MEASUREMENTS

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 order to determine the noise temperature of a one-port and the noise figure of a two-port, basic physical relations and figures are given. A brief discussion of commonly used noise measurement techniques follows. In a subsequent section the fundamentals of the sampling technique, which has found widespread applications in "digital" oscilloscopes and sampling scopes, are discussed. Here the emphasis is mainly on hardware aspects. The spectrum analyzer has become an absolutely indispensable tool for rf signal analysis. As the rf section of modern spectrum analyzers has a rather complex structure the reasons for this complexity and certain working principles as well as limitations are discussed. The paper concludes with an overview of the development of scalar and vectorial network analyzers.

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 coaxial line 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 appearance of the vector network analyzer. At the same time the first sampling scopes with 1 GHz bandwidth came 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 analyzers as well as sensitive (low noise) receivers. A spectrum analyzer is in principle very similar to a common superheterodyne broadcast receiver, except for the requirements of choice of functions and change of parameters. Thirty years ago, instruments were set manually and had some sort of analogue 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 be always aware what is really going on in the analog section before all data are digitized. The basis of these analogue sections are discussed now.

2. NOISE MEASUREMENTS

2.1 Thermal Noise

The conception of "noise" was applied originally to the type of audible sound caused by statistical variations of the air pressure with a wide flat spectrum (white noise). It is now also applied to electrical signals, the noise "floor" determining the lower limit of signal transmission.

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

Different categories of noise can be defined as

White ,	which has a flat spectrum,
Pink ,	being low-pass filtered, and
Blue ,	being high-pass filtered.

In addition to the spectral distribution, the amplitude density distribution is also required in order to characterize a stochastic signal. For signals coming from very many independent sources the amplitude density has a Gaussian distribution.

The noise power density delivered to a load by a black body is given by Planck's formula

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

where N_L is the noise power delivered to a load,
$h = 6.625 \times 10^{-34}$ Js,
$k = 1.38056 \times 10^{-23}$ J/K.

Equation (1) 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 [1]

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

For $hf \ll kT$ the Rayleigh-Jeans approximation of Eq. (1) holds

$$N_L = kT\,\Delta f \tag{3}$$

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

$$\overline{u^2(t)} = 4kT\,R\Delta f \tag{4.1}$$

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

$$\overline{i^2(t)} = 4\frac{kT\,\Delta f}{R} = 4kT\,G\Delta f, \tag{4.2}$$

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. 1)

$$\frac{\overline{u^2(t)}}{4R} = kT\,\Delta f. \tag{4.3}$$

We define a spectral density function [2]

$$\begin{aligned} W_u(f) &= 4kTR \\ W_i(f) &= 4kTG \\ \overline{u^2(t)} &= \int_{f_1}^{f_2} W_u(f)\,df \end{aligned} \tag{4.4}$$

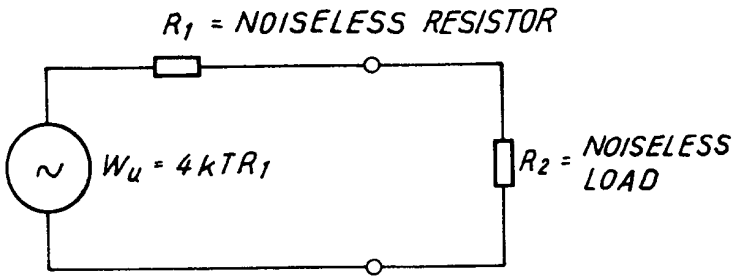

Fig. 1 Equivalent circuit for a noisy resistor R_1 terminated by a noiseless load R_2

Note: $u^2(t) = \int W_u(f)\,df$ = no load voltage.
When R_2 ($R_2 = R_1$) as external load is connected, a factor 2 voltage drop occurs.

A noisy resistor may be composed of many elements (resistive network). In general a resistor is made from many carbon grains (homogeneous temperatures). What happens for several resistors at different temperatures (inhomogeneous temperature distribution (Fig. 2))?

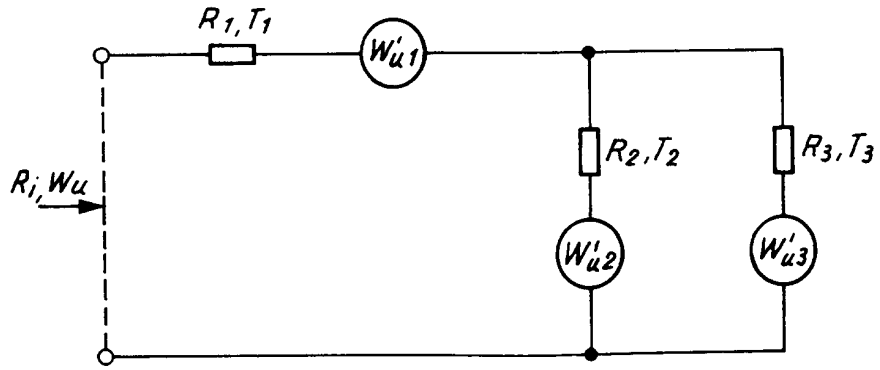

Fig. 2 Noisy one-port (resistors at different temperatures) [1, 2]

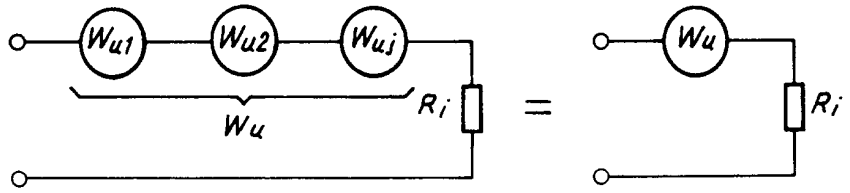

Fig. 3 Equivalent sources for the circuit of Fig. 2 [1, 2]

The total input impedance of the network (Fig. 2) is easy to work out as R_i [2]. Some caution is needed for the noise sources W_{uj}' to the input of the network (power divider).

Assume all W_{uj} are uncorrelated

$$W_u = \sum_j W_{uj} = 4kT_n R_i \tag{5}$$

$$T_n = \sum_j \beta_j T_j. \tag{6}$$

To visualize the physical meaning of the coefficients β_j consider a reciprocal network and feed some power to the input. Then β_j indicates the fractional part of this power dissipated in R_j. The relative contribution (β_j) of a lossy element to the total noise temperature T_n is equal to the relative dissipated power multiplied by its temperature [2].

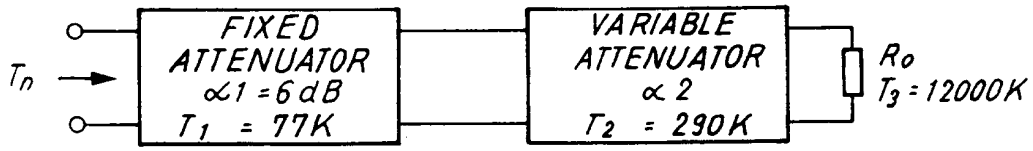

Fig. 4 Variable noise source [2]

$$T_n = \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3$$

Power at R_0 is $P\alpha_1\alpha_2 \rightarrow \beta_3 = \alpha_1\alpha_2$, $\beta_2 = \alpha_1(1 - \alpha_2)$, $\beta_1 = 1 - \alpha_1$, $\alpha_1 = 1/4$

$$T_n = (1 - \alpha_1)T_1 + \alpha_1 \cdot (1 - \alpha_2)T_2 + \alpha_1\alpha_2 T_3.$$

A similar calculation is possible for a directional antenna (satellite receiver). The noise temperature of (free) space amounts roughly to 3 K. But there are some losses in the atmosphere (air layer of 10 ... 20 km) causing a noise temperature at the antenna output of about 10 to 50 K which is still much less than 290 K.

So far, we have considered only pure resistors, but what about complex impedances? From the dissipation, we see that losses occur only in $Re(Z)$. The available noise power is independent of the magnitude of $Re(Z)$ with $Re(Z) > 0$. For Figs. 2 and 3, Eq.(5) still applies, except that R_i is replaced by $Re(Z_i)$. However, it must be remembered that in complex impedance networks the spectral power density W_u becomes frequency dependent [1].

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 [2]. 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, and noise temperatures of 50 K have been observed on a FET transistor at 290 K physical temperature.

2.2 Noise Figure Measurements

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 [3, 4]:

$$P_a = kT_0 \Delta f G_a. \tag{7}$$

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,

$$F = \frac{S_i / N_i}{S_0 / N_0}. \tag{8}$$

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})

$$F = \frac{N_{ad} + kT_0 \Delta f\, G_a}{kT_0 \Delta f\, G_a}. \tag{9}$$

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 (cf. variable noise source, see below) and measure the linear relation between amplifier output power and input termination noise temperature (T_s = T_{source}).

