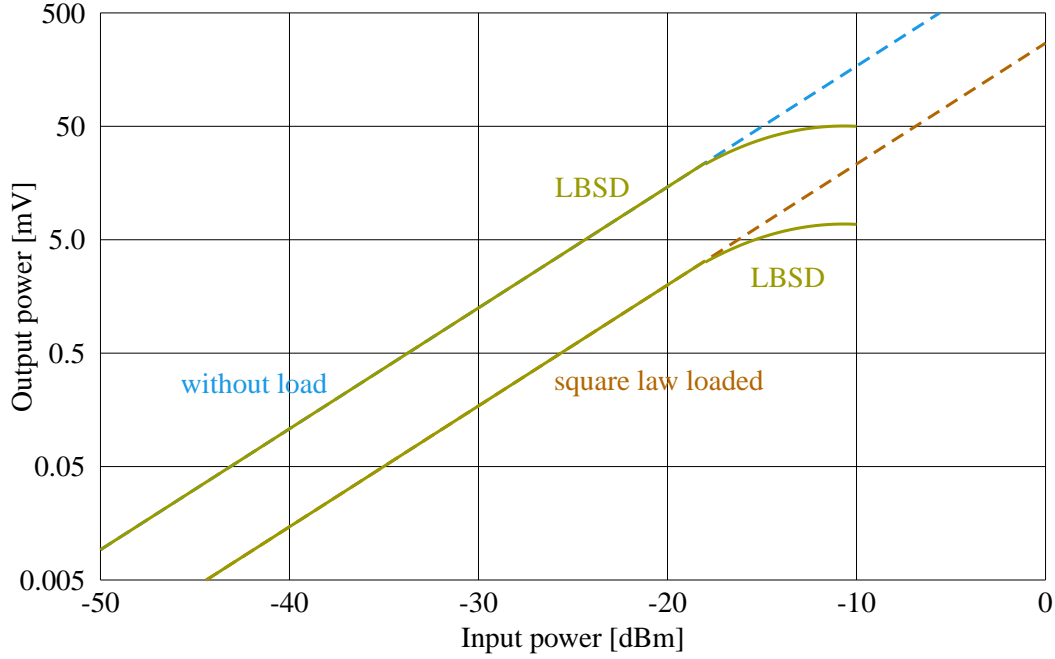
4

Fig. 5: Relation between input power and output voltage.

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

The transition between the linear region and the square-law region is typically between -10 and -20 dB (Fig. 5).

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

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

To avoid these limitations, another method of operating such diodes is needed.

2.3 Mixer

For the detection of very small RF signals a device that has a linear response over the full range (from 0 dBm ($=1$mW) down to thermal noise $= -174$ dBm/Hz $= 4\cdot10^{-21}$ W/Hz) is preferred. An RF mixer provides these features using 1, 2, or 4 diodes in different configurations (Fig. 6). A mixer is essentially a multiplier with a very high dynamic range implementing the function:

$$f_1(t)f_2(t) \quad \text{with } f_1(t) = \text{ RF signal} \quad \text{and } f_2(t) = \text{ LO signal} \ , \tag{11}$$

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

$$a_1 \cos(2\pi f_1 t + \varphi) \cdot a_2 \cos(2\pi f_2 t) = \frac{1}{2}a_1 a_2 \left[\cos((f_1 + f_2)t + \varphi) + \cos((f_1 - f_2)t + \varphi)\right] \ . \tag{12}$$

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

Examples of different mixer configurations are shown in Fig. 6.
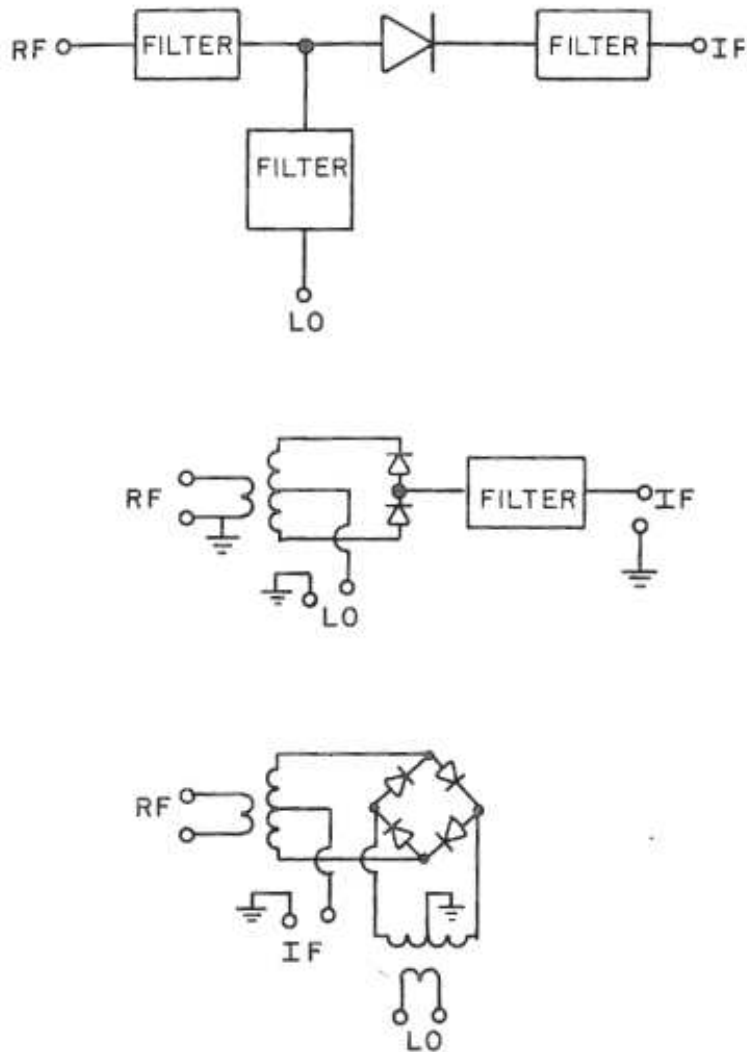
5

Fig. 6: Examples of different mixer configurations

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

The response of a mixer in time domain is depicted in Fig. 8.

The output signal is always in the 'linear range' provided that the mixer is not in saturation with respect to the RF input signal. Note that for the LO signal the mixer should always be in saturation to make sure that the diodes work as a nearly ideal switch. The phase of the RF signal is conserved in the output signal available form the RF output.

2.4 Amplifier

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

$$G[\text{dB}] = 20 \cdot \frac{V_{\text{RFout}}}{V_{\text{RFin}}} \quad \text{or} \quad \frac{V_{\text{RFout}}}{V_{\text{RFin}}} = 20 \cdot \log G[\text{lin}] \quad . \tag{13}$$

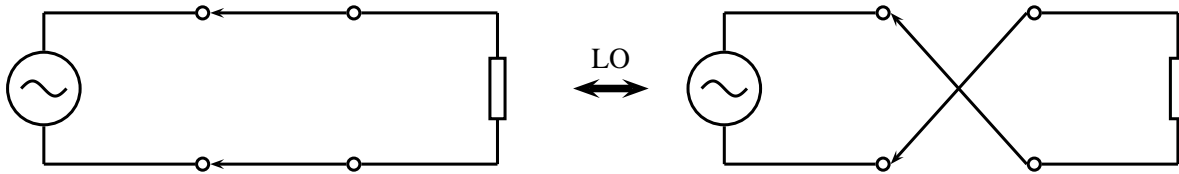The circuit symbol of an amplifier is shown in Fig. 9 together with its S-matrix.