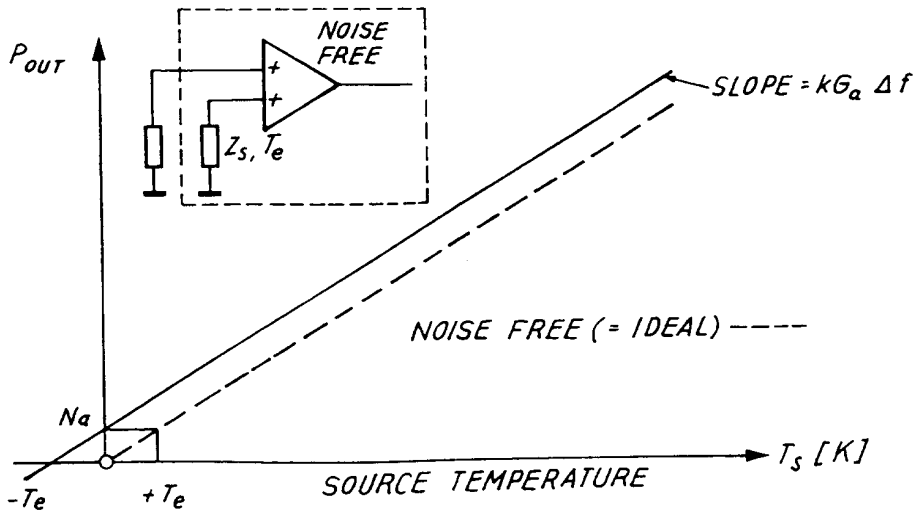

Fig. 5 Relation between source noise temperature T_S and output power P_{out} for an ideal and a real amplifier [3, 4]

In a similar way one has defined a factor Y as (for T_e = effective input noise temperature; see Fig. 5). T_H and T_c are the noise temperatures of a hot or cold input termination, respectively.

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

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

$$F = \frac{\left[(T_H/290) - 1\right] - Y\left[(T_c/290) - 1\right]}{Y - 1} \tag{10}$$

and with $T_o = T_c = 300$ K

$$F = \frac{T_H/T_0 - 1}{Y - 1} = \frac{T_{ex}}{T_0(Y - 1)}. \tag{11}$$

Some noise figure meters switch with a rate of about 1 kHz between hot and cold, i.e. they turn ON and OFF the "hot" source. The noise source is a specially designed Zener diode operated in the avalanche regime when "hot" [3]. We define an excess noise ratio (ENR) as 10 log$[(T_H\text{-}T_0)/T_0$. The ENR is usually around 16 dB. On modern instruments the evaluation of noise figure and gain is done quickly and automatically by an appropriate computer code. We just need to find two suitable points on the solid straight line in Fig. 5. This is the basis for most noise figure measurements.

The added noise power N_{ad} from the amplifier may be considered to come from an input termination at T_e (with the effective input noise temperature)

$$T_e = \frac{N_{ad}}{kG_a \Delta f}. \tag{12}$$

To find the two points on the straight line of Fig. 5 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 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 ≈ 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 $T_c - T_H$. 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 analyzer. However, spectrum analyzers have high noise figures (20...40 dB) and the use of a low-noise preamplifier is recommended. This "total power radiometer" [5] is not very sensitive but often sufficient, e.g. for observation of the Schottky noise of a charged particle beam. Note that the spectrum analyzer may also be used for two-port noise figure measurements. An improvement of this "total power radiometer" is the "Dicke Radiometer" [5]. It uses a 1 kHz switch between the unknown one port and a controllable reference source. The reference source via a feedback loop is made equal to the unknown, and one obtains a resolution of about 0.2 K. Unfortunately, switch spikes sometimes appear, nowadays switch-free correlation radiometers with the same performance are available [6].

The noise figure of a cascade of amplifiers is [1-4, 6]

$$F_{total} = F_1 + \frac{F_2 - 1}{G_{a1}} + \frac{F_3 - 1}{G_{a1} G_{a2}} + \cdots \tag{13}$$

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, use the noise measure M,

$$M = \frac{F - 1}{1 - (1/G_a)}. \tag{14}$$

The amplifier with the smallest M has to be the first in the cascade [4].

3. SAMPLING SCOPE

The bandwidth of real-time oscilloscopes is limited in most cases to 1 GHz. Special (expensive) instruments permit single-shot recording up to 6 GHz. For higher bandwidth on repetitive signals the sampling technique has been in use since about 1960.

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 $p(t)$ [7]

$$p(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nT_s) = \text{III}(t / T_s) \tag{15}$$

where the symbol III is derived from the Russian letter III and is pronounced "sha". It represents a series of δ-pulses.

The sampled time function $s_s(t)$ is

$$s_S(t) = s(t)p(t) = s(t)\text{III}(t / T_s)$$

$$S_s(f) = S(f) * \frac{1}{T_s}\text{III}(T_s f) \tag{16}$$

$$S_s(f) = \frac{1}{T_s} \sum_{n=-\infty}^{+\infty} S(f - mF) \quad \text{with } F = \frac{1}{T_s}. \tag{17}$$

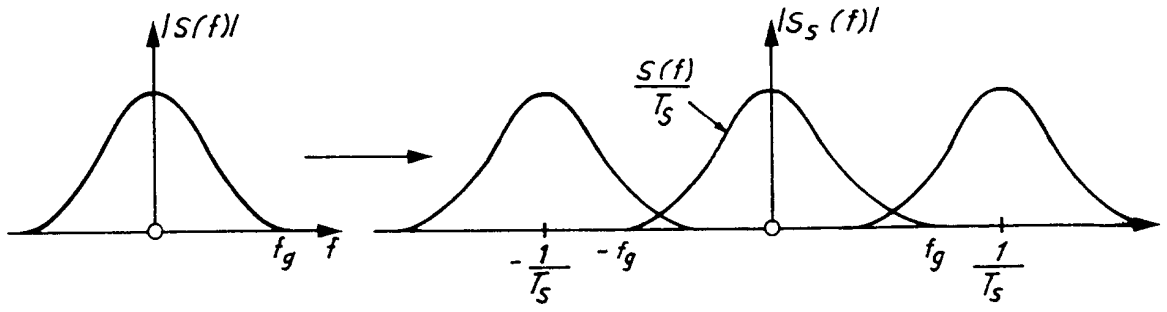

Fig. 6 Periodically repeated component of the Fourier Transform of $s_s(t)$ [7]

Note that the spectrum is repeated periodically by the sampling process. For proper reconstruction, one assures that overlapping as in Fig. 6 does not occur.

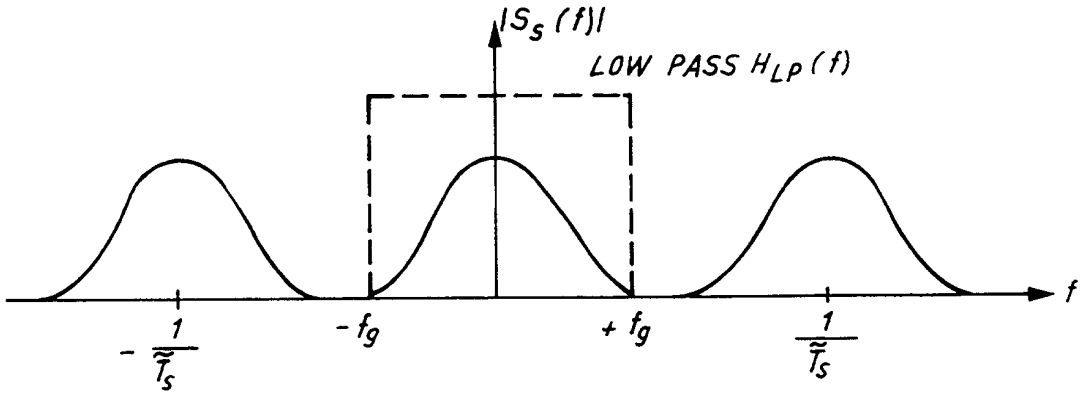

Fig. 7 Reconstruction of $S(f)$ via ideal lowpass from $S_S(f)$ (slightly oversampled)

If the spectra overlap as in Fig. 6 we have undersampling, the sampling rate is too low. If big gaps occur between the spectra (Fig. 7) 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. Nowadays (1991), with several ADCs in time-multiplex one obtains 8-bit vertical resolution at 2 *GSa/s* = 1 GHz bandwidth.

Another way to look at the sampling theorem (Nyquist) is to consider the sampling gate as a harmonic mixer.

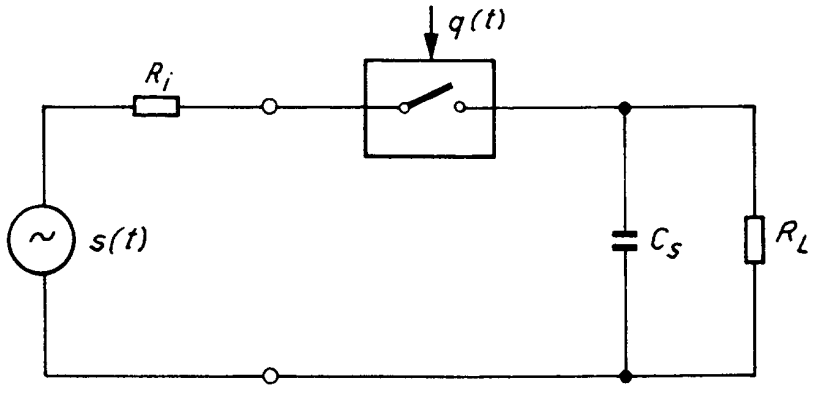

Fig. 8 Sampling gate as harmonic mixer; C_s = sampling capacitor [6]

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. 8 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. 9). The highest possible 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.