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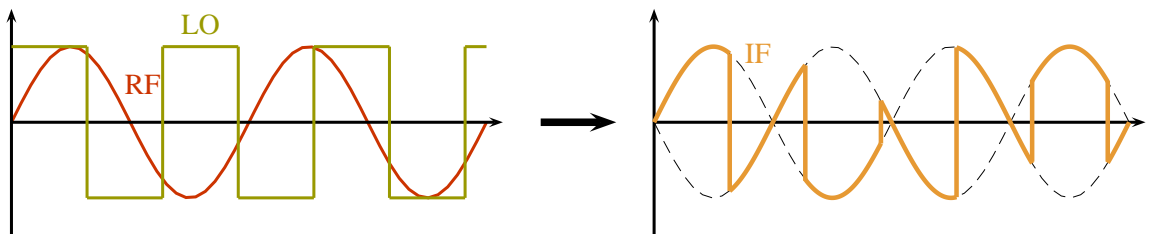

Fig. 8: Time domain response of a mixer

$$S = \begin{pmatrix} 0 & 0 \\ G & 0 \end{pmatrix}$$
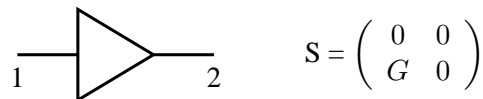
Fig. 9: Circuit symbol an S-matrix of an ideal amplifier

The bandwidth of an amplifier specifies the frequency range where it is usually operated. This frequency range is defined by the $-3\,\mathrm{dB}$ points[2] with respect to its maximum or nominal transmission gain.

In an ideal amplifier the output signal would be proportional to the input signal. However, a real amplifier is nonlinear, such that for larger signals the transfer characteristic deviates from its linear properties valid for small signal amplification. When increasing the output power of an amplifier, a point is reached where the small signal gain becomes reduced by 1 dB (Fig. 10). This output power level defines the 1 dB compression point, which is an important measure of quality for any amplifier (low level as well as high power).

The transfer characteristic of an amplifier can be described in terms which are commonly used for RF engineering, i.e., the S-matrix (for further details see the paper on S-matrices of this school). As implicitly contained in the S-matrix, the amplitude and phase information of any spectral component are

---

[2]The $-3\,\mathrm{dB}$ points are the points left and right of a reference value (e.g. a local maximum of a curve) that are 3 dB lower than the reference.
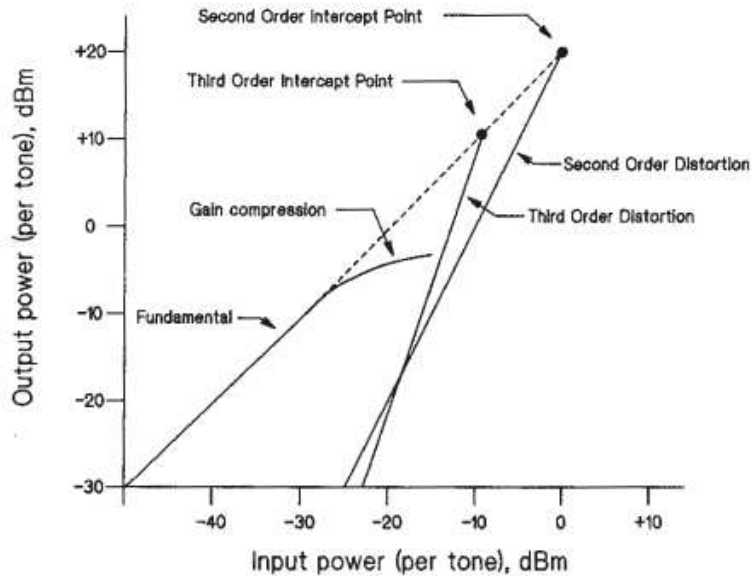
Fig. 10: Example for the 1dB compression point [4].

preserved when passing through an ideal amplifier. For a real amplifier the element $G = S_{21}$ (transmission from port 1 to port 2) is not a constant but a complex function of frequency. Also the elements S_{11} and S_{22} are not 0 in reality.

2.5 Interception points of nonlinear devices

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

One of the most relevant interception points is the interception point of 3rd order (IP3 point). Its importance derives from its straightforward determination, plotting the input versus the output power in logarithmic scale (Fig. 10). The IP3 point is usually not measured directly, but extrapolated from measurement data at much smaller power levels in order to avoid overload and damage of the device under test (DUT). If two signals $(f_1, f_2 > f_1)$ which are closely spaced by Δf in frequency are simultaneously applied to the DUT, the intermodulation products appear at $+ \Delta f$ above f_2 and at $- \Delta f$ below f_1.

The transfer functions or weakly nonlinear devices can be approximated by Taylor expansion. Using n higher order terms on one hand and plotting them together with an ideal linear device in logarithmic scale leads to two lines with different slopes ($x^n \overset{\log}{\to} n \cdot \log x$). Their intersection point is the intercept point of nth order. These points provide important information concerning the quality of nonlinear devices.

In this context, the aforementioned 1 dB compression point of an amplifier is the intercept point of first order.

Similar characterization techniques can also be applied with mixers which with respect to the LO signal cannot be considered a weakly nonlinear device.

2.6 The superheterodyne concept

The word superheterodyne is composed of three parts: super (Latin: over), $\epsilon\tau\epsilon\rho\omega$ (hetero, Greek: different) and $\delta\upsilon\nu\alpha\mu\iota\sigma$ (dynamis, Greek: force) and can be translated as two forces superimposed[3]. Different

---

[3]The direct translation (roughly) would be: Another force becomes superimposed.

abbrevations exist for the superheterodyne concept. In the US it is often referred to by the simple word 'heterodyne' and in Germany one can find the terms 'super' or 'superhet'. The 'weak' incident signal is subjected to nonlinear superposition (i.e., mixing or multiplication) with a 'strong' sine wave from a local oscillator. At the mixer output we then get the sum and difference frequencies of the signal and local oscillator. The LO signal can be tuned such that the output signal is always at the same frequency or in a very narrow frequency band. Therefore a fixed frequency bandpass with excellent transfer characteristics can be used which is cheaper and easier than a variable bandpass with the same performance. A well-known application of this principle is any simple radio receiver (Fig. 11).
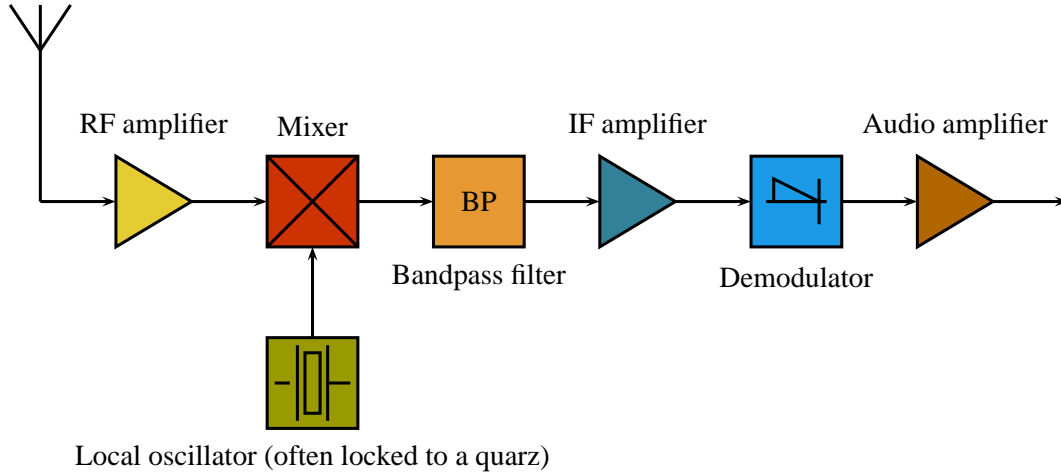
Fig. 11: Schematic drawing of a superheterodyne receiver

3 Oscilloscope

An oscilloscope is typically used for acquisition, display, and measurement of signals in time domain. The bandwidth of real-time oscilloscopes is limited in most cases to 10 GHz. For higher bandwidth on repetitive signals the sampling technique has been in use since about 1960. One of the many interesting features of modern oscilloscopes is that they can change the sampling rate through the sweep in a programmed manner. This can be very helpful for detailed analysis in certain time windows. Typical sampling rates are between a factor 2.5 and 4 of the maximum frequency (according to the Nyquist theorem a real-time minimum sampling rate of twice the maximum frequency f_{max} is required).

Sequential sampling (Fig. 12) requires a pre-trigger (required to open the sampling gate) and permits a non-real-time bandwidth of more than 100 GHz with modern scopes.

Random sampling (rarely used these days, Fig. 13) was developed about 40 years ago (around 1970) for the case no pre-trigger was available and relying on a strictly periodic signal to predict a pre-trigger from the measured periodicity.

Sampling is discussed in more detail in the following. Consider a bandwidth-limited time function s(t) and its Fourier transform S(f). The signal s(t) is sampled (multiplied) by a series of equidistant δ-pulses p(t) [5]:

$$\mathrm{p}(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nT_{\mathrm{s}}) = III(t/T_{\mathrm{s}}) \qquad (14)$$

where the symbol III is derived from the Russian letter Ш and is pronounced 'sha'. It represents a series of δ-pulses.
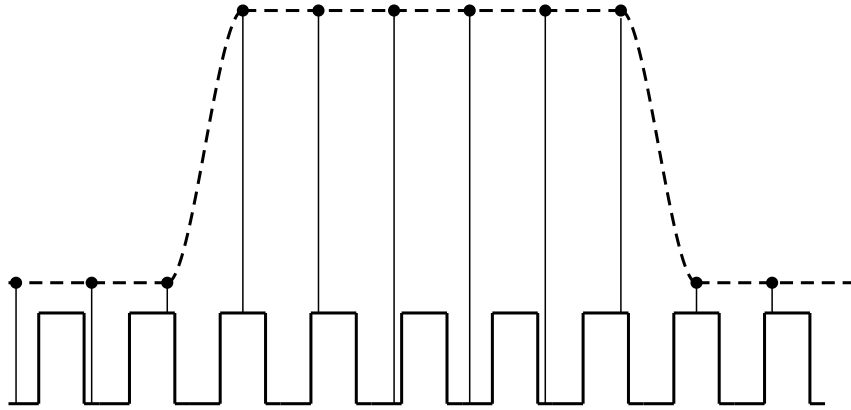
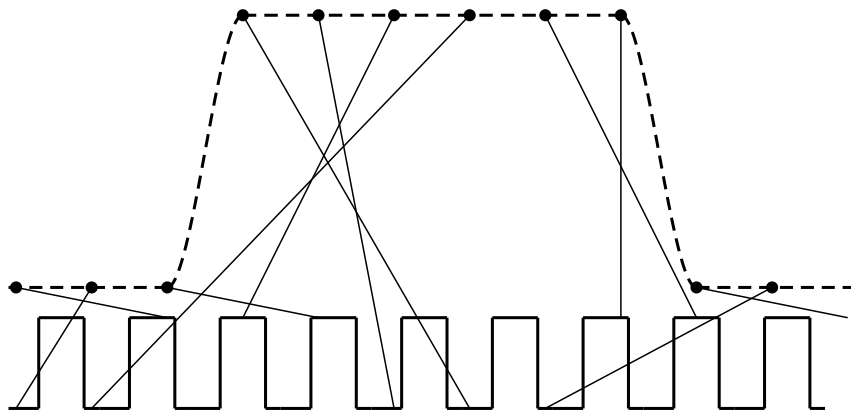**Fig. 12:** Illustration of sequential sampling

Fig. 13: Illustration of random sampling

The sampled time functions $s_s(t)$ is

$$s_s(t) = s(t)p(t) = s(t)III(t/T_s)$$

$$S_s(f) = S(f) * \frac{1}{T_s}III(T_sf) \tag{15}$$

$$S_s(f) = \frac{1}{T_s}\sum_{n=-\infty}^{+\infty} S(f - mF) \text{ with } F = \frac{1}{T_s} \quad . \tag{16}$$

Note that the spectrum is repeated periodically by the sampling process. For proper reconstruction, one ensures that overlapping as in Fig. 14 does not occur.

If the spectra overlap as in Fig. 14 we have undersampling, the sampling rate is too low. If big gaps occur between the spectra (Fig. 15) we have oversampling, the sampling rate is too high. But this scheme applies in most cases. In the limit we arrive at a Nyquist rate of $1/T_s = 2f_g = F$.

The rules mentioned above are of great importance for all 'digital' oscilloscopes. The performance (conversion time, resolution) of the input ADC (analog digital converter) is the key element for single-shot rise time. With several ADCs in time-multiplex one obtains these days 8-bit vertical resolution at $20\,\text{GSa/s} = 10\,\text{GHz}$ bandwidth.

Another way to look at the sampling theorem (Nyquist) is to consider the sampling gate as a
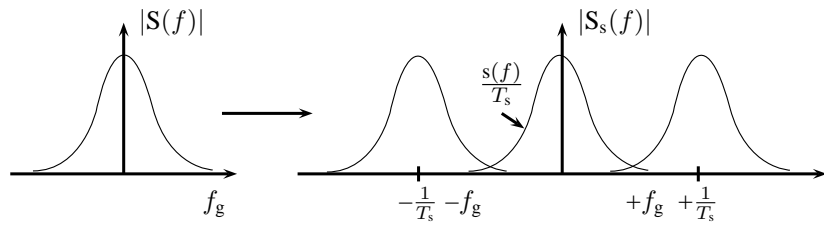
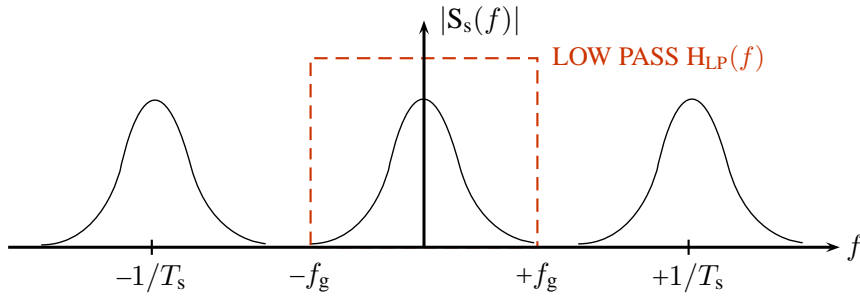**Fig. 14:** Periodically repeated component of the Fourier Transform of $s_s(t)$ [5]

Fig. 15: Reconstruction of $S(f)$ via ideal lowpass from $S_s(f)$ (slightly oversampled)
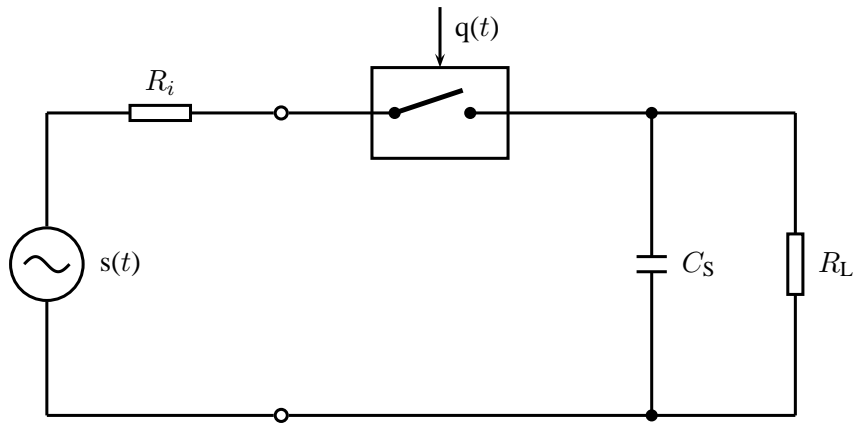
harmonic mixer (Fig. 16).