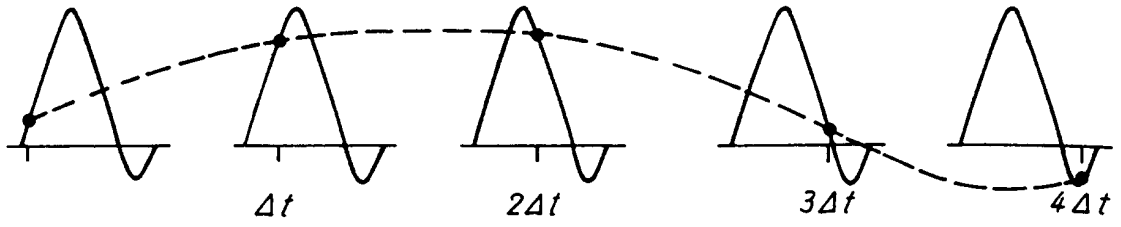

Fig. 9 Signal reconstruction with sampling shift by Δt per pulse [8]

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 10)

$$i_c(t) = \frac{s(t)}{R_i + R_d(t)}. \tag{18}$$

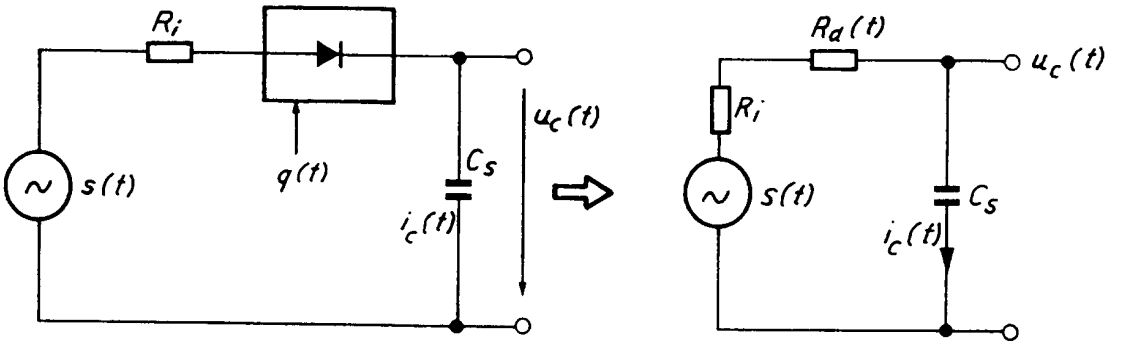

Fig. 10 Equivalent circuit for the sample-and-hold element [6]

After the sampling process we have [6]

$$u_c(t) = \frac{1}{C_s}\int_{-\infty}^{+\infty} i_c(t)\,dt = \int_{-\infty}^{+\infty} s(t)\frac{1}{C_s\left(R_i + R_d(t)\right)}dt \tag{19}$$

with the control signal for the Schottky diode being

$$q(t) = \frac{1}{C_s\left(R_i + R_d(t)\right)}. \tag{20}$$

The control or switching signal is moved by τ or $n\Delta t$ (Fig. 14) with respect to the signal to be sampled $s(t)$.

$$u_c(\tau) = \int_{-\infty}^{+\infty} s(t)q(t-\tau)\,dt. \tag{21}$$

Note that the time constant R_iC_s is much bigger than the length of $q(t)$. C_s is only charged to a fraction of $s(t)$.

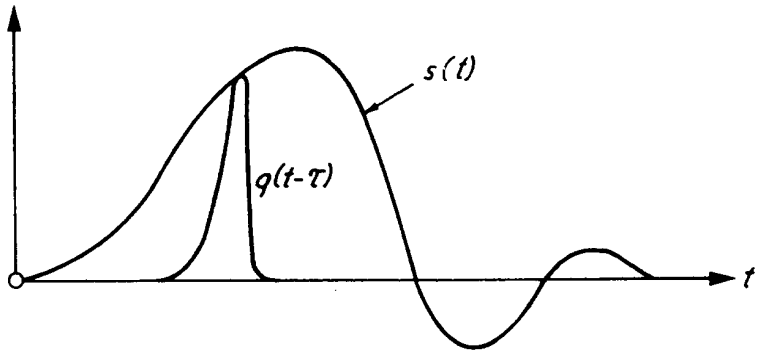

Fig. 11. Sampling with finite-width sampling pulse

The sampling efficiency η is defined as

$$\eta = \frac{u_c(\tau)}{s(\tau)}. \tag{22}$$

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.

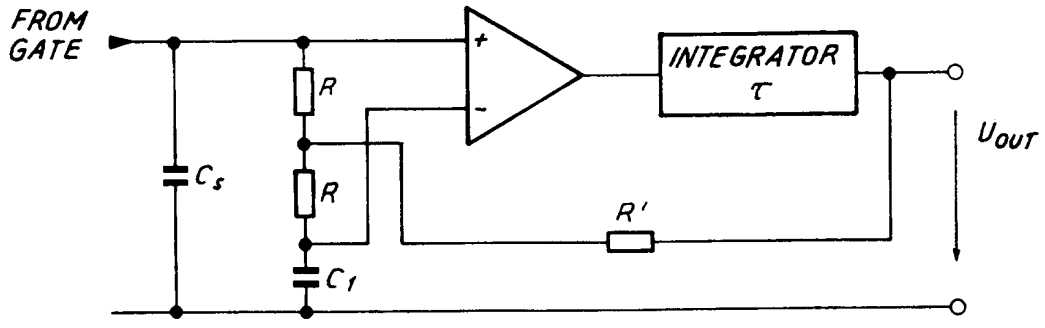

Fig. 12 Sampling-amplifier circuit [9]

The sampling gate is interesting from a technological point of view. As aperture times (Fig. 11) 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.

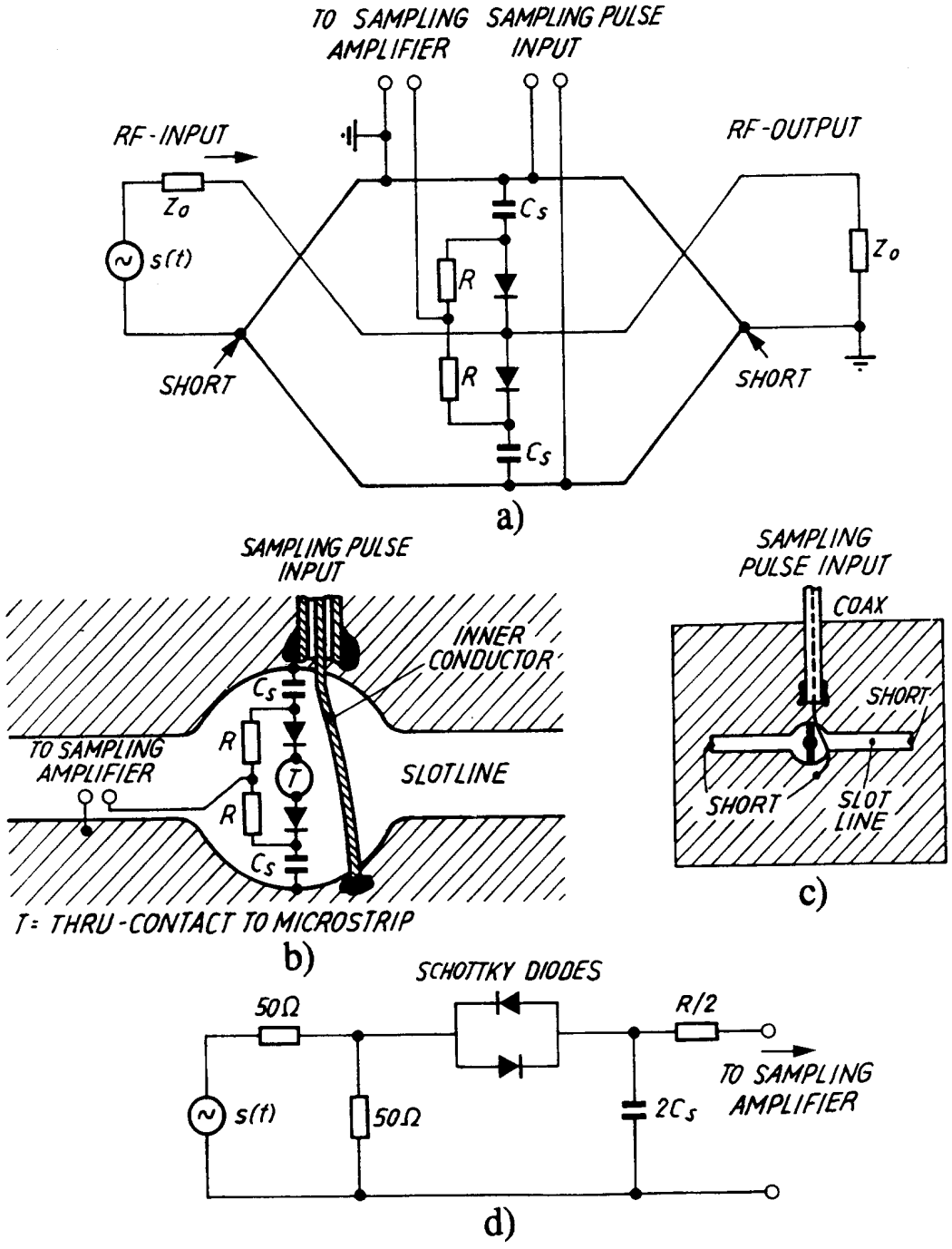

Fig. 13 Sampling circuits [8]
a) Simplified circuit of first sampling gate
b) Topology of sampling gate (detail)
c) Topology of sampling gate (overview)
d) Equivalent circuit to 13b, 13c

In MIC technology the sampling pulse is applied to a slotline in the ground-plane metallisation of a microstrip substrate (Fig. 13). This slot line has a length of some 10 mm and is shorted at both ends. With a voltage across the slotline the fast Schottky diodes open and connect the microstrip line via a through hole to the sampling capacitor C_s. Due to the particular topology of the circuit the signal line (microstrip) is decoupled from the sampling pulse line over a wide frequency range (Fig. 13).