Fig. 16: Sampling gate as harmonic mixer; C_s = sampling capacitor [6]

This is basically a nonlinear element (e.g., a diode) that gives product terms of two signals superimposed on its nonlinear characteristics.

The switch in Fig. 16 may be considered as a periodically varying resistor $R(t)$ actuated by $q(t)$. If $q(t)$ is not exactly a δ-function then the higher harmonics decrease with f and the spectral density becomes smaller at high frequencies.

For periodic signals one may apply a special sampling scheme. With each signal event the sampling time is moved by a small fraction Δt along the signal to be measured (Fig. 17). The highest possible
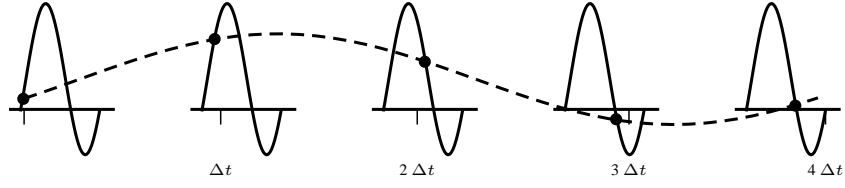
Fig. 17: Signal reconstruction with sampling shift by Δt per pulse [7]

signal frequency for this sequential sampling is linked to the width of the sampling pulse. This sampling or gating pulse should be as short as possible otherwise signal averaging during the 'gate-open' period would take place.

The sampling pulse is often generated by step-recovery diodes (snap-off diodes) which change their conductivity very rapidly between the conducting and non-conducting state. The actual switch (Schottky diode) becomes conductive during the gate pulse and charges a capacitor (sample and hold circuit) but not to the full signal voltage. Assuming a time constant $R_i C_s$ much bigger than the 'open' time of the sampling gate, we obtain approximately (Fig. 18)

$$i_c(t) = \frac{s(t)}{R_i + R_d(t)} \quad .$$

(17)

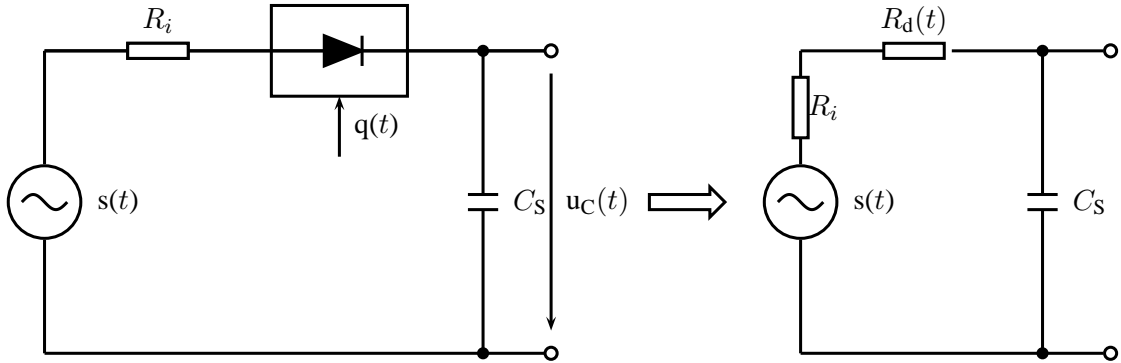After the sampling process we have [6]

Fig. 18: Equivalent circuit for the sample-and-hold element [6]

$$u_c(t) = \frac{1}{C_s} \int_{-\infty}^{+\infty} i_c(t)dt = \int_{-\infty}^{+\infty} s(t) \frac{1}{C_s(R_i + R_d(t))} \quad ,$$

(18)

with the control signal for the Schottky diode being

$$q(t) = \frac{1}{C_s(R_i + R_d(t))} \quad .$$

(19)

The control or switching signal is moved by τ or $n\Delta t$ (Fig. 21) with respect to the signal to be sampled $s(t)$:

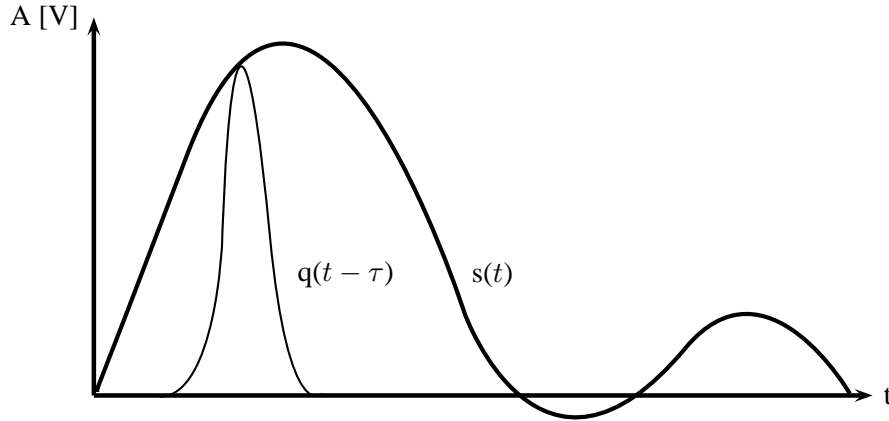$$u_c(\tau) = \int_{-\infty}^{+\infty} s(t)q(t - \tau)dt \quad .$$

(20)

12

Fig. 19: Sampling with finite-width sampling pulse

Note that the time constant $R_i C_s$ is much bigger than the length of q(t). C_s is only charged to a fraction of s(t) (Fig. 19). The sampling efficiency η is defined as

$$\eta = \frac{u_c(\tau)}{s(\tau)} \quad . \tag{21}$$

In order to circumvent the problem of poor sampling efficiency a feedback loop technique (integrator) can be used. This integrator amplifies the voltage step on the sampling capacitor, after the sampling gate is closed, exactly by a factor $1/\eta$. If the sampling gate has not moved with respect to the trigger, the sampling capacitor is already charged to the correct voltage u$_c(\tau)$ and there is no change. Otherwise the change in u$_c$ just amounts to the change in signal voltage.

The sampling gate is interesting from a technological point of view. As aperture times (Fig. 19) may be of the order of 10 ps, MIC (Microwave Integrated Circuit) technology has been used for many years. Today, the latest generation of sampling heads (50 GHz) is even one step further with MMIC (Monolithic Microwave Integrated Circuits) technology.

In MIC technology the sampling pulse is applied to a slotline in the ground-plane metallization of a microstrip substrate (Fig. 20). This slot line has a length of some 10 mm and is shorted at both ends.
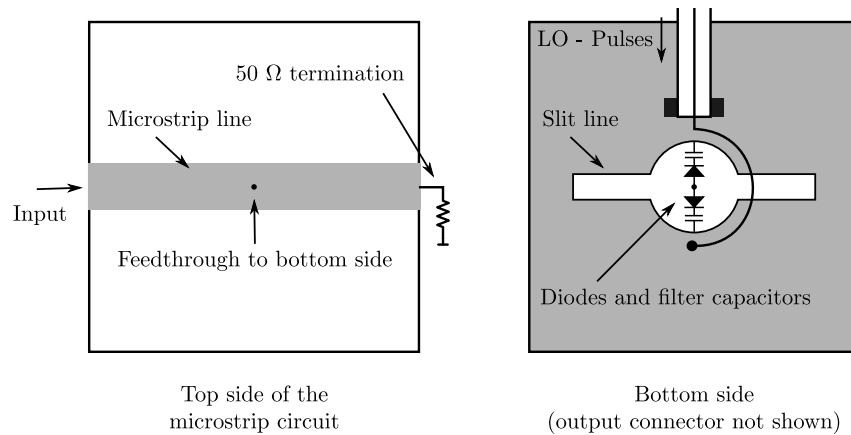


Fig. 20: Sampling circuits [7]

With a voltage across the slotline the fast Schottky diodes open and connect the microstrip line via a

Rise time: 7 ps	≈ 50 GHz
Jitter	≈ 1.5 ps
Static operation possible, no minimum repetition rate required.	
Optical sampling (mode-locked laser), 1 ps risetime	≈ 350 GHz

through hole to the sampling capacitor C_s. Owing to the particular topology of the circuit the signal line (microstrip) is decoupled from the sampling pulse line over a wide frequency range (Fig. 20).

To move the sampling pulse by Δt for each event requires a pre-trigger (several 10 ns ahead), to start a fast-ramp generator. The intersection (comparator) of the ramp generator output with a staircase-like reference voltage defines the sampling time and Δt (Fig. 21)
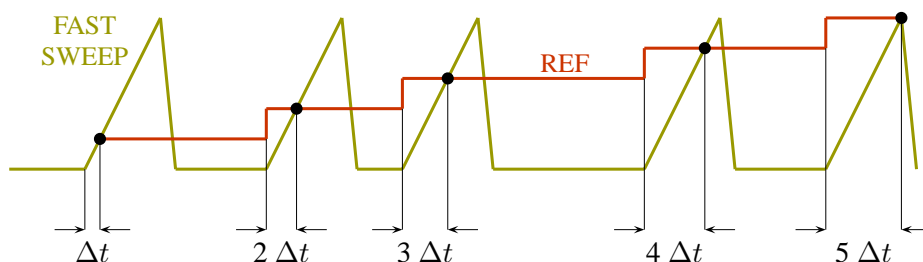
Fig. 21: Timing of sampling pulses [7]

The delay required for the pre-trigger has been a significant problem for many applications, since it may be as large as 70 ns on certain (older) instruments. A 70 ns delay-line leads to considerable signal distortions especially for the high-frequency components.

To avoid the delay for the pre-trigger a technique named 'random sampling' was developed about 45 years ago. It requires a strictly periodic signal rather than just the repetitive one for sequential sampling (Fig. 21). By measuring the (constant) repetition frequency of this strictly periodic signal, a prediction of the next pulse arrival time can be given in order to generate a trigger. Today there is little interest in random sampling, as pre-trigger delays are drastically reduced (12 ns). There are also problems with jitter, and random sampling needs repetition rates of serveral kHz [7].

Features of modern sampling scopes are summarized in Table 2.

4 Spectrum analyser

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

In general there are two types of spectrum analysers

– Scalar spectrum analysers (SA) and
– Vector spectrum analysers (VSA)

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

4.1 Scalar spectrum analysers

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

One of the major advantages of the frequency-domain display is the sensitivity to periodic perturbations. For example, 5% distortion is already difficult to see in the time domain but in the frequency domain the sensitivity to such 'sidelines' (Fig. 22) is very high (-120 dB below the main line). A very
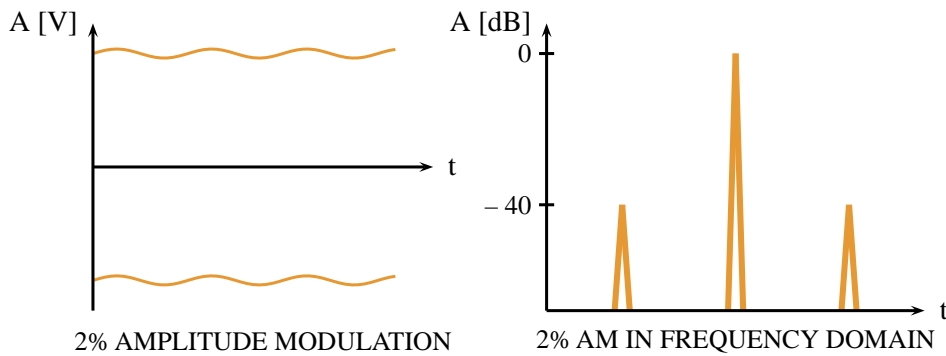

Fig. 22: Example of amplitude modulation in time and frequency domain

faint amplitude modulation (AM) of 10^{-12} (power) on some sinusoidal signals would be completely invisible on the time trace, but can be displayed as two sidelines 120 dB below the carrier in the frequency domain [8].

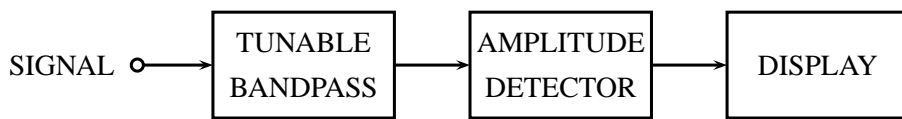We will now consider only serial processing or swept tuned analysers (Fig. 23).

Fig. 23: A tunable bandpass as a simple spectrum analyser (SA)

The easiest way to design a swept tuned spectrum analyser is by using a tunable bandpass. This may be an LC circuit, or a YIG filter (YIG = Yttrium-Iron-Garnet) beyond 1 GHz. The LC filter exhibits poor tuning, stability and resolution. YIG filters are used in the microwave range (as preselector) and for YIG oscillators. Their tuning range is about one decade, with Q-values exceeding 1000.

For much better performance the superheterodyne principle can be applied (Fig. 11).

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

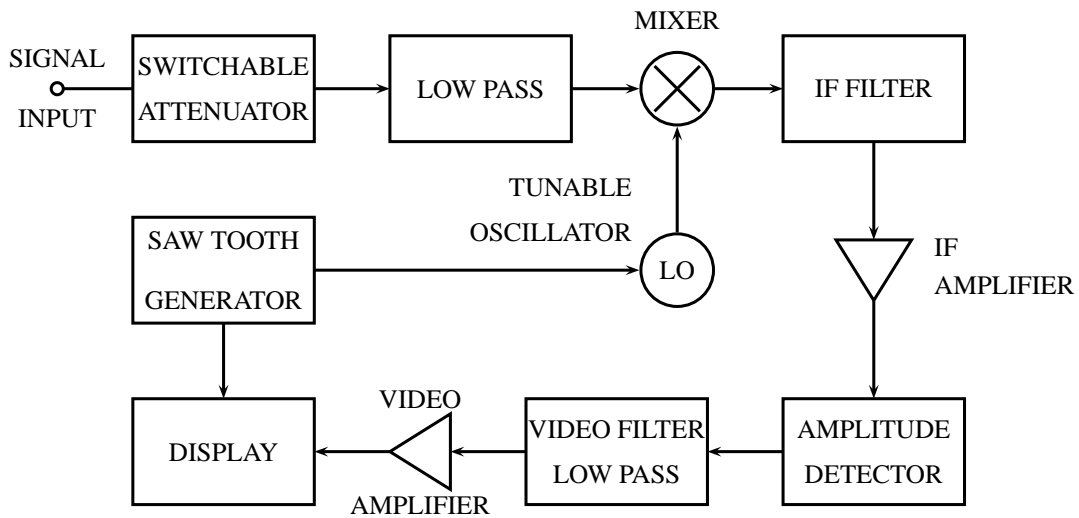$$f_s = f_{LO} \pm f_{IF} \ . \tag{22}$$

Fig. 24: Block diagram of a spectrum analyser

Assuming a signal range from 0 to 1 GHz for the spectrum analyser depicted in Fig. 24 and f_{LO} between 2 and 3 GHz we get the frequency chart shown in Fig. 25.
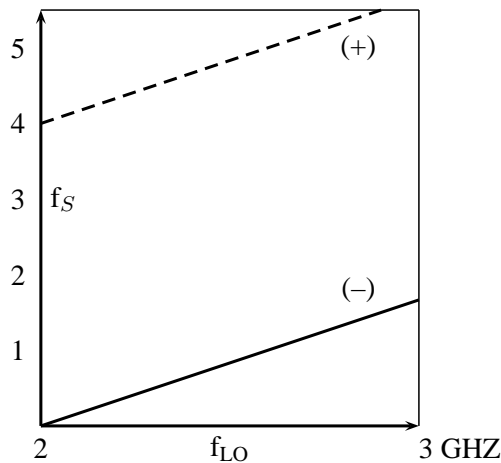