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. 14).

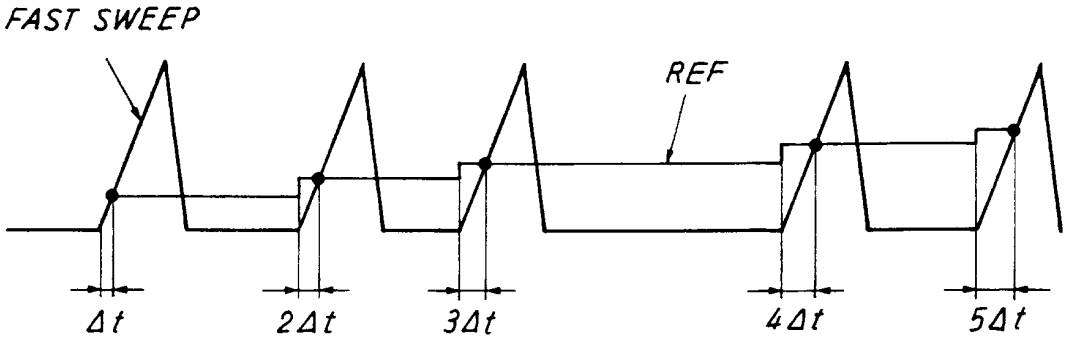

Fig. 14 Timing of sampling pulses [8]

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 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 25 years ago. It requires a strictly periodic signal rather than just the repetitive one for sequential sampling (Fig. 14). 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 [8].

Features of modern sampling scopes:

Rise time: 7 ps	$\hat{=}$ 50 GHz
Jitter	$\approx$ 1.5 ps
Static operation possible, no minimum repetition rate required.	
Superconducting sampling:	70 GHz
Optical sampling (mode-locked laser), 1 ps risetime	$\hat{=}$ 350 GHz

4. SPECTRUM ANALYZER

A common oscilloscope displays a signal in the amplitude-time plane (time domain). Another possible way to characterize a signal is a display in the amplitude-frequency plane. This type of instrument is called a spectrum analyzer. There are several ways to build a spectrum analyzer. For real-time analysis one may use a filter bank consisting of many slightly overlapping fixed-frequency tuned bandpass filters. Nowadays the fast Fourier transform (FFT) is applied at a real time spectrum analyzer on a digitized amplitude time trace. A periodic signal in the time-domain gives a number of discrete lines in the frequency domain. A single pulse returns a continuous spectrum.

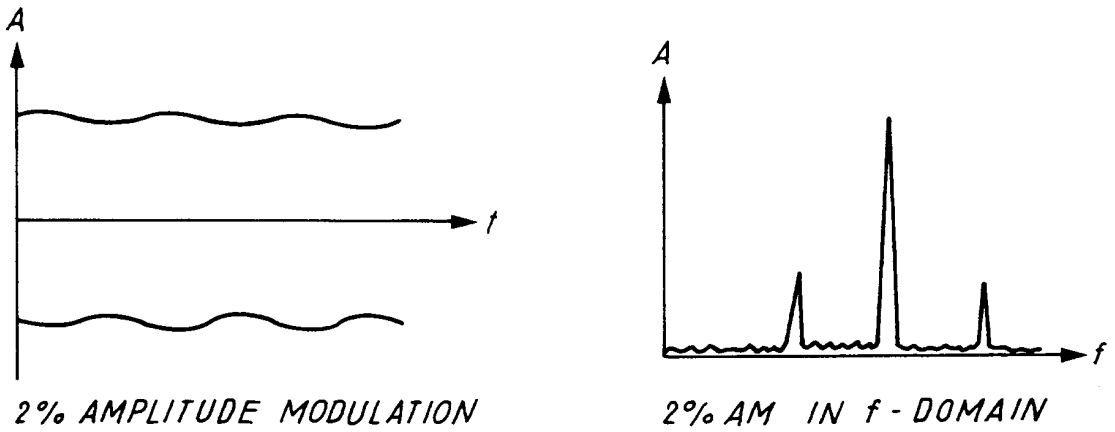

Fig. 15 Amplitude-modulated signal

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. 15) is very high (-120 dB below the main line). A very faint amplitude modulation (AM) of 10^{-12} (power) on some sinusoidal signals would be completely invisible on the time trace, but can be displayed as two sidelines 120 dB below the carrier in the frequency domain [11].

We will now consider only serial processing or swept tuned analyzers (Fig. 16).

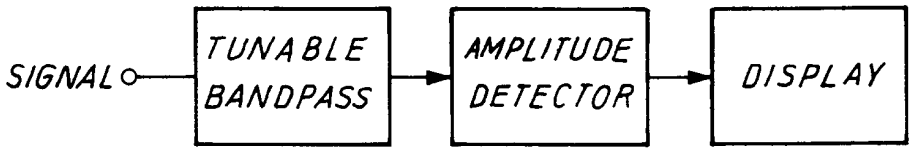

Fig. 16 A tunable bandpass as a simple spectrum analyzer (SPA)

The easiest way to design a swept tuned spectrum analyzer 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 1 000.

For much better performance the superheterodyne principle can be applied (Fig. 17).

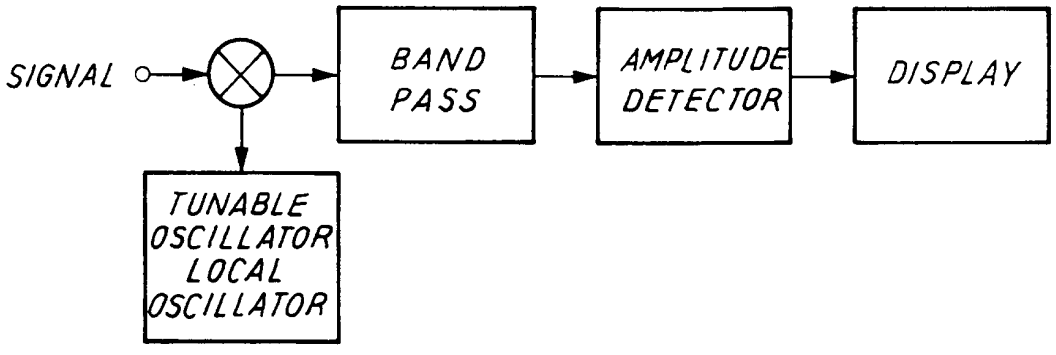

Fig. 17 Superheterodyne receiver [1]

The "weak" incident signal is subjected to nonlinear superposition (i.e. mixing or multiplication) with a "strong" sine wave from a local oscillator (LO). At the mixer output we then get the sum and difference frequencies of the signal and local oscillator. A fixed-frequency bandpass is better than a variable bandpass in terms of stability and selectivity. In fact Fig. 17 is the block diagram of any AM or FM broadcast receiver

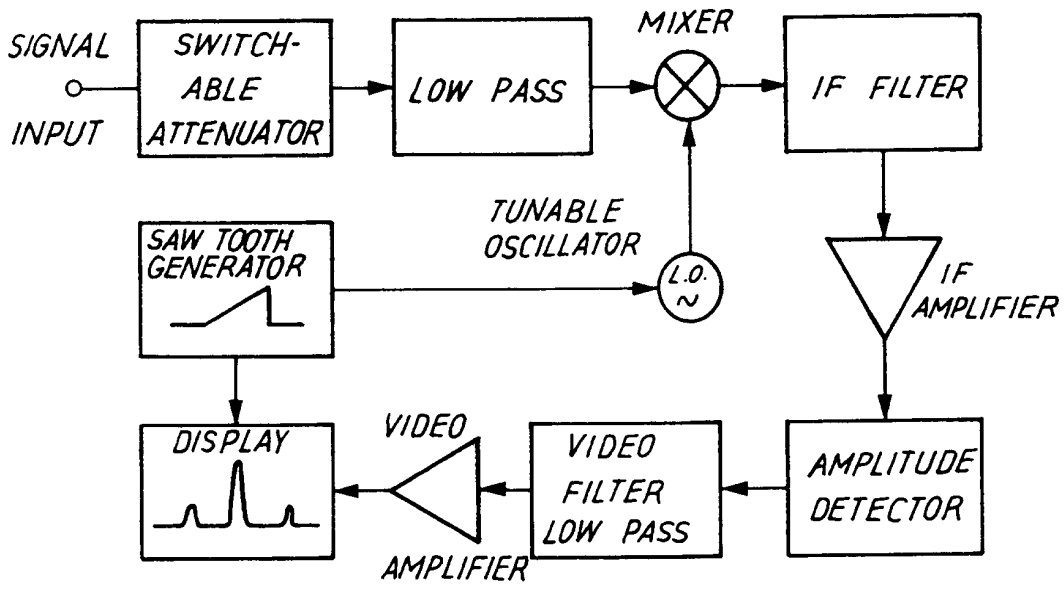

Fig. 18 Spectrum analyzer (block diagram) [12]

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

$$f_s = f_{LO} \pm f_{IF} \tag{23}$$

Assuming a signal range from 0 to 1 GHz for the spectrum analyzer depicted in Fig. 18 and f_{LO} between 2 and 3 GHz we get the frequency chart shown in Fig. 19.

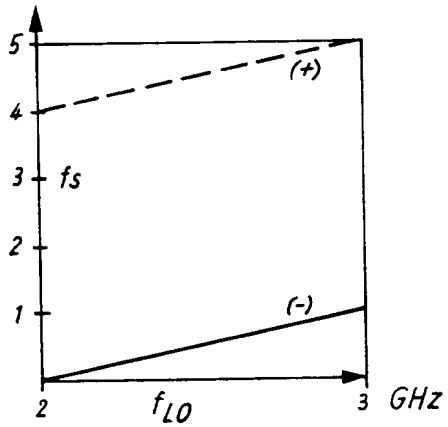

Fig. 19 Frequency chart of the SPA of Fig. 18, intermediate frequency = 2 GHz [11]

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 SPA's and high-quality receivers use more than one IF. Certain SPA's have four different LO's, 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.

The requirements of a modern SPA 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)

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

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 SPA's 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

$$f_s = Nf_{LO} \pm f_{IF} \quad n = 1, 2, 3, \ldots$$

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 analyzers 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 analyzed.

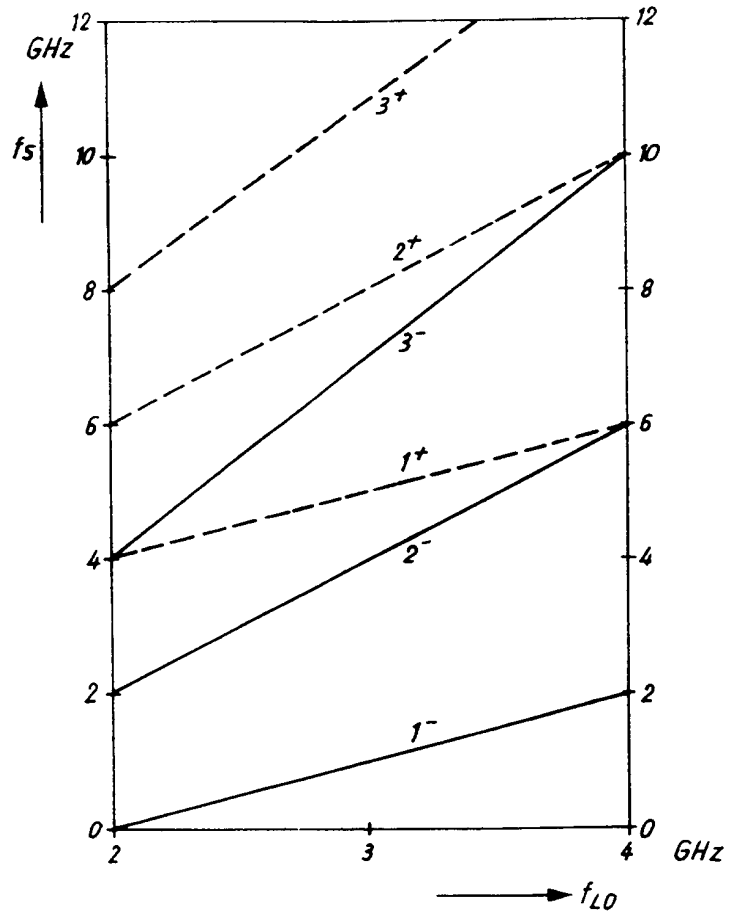

Fig. 20 Frequency chart for a SPA with harmonic mixing [11]

SPA's 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 SPA display are evaluated with methods discussed in Section 2 [2, 3, 5, 13, 14].

Spectrum analyzers can also be used to directly measure the phase noise of an oscillator provided that the LO phase noise in the SPA is much lower than that of the device under test [2]. For higher resolution, set-ups with delay lines and additional mixers (SPA at low frequencies or FFT) are advised.

5. SCALAR NETWORK ANALYZER

For many measurement problems it is sufficient to know only the modulus of a complex transmission or reflection coefficient. A very simple measurement set-up is depicted in Fig. 21 where DUT represents the device under test or equipment being measured.

The detector may be a Schottky diode or any other rf power measuring device such as a thermo-element producing a voltage between a junction of two metals at different temperatures, or a resistor with a high temperature-coefficient (thermistor). The small signal response of a Schottky diode detector (<-10 dBm) delivers an output voltage (U_2) proportional to the rf power.

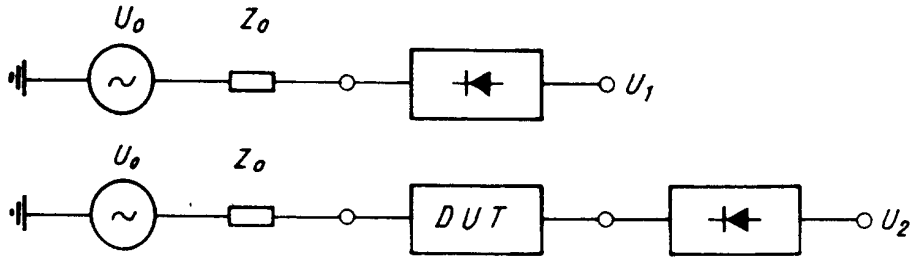

Fig. 21 Substitution measurement for the determination of insertion loss

A logarithmic amplifier often follows the diode detector to provide a reading in dB. Diode detectors may require some external resistors in order to be reasonably matched to 50 Ω on the rf side. As the signal strength U_0 of the rf generator is unlikely to be constant over a wide frequency range, either some levelling (feedback loop) or some other calibration measurements are needed.

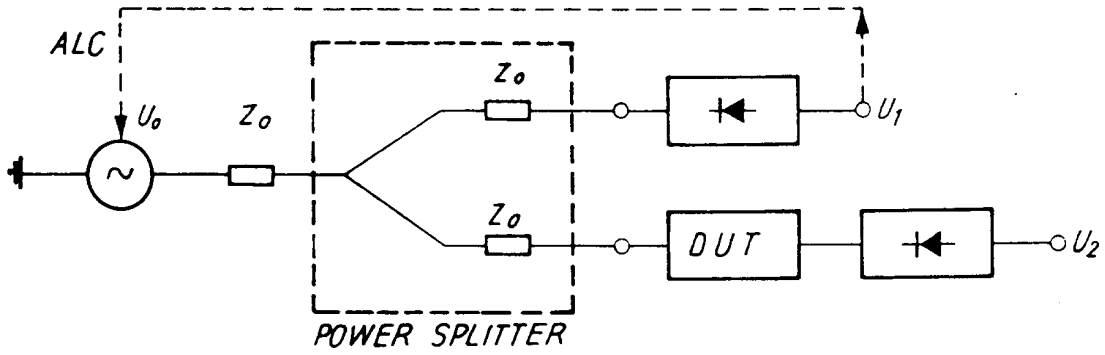

Fig. 22 Measurement of insertion loss with a power divider [14]
(ALC = Automatic Level Control)

A (resistive) power divider may provide a reference signal for any frequency usable for ALC or automatic level control, and the ratio U_2/U_1 cancels amplitude variations of the generator. This type of power divider (Fig. 22) is best suited to ALC rather than the symmetric $16^2/_3$ Ω version (Z_0 = 50 Ω). However, the symmetric resistive power divider has losses of 6 dB.

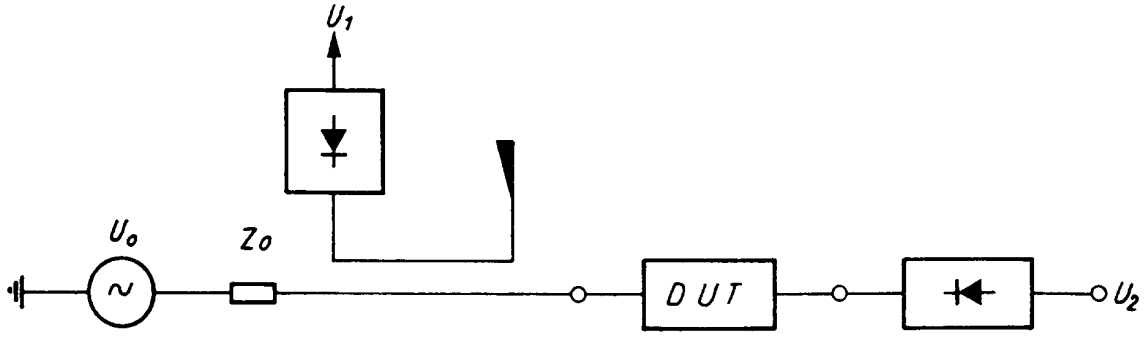

Fig. 23 Measurement of insertion loss with a directional coupler

Using a broadband directional coupler instead of a resistive power divider provides the same information as the resistive power divider, but with much less insertion loss (Fig. 23). However, the flatness of the frequency response of directional couplers may be worse than that of the resistive divider.