Fig. 25: Frequency chart of the SA of Fig. 24, intermediate frequency = 2 GHz

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

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

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

16

The requirements of a modem SA with respect to frequency are

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

and with respect to amplitude response are

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

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

$$\frac{\mathrm{d}f}{\mathrm{d}t} < (\Delta f)^2 \ . \tag{23}$$

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

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

Certain SAs do not use a sinusoidal LO signal but, rather, periodic short pulses or a comb spectrum (harmonic mixer). This is very closely related to a sampling scope, except that the spacing of the comb lines is different

$$f_s = N f_{\mathrm{LO}} \pm f_{\mathrm{IF}} \quad n = 1, 2, 3, ... \tag{24}$$

A single, constant input-frequency line may appear several times on the display. This difficulty (multiple response) was a particular problem with older instruments. Certain modulation and sweep modes permit the identification and rejection of these 'ghost' signals. On modern spectrum analysers the problem does not occur, except at frequencies beyond 60 GHz, when a tracking YIG filter may need to be installed.

Caution is advised when applying, but not necessarily displaying, two or more strong (> 10 dBm) signals to the input. Intermodulation 3rd-order products may appear (from the first mixer or amplifier) and could lead to misinterpretation of the signals to be analysed.

SAs usually have a rather poor noise figure of 20–40 dB as they often do not use preamplifiers in front of the first mixer (dynamic range, linearity). But with a good preamplifier the noise figure can be reduced to almost that of the preamplifier. This configuration permits amplifier noise figure measurements to be made with reasonable precision of about 0.5 dB. The input of the amplifier to be tested is connected to a hot and a cold termination and the corresponding two traces on the SA display are evaluated [9–13].

Spectrum analysers can also be used to directly measure the phase noise of an oscillator provided that the LO phase noise in the SA is much lower than that of the device under test [9]. For higher resolution, set-ups with delay lines and additional mixers (SA at low frequencies or FFT) are advised.

5   Vector spectrum and FFT analyser

The modern vector spectrum analyser (VSA) is essentially a combination of a two-channel digital oscilloscope and a spectrum analyser FFT display. The incoming signal gets down-mixed, bandpass (BP) filtered and passes an ADC (generalized Nyquist for BP signals; $f_{\mathrm{sample}} = 2$ BW). A schematic drawing of a modern VSA can be seen in Fig. 26. The digitized time trace then is split into an I (in phase) and Q
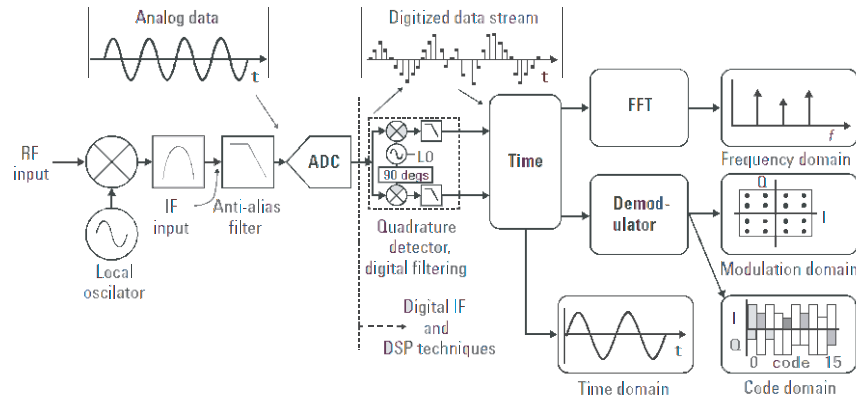
Fig. 26: Block diagram of a vector spectrum analyser

(quadrature, 90 degree offset) component with respect to the phase of some reference oscillator. Without this reference, the term vector is meaningless for a spectral component.

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

An example of vector spectrum analyser display and performance is given in Fig. 27 and Fig. 28. Both figures were obtained during measurements of the electron cloud in the CERN SPS.
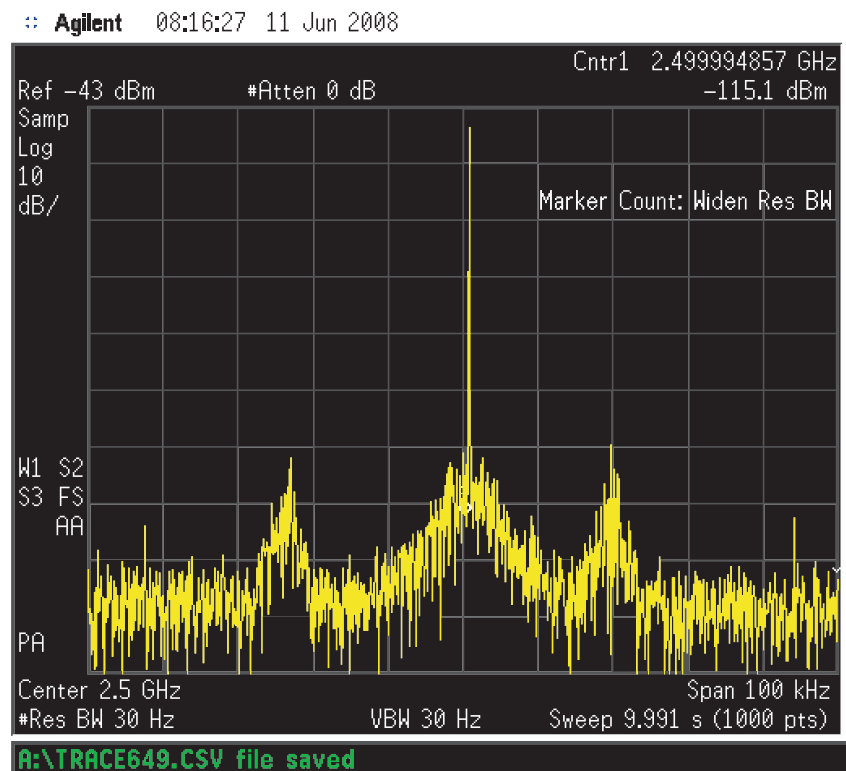

Fig. 27: Single-sweep FFT display similar to a very slow scan on a swept spectrum analyser

6 Noise basics

The concept of 'noise' was applied originally to the type of audible sound caused by statistical variations of the air pressure with a wide flat spectrum (white noise). It is now also applied to electrical signals, the noise 'floor' determining the lower limit of signal transmission. Typical noise sources are: Brownian
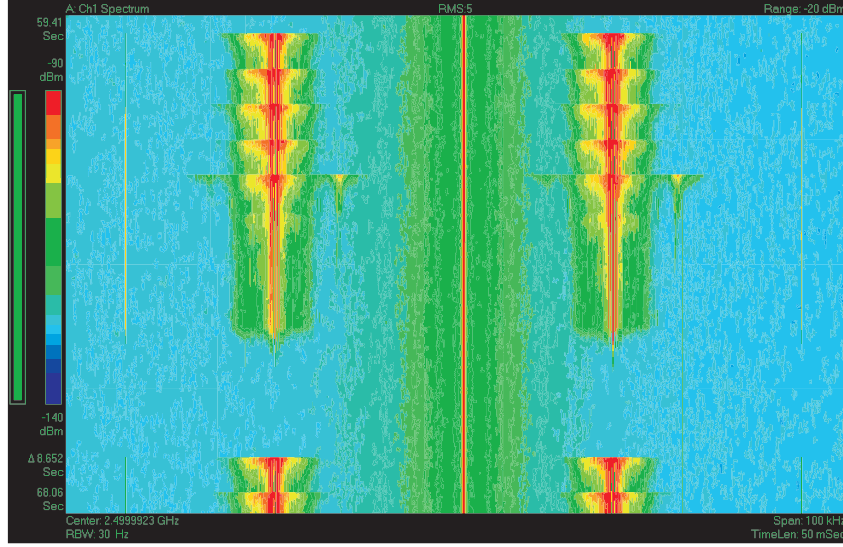
Fig. 28: Spectrogram display containing about 200 traces as shown on the left side in colour coding. Time runs from top to bottom.

movement of charges (thermal noise), variations of the number of charges involved in the conduction (flicker noise), and quantum effects (Schottky noise, shot noise). Thermal noise is only emitted by structures with electromagnetic losses which, by reciprocity, also absorb power. Pure reactances do not emit noise (emissivity = 0).

Different categories of noise can be defined:

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

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

$$\frac{N_{\mathrm{L}}}{\Delta f} = hf \left(\mathrm{e}^{hf/kT} - 1 \right)^{-1} \ , \tag{25}$$

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

Equation (25) indicates constant noise power density up to about 120 GHz (at 290 K) with 1% error. Beyond, the power density decays and there is no 'ultraviolet catastrophy', i. e., the total noise power is finite.

The radiated power density of a black body is given as

$$W_{\mathrm{r}}(f, T) = \frac{hf^3}{c^2 \left[\mathrm{e}^{hf/kT} - 1 \right]} \ . \tag{26}$$

For $hf << kT$ the Rayleigh–Jeans approximation of equation (25) holds:

$$N_{\mathrm{L}} = kT\Delta f \ , \tag{27}$$

19

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

$$\overline{u^2(t)} = 4kTR\Delta f \tag{28}$$

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

$$\overline{i^2(t)} = 4\frac{kT\Delta f}{R} = 4kTG\Delta f \ , \tag{29}$$

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

The available power (which is independent of R) is given by (Fig. 29)

$$\frac{\overline{u^2(t)}}{4R} = kT\Delta f \ . \tag{30}$$
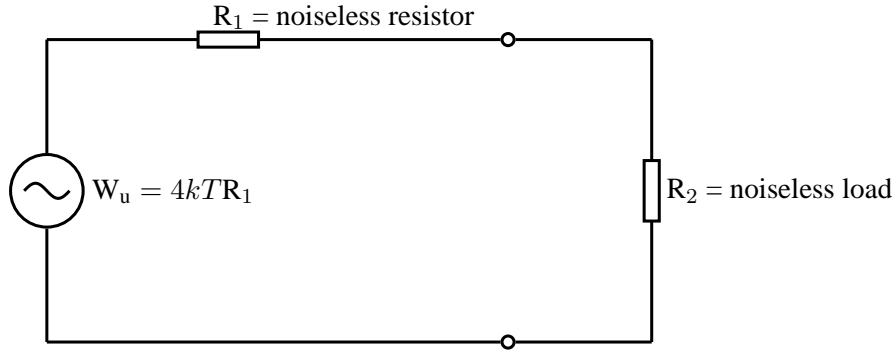

Fig. 29: Equivalent circuit for a noisy resistor R_1 terminated by a noisless load R_2

We define a spectral density function [9]

$$
\begin{aligned}
W_u(f) &= 4kTR \\
W_i(f) &= 4kTG \\
\overline{u^2(t)} &= \int_{f_1}^{f_2} W_u(f)\mathrm{d}f \ .
\end{aligned}
\tag{31}
$$

A noisy resistor may be composed of many elements (resistive network). In general, it is made from many carbon grains which have homogeneous temperatures. But if we consider a network of resistors with different temperatures and hence with an inhomogeneous temperature distribution (Fig. 30) the spectral density function changes to

$$W_u = \sum_j W_{uj} = 4kT_nR_i \ , \tag{32}$$

$$T_n = \sum_j \beta_j T_j \ , \tag{33}$$

where W_{uj} are the noise sources (Fig. 31), T_n is the total noise temperature, R_i the total input impedance, and β_j are coefficients indicating the fractional part of the input power dissipated in the resistor R_j. It is assumed that the W_{uj} are uncorrelated for reasons of simplicity.
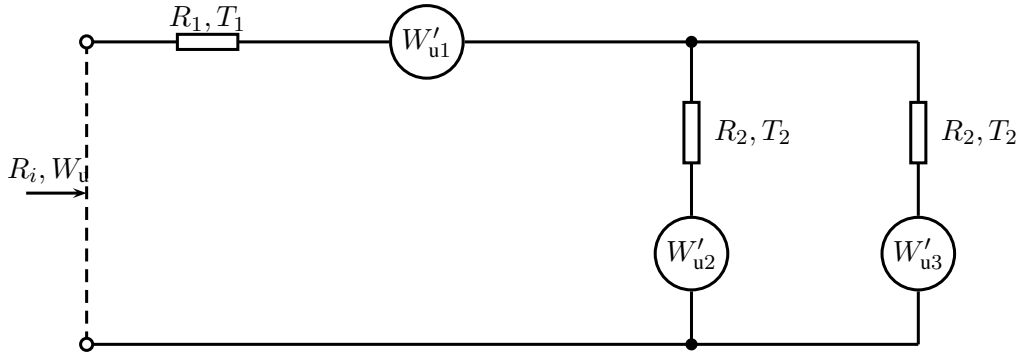
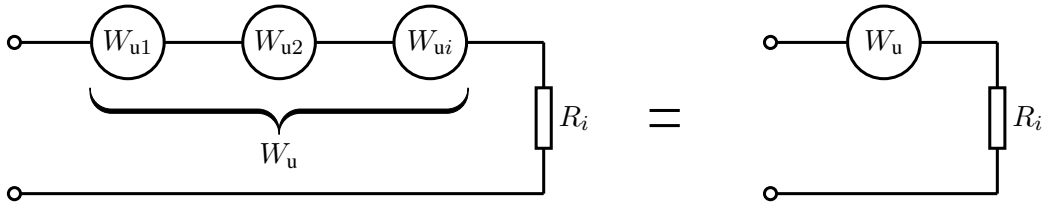**Fig. 30:** Noisy one-port with resistors at different temperatures [9, 14]

Fig. 31: Equivalent sources for the circuit of Fig. 30.

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

$$T_{\mathrm{n}} = \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \cdots \tag{34}$$

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

So far only pure resistors have been considered. Looking at complex impedances, it can be seen that losses from dissipation occur in $Re(\mathrm{Z})$ only. The available noise power is independent of the magnitude of $Re(\mathrm{Z})$ with $Re(\mathrm{Z}) > 0$. For Figs. 30 and 31 Eq.(33) still applies, except that R_{i} is replaced by $Re(\mathrm{Z_i})$. However, it must be remembered that in complex impedance networks the spectral power density W_{u} becomes frequency dependent [14].

The rules mentioned above apply to passive structures. A forward-biased Schottky diode (external power supply) has a noise temperature of about $T_0/2 + 10\%$. A biased Schottky diode is not in thermodynamic equilibrium and only half of the carriers contribute to the noise [9]. But it represents a real 50 Ω resistor when properly forward biased. For transistors and in particular field-effect transistors (FETs), the physical mechanisms are somewhat more complicated. Noise temperatures of 50 K have been observed on a FET transistor at 290 K physical temperature.