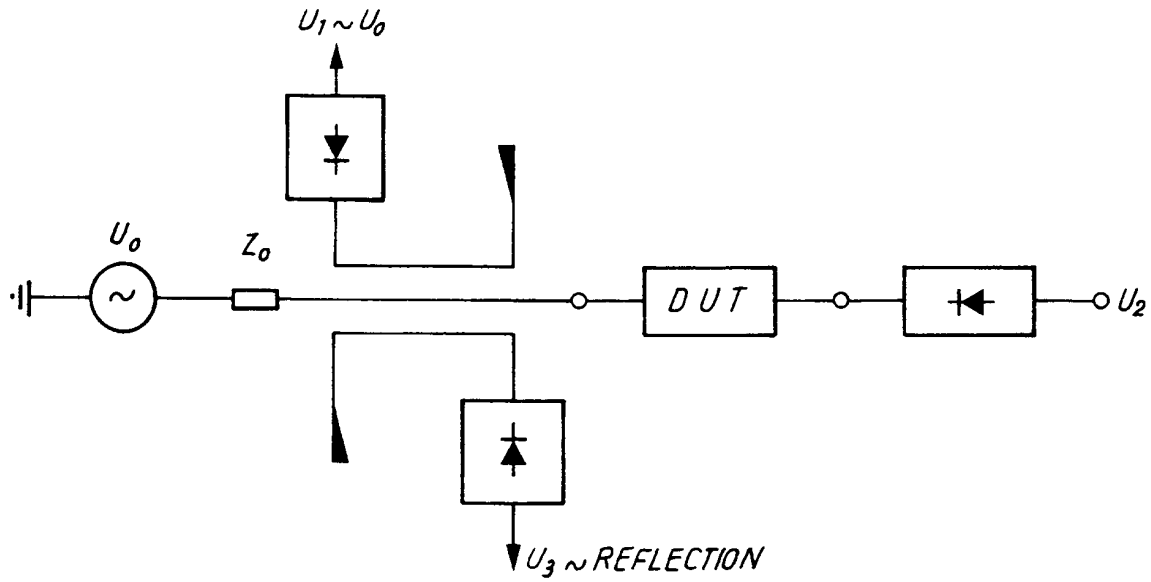

Fig. 24 Measurement of reflection with a dual directional coupler [3, 12]

So far, we have been looking at transmission properties only. For reflection measurements, a simple extension of the directional coupler, the bi-directional coupler, also provides a signal proportional to the reflected wave. With the set-up in Fig. 24 scalar data on the reflection coefficient S_{11} and the transmission coefficent S_{21} can be obtained. Only with power detectors is it possible to obtain the full vectorial information on the reflection coefficient. This configuration is known as a six-port reflectometer (Fig. 25) [6, 12, 16-18].

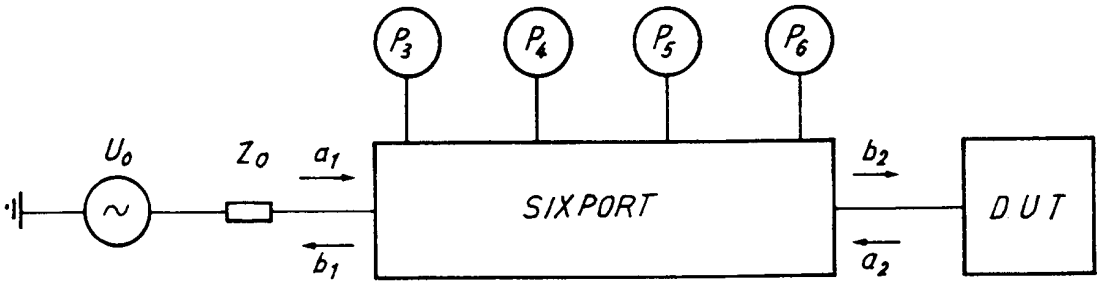

Fig. 25 Six-port reflectometer

It has been shown that, in order to obtain the vectorial information on S_{11} of the device under test, at least three power measurements are required. Usually, four power measurements are carried out (Fig. 26), and one may take advantage of the resulting redundancy.

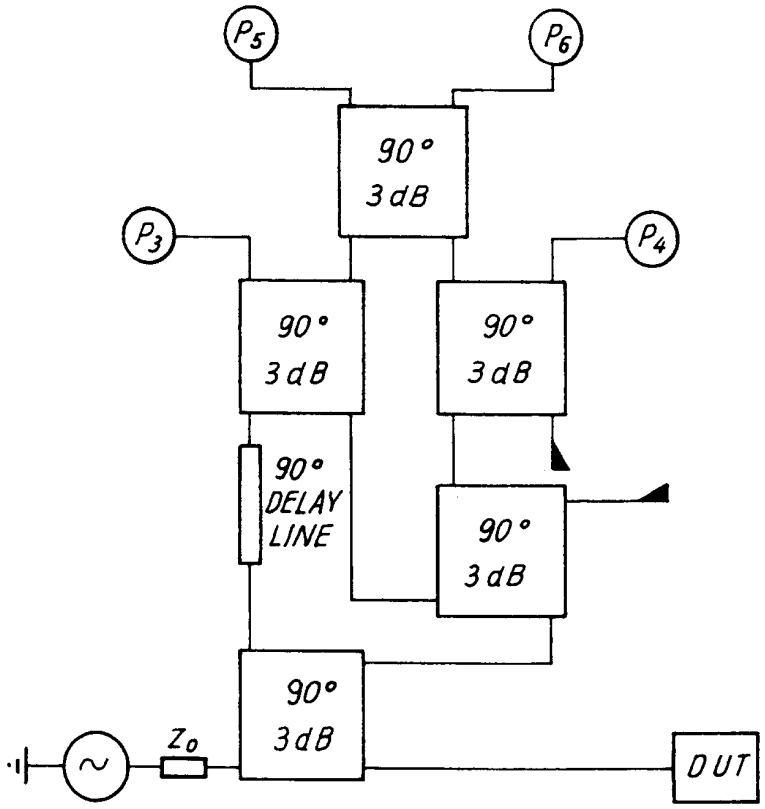

Fig. 26 Configuration of a six-port reflectometer [16]

Another way to measure the S-parameters at a fixed frequency is by using measurement bridges both in transmission and in reflection. For the transmission measurement a calibrated attenuator and phase shifter are set such that the reading of the detector becomes minimal or zero. The S_{21} of the device under test (Figs. 27, 28) then corresponds to the setting of the attenuator and the phase shifter.

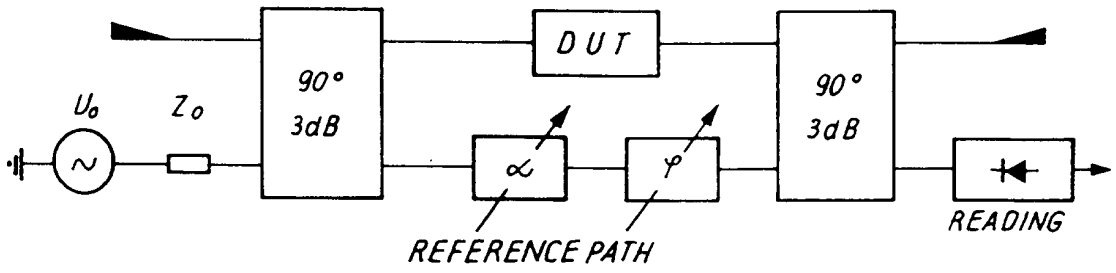

Fig. 27 Transmission measurement bridge [12]

In a similar way the reflection measurement bridge (Fig. 28) measures S_{11} from a variable attenuator and a variable short [6, 12].

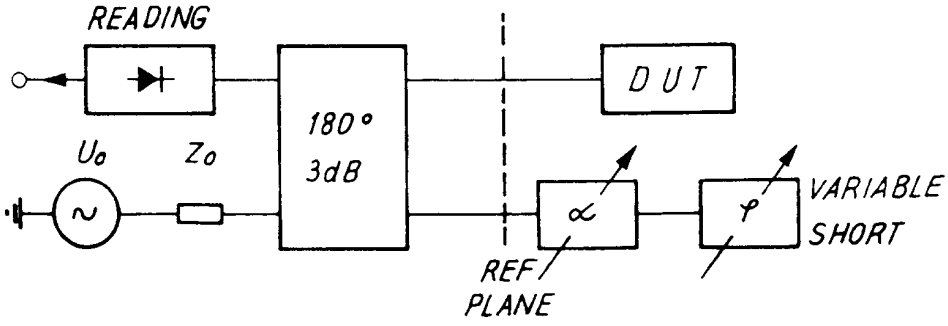

Fig. 28 Reflection measurement bridge

In practice the generator and the load (detector) are not perfectly matched. This leads to certain systematic measurement errors, but the effect of the mismatch may be removed by taking it into account by means of an error two-port (Fig. 29).

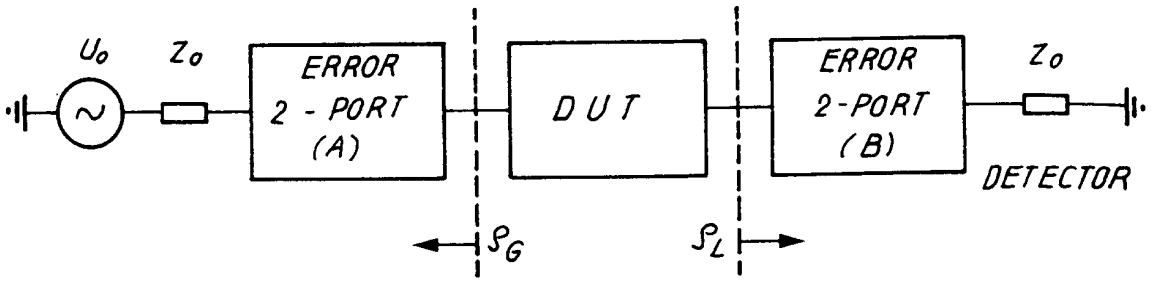

Fig. 29 Description of systematic errors using error two-ports [6]

The S-matrix of the error two-port on the generator side is often denoted as (A) and on the detector side as (B). As with all power detectors, no direct phase information is available, only certain calibration measurements may be carried out. They may not necessarily produce all the complex elements of the error two-ports depending on the calibration method employed.

For scalar network analyzers, a Fourier transform from the frequency into the time domain is generally not well defined, since the phase information is missing. However, some analyzers offer the possibility to apply an FFT on the frequency-domain data. These then make assumptions on the phase or carry out additional measurements to obtain indirect phase data.

Normally the RF signal in scalar network analyzers is modulated (ON, OFF) at about 20 kHz in order to eliminate DC drifts (flicker noise) of the detector diodes.

Power detectors do not discriminate between harmonics, e.g., from the generator and from the device under test. These harmonics may significantly alter the detector reading and to avoid this, a tracking (YIG) filter ahead of the detector is often applied.

6. VECTOR NETWORK ANALYZER

Having discussed scalar two-port measurement methods and spectrum analyzer techniques, it is just a small step to arrive at the vector-network analyzer (superheterodyne vector NWA). Where two rf generators are used. Both are variable in frequency (sweep, step mode), but are locked to each other via a phase-locked loop (PLL) such that their frequency

difference is kept constant and just equal to an IF. After mixing the IF contains the full phase and amplitude information of the rf signal to be measured (Fig. 30).

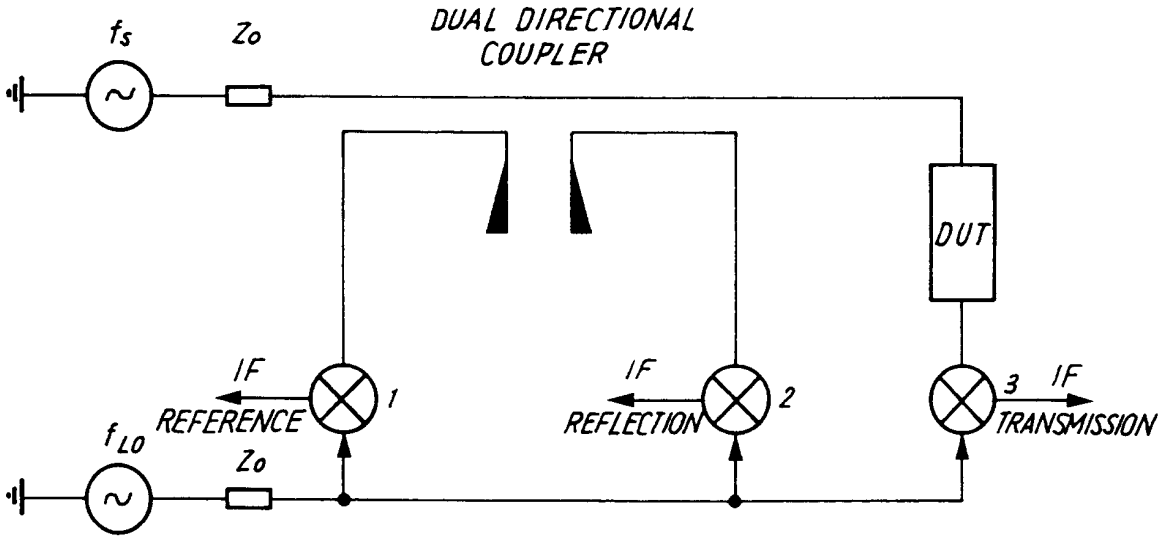

Fig. 30 Simplified circuit diagram of a superheterodyne vector NWA [3, 6]

For reasons similar to the case of the spectrum analyzer, a vector NWA has IF's at 20 MHz and 100 kHz. The IF around 20 MHz has a several MHz filter bandwidth in order to permit a fast sweep and to allow the PLL to lock in. For the second IF the filter bandwidth is often variable from a few kHz down to about 10 Hz and this is achieved by digital averaging.

The three mixers in Fig. 30 are characteristic of many NWA's and are in general well matched. Mixer 1 gives a signal to a PLL which is nearly independent of the device under test and permits the LO to be kept at the required frequency difference from the signal f_s. Mixer 2 delivers an IF signal with the information on the complex reflection coefficient of the testpiece, and Mixer 3 has the complex transmission information. The LO signals for the three mixers are "large", around 10 dBm, while the signals to be measured are "small" (<-10 dBm).

The phase and amplitude information is not lost during the double mixing process; the measurements being carried out at 100 kHz. A high quality logarithmic amplifier for the display in dB, and a phase sensitive detector measuring the phase of the second IF against a 100 kHz reference oscillator, provide the data that finally arrives at the CRT display. For proper definition of reference planes of the device under test, mechanical line extensions have been used in the past. But these are now replaced by electronic extensions consisting of another PLL and frequency dividers.

Similarly to spectrum analyzers, certain NWA also use harmonic mixers i.e., the LO signal has a comb spectrum or appears as a sequence of short Gaussian pulses in the time domain. Harmonic mixing increases the frequency range, but also the risk of undesired mixing products from certain comb frequencies.

The directional couplers and other components such as connectors, cables and mixers of the NWA are not perfect due to finite directivity, mismatch and cross talk. In order to remove these systematic errors several calibration methods have been introduced in the last 10-15 years [16].

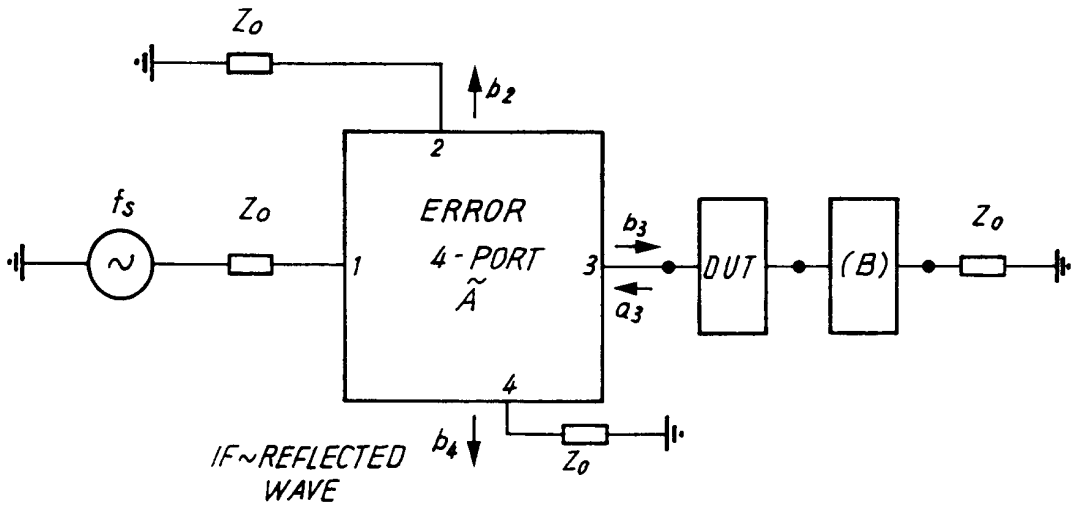

Fig. 31 Vector NWA representation via S-matrices

The situation is slightly different for a dual reflectometer (Fig. 32). In this case, the input and output reflection coefficients of the device under test as well as its forward and reverse transmission can be measured without disconnecting it.