7 Noise-figure measurement with the spectrum analyser

Consider an ideal amplifier (noiseless) terminated at its input (and output) with a load at 290 K with an available power gain (G_a). We measure at the output [10, 15]:

$$P_a = kT_0 \Delta f G_a \ .$$ (35)

For $T_0 = 290$ K (or often 300 K) we obtain $kT_0 = -174$ dBm/Hz ($-$dBm = decibel below 1 mW). At the input we have for some signal S_i a certain signal/noise ratio S_i/N_i and at the output S_0/N_0. For an ideal (= noiseless) amplifier S_i/N_i is equal to S_0/N_0, i. e., the signal and noise levels are both shifted by the same amount. This gives the definition of the noise figure F:

$$F = \frac{S_i/N_i}{S_0/N_0} \ .$$ (36)

The ideal amplifier has F = 1 or F = 0 dB and the noise temperature of this amplifier is 0 K. The real amplifier adds some noise which leads to a decrease in S_0/N_0 due to the noise added (= N_{ad}):

$$F = \frac{N_{ad} + kT_0 \Delta f G_a}{kT_0 \Delta f G_a} \ .$$ (37)

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

In a similar way a factor Y can be defined:

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

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

$$F = \frac{[T_{(H)}/290) - 1] - Y[(T_C/290) - 1]}{Y - 1} \ ,$$ (38)

where T_e is the effective input noise temperature (see Fig. 32) and T_H and T_C are the noise temperatures of a hot or cold input termination. To find the two points on the straight line of Fig. 32 one may switch between two input terminations at 373 K (100° C) and 77 K. For precise reading of RF power, calibrated piston attenuators in the IF path (Intermediate Frequency Superheterodyne Receiver) are in use. This is the hot/cold method. The difference between the Y-factor and the hot/cold method is that for the latter the input of the amplifier becomes physically connected to resistors at different temperatures (77, 373 K). For the Y-factor, the noise temperature of the input termination is varied by electronic means between 300 K and 12000 K (physical temperature always around 300 K).

As a variant of the 3-dB method with a controllable noise source, the excess noise temperature definition ($T_{ex} = T_H - T_C$) is often applied. A switchable 3-dB attenuator at the output of the amplifier just cancels the increase in noise power from $T_H - T_C$. Thus the influence of nonlinearities of the power meter is eliminated. To measure the noise of one port one may also use a calibrated spectrum analyser. However, spectrum analysers have high noise figures (20–40 dB) and the use of a low-noise preamplifier is recommended. This 'total power radiometer' [11] is not very sensitive but often sufficient, e. g., for observation of the Schottky noise of a charged particle beam. Note that the spectrum analyser may also be used for two-port noise figure measurements. An improvement of this 'total power radiometer' is the 'Dicke Radiometer' [11]. It uses a 1 kHz switch between the unknown one port and a controllable reference source. The reference source is made equal to the unknown via a feedback loop, and one
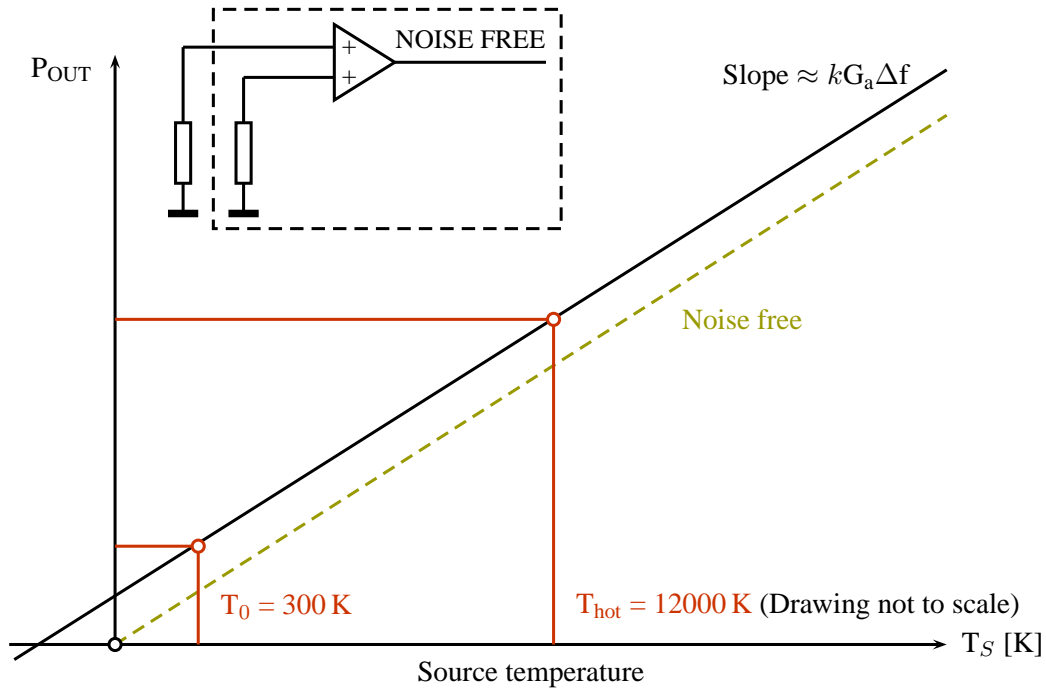
Fig. 32: Relation between source noise temperature T_s and output power P_{out} for an ideal (noise free) and a real amplifier [10, 15]

obtains a resolution of about 0.2 K. Unfortunately, switch spikes sometimes appear. Nowadays, switch-free correlation radiometers with the same performance are available [6].

The noise figure of a cascade of amplifiers is [6, 9, 10, 14, 15]

$$F_{total} = F_1 + \frac{F_2 - 1}{G_{a1}} + \frac{F_3 - 1}{G_{a1} G_{a2}} + \cdots \tag{39}$$

As can be seen from Eq. (39) the first amplifier in a cascade has a very important effect on the total noise figure, provided G_{a1} is not too small and F_2 is not too large. In order to select the best amplifier from a number of different units to be cascaded, one can use the noise measure M:

$$M = \frac{F - 1}{1 - (1/G_a)} \quad . \tag{40}$$

The amplifier with the smallest M has to be the first in the cascade [15].

References

[1] F. Caspers, RF engineering basic concepts: the Smith chart, CAS proceedings, 2010.

[2] G. D. Vendelin, A. M. Pavio and U. L. Rohde, Microwave Circuit Design Using Linear and Non-linear Techniques, second edition Wiley-Interscience, 2005, ISBN-10 0-471-41479-4

[3] M. Thumm, W. Wiesbeck and S. Kern, Hochfrequenzmesstechnik, B.G. Teubner Stuttgart · Leipzig, 1998, ISBN 3-519-16360-8

[4] R. A. Witte, Spectrum and Network Measurements, Prentice-Hall Inc., 1991, ISBN 0-13-826959-9

[5] H. D. Lueke, Signaliibertragung, Springer, 1975, ISBN 3-540-07125-3.

[6] B. Schiek, Messsysteme der Hochfrequenztechnik, Hiithig 1984, ISBN 3-7785-1045-2.

[7] K. Lipinski, Moderne Oszillographen, VDE, 1974.

[8] W. O. Schleifer, Hochfrequenz und Mikrowellenmesstechnik in der Praxis, Hiithig, 1981, ISBN 3-7785-0675-7.

[9] B. Schiek and H. J. Sieveris, Rauschen im Hochfrequenzschaltungen, Hiithig, 1984, ISBN 3-7785-2007-5.

[10] P. C. L. Yip, High frequency circuit design and measurement, Chapman and Hall, 1990, ISBN 0-412-34160-3.

[11] G. Evans and C. W. McLeisch, RF-Radiometer Handbook, Artech, 1977, ISBN 0-89006-055-X.

[12] F. R. Connor, Noise, Edward Arnold, 1973, ISBN 0-7131-3306-6.

[13] F. Landstorfer and H. Graf, Rauschprobleme der Nachrichtentechnik, Oldenbourg, 1981, ISBN 3-486-24681-X.

[14] O. Zinke and H. Brunswig, Lehrbuch der Hochfrequenztechnik, Zweiter Band, Springer, 1974, ISBN 3-540-06245-9.

[15] Fundamentals of RF and microwave noise figure measurements, Hewlett-Packard Application Note 57-1.