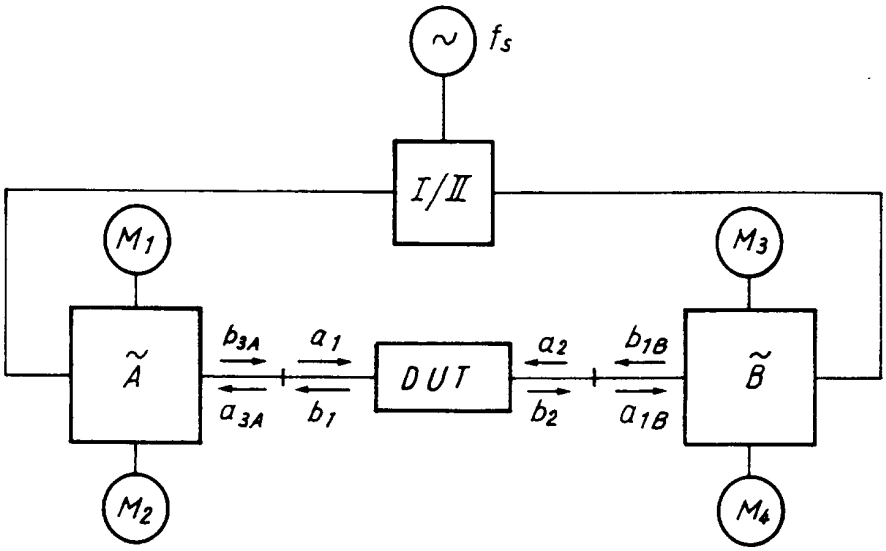

Fig. 32 Block diagram of a dual reflectometer [19]

For an error analysis the four-ports $\tilde{A}$ and $\tilde{B}$ can be reduced into two-ports A and B with an additional term for crosstalk (Fig. 33).

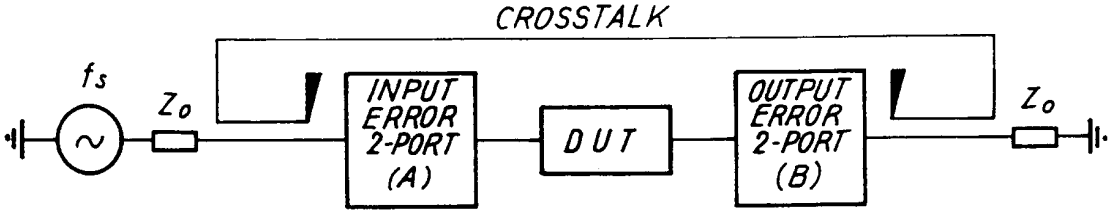

Fig. 33 Reduced-error model for vector NWA

The purpose of the calibration routines for vector NWA's is to determine the elements of A and B by measuring several calibration standards. Once the complex and frequency dependent elements and the crosstalk are known, the device under test is inserted and the measured data corrected. Then we have (without crosstalk) a cascade of three S-matrices, two of them known (A, B). This permits us to evaluate the unknown S-matrix of the testpiece. For a calibration measurement a very good open circuit, short circuit and load are often used, but there are many other methods with particular advantages and disadvantages (Fig. 34) [19, 20].

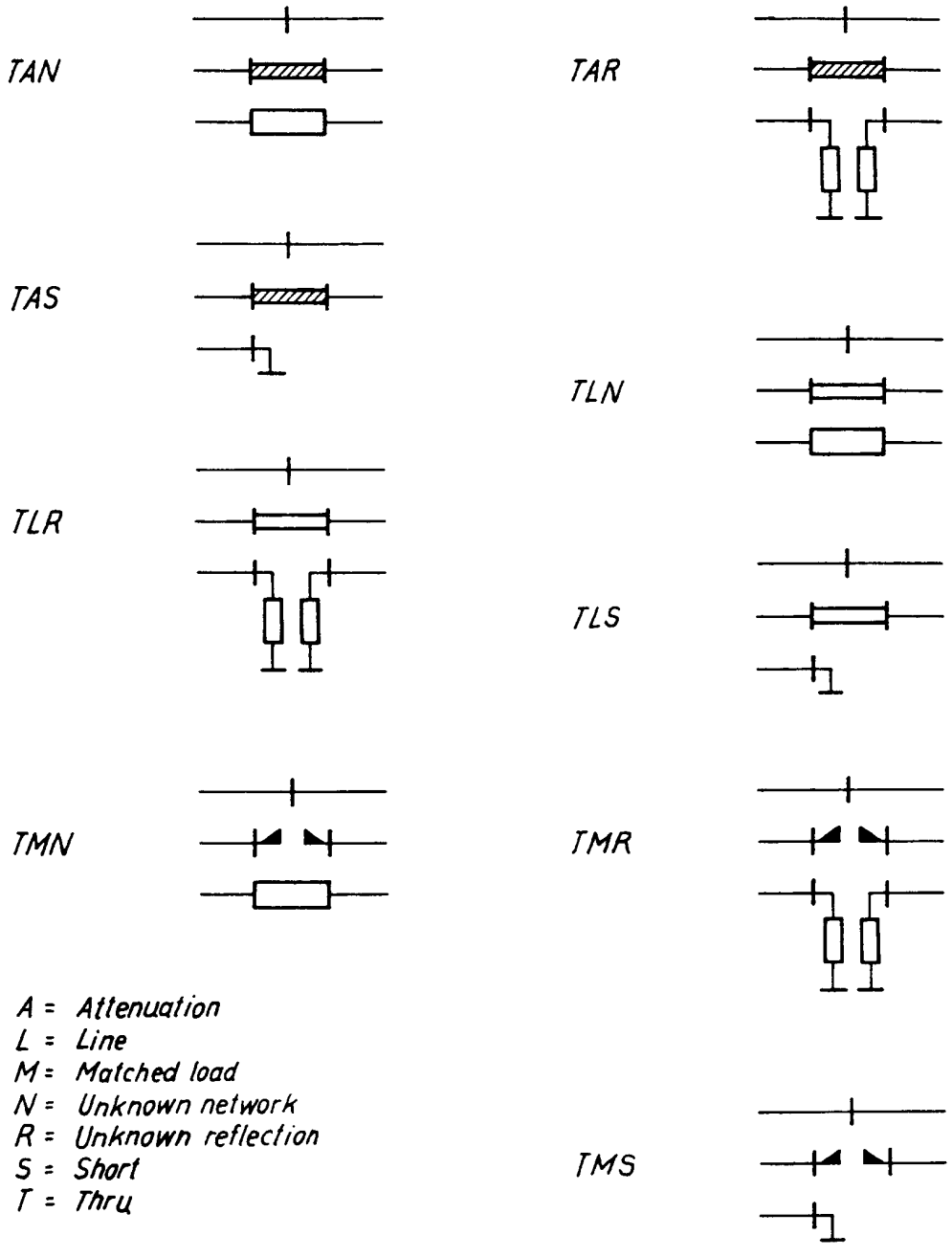

Fig. 34 Calibration routines [19]

Finally it should be mentioned that vector NWA's permit not only the evaluation of the frequency-domain reaponse for linear equipment but also the equivalent time-domain response.

In contrast to scalar NWA's, this transformation is meaningful and exact, as both amplitude and phase data are available.

In the time domain several interesting manipulations may be carried out (time filter) and the modified data can be transformed back into the frequency domain. This is a very efficient means to remove the effect of multiple reflections.

Certain vector NWA's offer the possibility of doing nonlinear circuit analysis by measuring higher-order harmonics. Since the circuitry of vector NWA's and spectrum analyzers are so closely related, dual function instruments (network-spectrum analyzers) are also on the market now.

7. HOMODYNE NETWORK ANALYZERS

The homodyne NWA does not have a frequency-locked tracking LO following the main (signal) oscillator. It generates its offset frequency by means of a single-sideband generator and uses a single-sideband detector as a receiver [6]. It is in a way the poor man's NWA, as the hardware is very cheap and does not need several IF's. However, the bandwidth is limited and the dynamic range lower than that of a superheterodyne NWA.

ACKNOWLEDGEMENTS

The author would like to thank E. Jensen, C. Taylor and S. Turner for stimulating discussions and for reading the manuscript, and B. Schiek for valuable hints.

REFERENCES

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

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

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

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

[5] G. Evans and C.W. McLeisch, RF-radiometer handbook, Artech 1977, ISBN 0-89006-055-X.

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

[7] H.D. Lüke, Signalübertragung, Springer, 1975, ISBN 3-540-07125-3.

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

[9] Hewlett-Packard Journal, Nov. 1971, author not known.

[10] U. Tietze and C. Schenk, Halbleiterschaltungstechnik, Springer 1986.

[11] W.D. Schleifer, Hochfrequenz und Mikrowellenmesstechnik in der Praxis, Hüthig 1981, ISBN 3-7785-0675-7.

[12] Meinke, Gundlach, Taschenbuch der Hochfrequenztechnik, 4. Auflage, Springer 1986, ISBN 3-540-15393-4.

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

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

[15] P. Gerdsen, Hochfrequenzmesstechnik Messgeräte und Messverfahren, Teubner 1982, ISBN 3-519-00092-X.

[16] J.F. Zürcher, M. Borgeaud and F.E. Gardiol, A compact portable six-port reflectometer, Mikrowellen Magazin Vol. 9, No. 2 (1983) 168.

[17] U. Stumper, Sechstorschaltungen zur Bestimmung von Streukoeffizientem, Mikrowellen Magazin, Vol. 9, No. 6 (1983) 669.

[18] G.F. Engen, A least squares solution for use in the six-port measurement technique, IEEE-MTT-28, No. 12 (Dec. 1980) 1473.

[19] H. Eul and B. Schiek, Experimental results of new self-calibration procedures for network analyzers, Proc. 20th European Microwave Conf., Budapest 1990, p. 1461.

[20] R.R. Pantoja, M.J. Howes, J.R. Richardson and R.D. Polland, Improved calibration and measurement of the scattering parameters of microwave integrated circuits, IEEE-MTT-28, No. 11 (Nov. 1989) 1675.