Fundamentals of Spectrum Analysis

Christoph Rauscher

ROHDE & SCHWARZ

Christoph Rauscher
Volker Janssen, Roland Minihold

Fundamentals of Spectrum Analysis

Mühldorfstrasse 15
81671 München
Germany
www.rohde-schwarz.com

Sixth edition 2008
Printed in Germany

ISBN 978-3-939837-01-5
PW 0002.6635.00

Table of Contents

Measurement Tips

1 Introduction

One of the most frequent measurement tasks in radiocommunications is the examination of signals in the frequency domain. Spectrum analyzers required for this purpose are therefore among the most versatile and widely used RF measuring instruments. Covering frequency ranges of up to 40 GHz and beyond, they are used in practically all applications of wireless and wired communication in development, production, installation and maintenance efforts. With the growth of mobile communications, parameters such as displayed average noise level, dynamic range and frequency range, and other exacting requirements regarding functionality and measurement speed come to the fore. Moreover, spectrum analyzers are also used for measurements in the time domain, such as measuring the transmitter output power of time multiplex systems as a function of time.

This book is intended to familiarize the uninitiated reader with the field of spectrum analysis. To understand complex measuring instruments it is useful to know the theoretical background of spectrum analysis. Even for the experienced user of spectrum analyzers it may be helpful to recall some background information in order to avoid measurement errors that are likely to be made in practice.

In addition to dealing with the fundamentals, this book provides an insight into typical applications such as phase noise and channel power measurements.

For further discussions of this topic, refer also to Engelson [1-1] and [1-2].

2 Signals

2.1 Signals displayed in time domain

In the time domain the amplitude of electrical signals is plotted versus time – a display mode that is customary with oscilloscopes. To clearly illustrate these waveforms, it is advantageous to use vector projection. The relationship between the two display modes is shown in Fig. 2-1 by way of a simple sinusoidal signal.

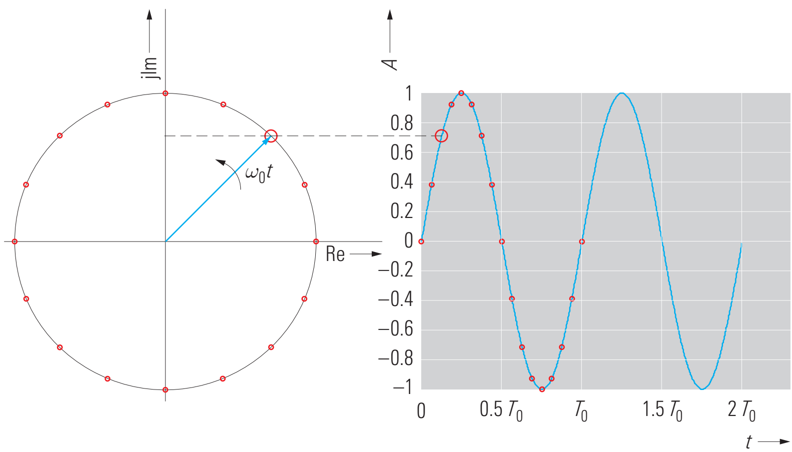


Fig. 2-1 Sinusoidal signal displayed by projecting a complex rotating vector on the imaginary axis

The amplitude plotted on the time axis corresponds to the vector projected on the imaginary axis (jIm). The angular frequency of the vector is obtained as:

$$\omega_0 = 2 \cdot \pi \cdot f_0 \qquad \textit{(Equation 2-1)}$$

where ω_0 angular frequency
f_0 signal frequency

A sinusoidal signal with $x(t) = A \cdot \sin(2 \cdot \pi \cdot f_0 \cdot t)$ can be described as $x(t) = A \cdot \mathrm{Im}\{e^{j \cdot 2\pi \cdot f_0 \cdot t}\}$.

2.2 Relationship between time and frequency domain

Electrical signals may be examined in the time domain with the aid of an oscilloscope and in the frequency domain with the aid of a spectrum analyzer (see Fig. 2-2).

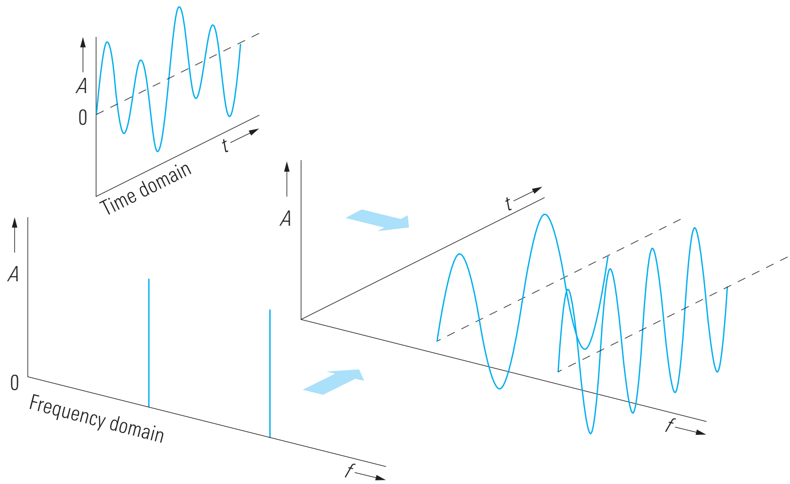


Fig. 2-2 Signals examined in time and frequency domain

The two display modes are related to each other by the Fourier transform (denoted F), so each signal variable in the time domain has a characteristic frequency spectrum. The following applies:

$$\underline{X}_{\mathrm{f}}(f) = F\left\{x(t)\right\} = \int_{-\infty}^{+\infty} x(t) \cdot \mathrm{e}^{-j2\pi ft}\,\mathrm{d}t \qquad \textbf{\textit{(Equation 2-2)}}$$

and

$$x(t) = F^{-1}\left\{\underline{X}_{\mathrm{f}}(f)\right\} = \int_{-\infty}^{+\infty} \underline{X}_{\mathrm{f}}(f) \cdot \mathrm{e}^{j2\pi ft}\,\mathrm{d}t \qquad \textbf{\textit{(Equation 2-3)}}$$

where $F\{x(t)\}$ Fourier transform of $x(t)$
$F^{-1}\{\underline{X}(f)\}$ inverse Fourier transform of $X(f)$
$x(t)$ signal in time domain
$\underline{X}_{\mathrm{f}}(f)$ complex signal in frequency domain

To illustrate this relationship, only signals with periodic response in the time domain will be examined first.

Periodic signals

According to the Fourier theorem, any signal that is periodic in the time domain can be derived from the sum of sine and cosine signals of different frequency and amplitude. Such a sum is referred to as a Fourier series. The following applies:

$$x(t) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cdot \sin(n \cdot \omega_0 \cdot t) + \sum_{n=1}^{\infty} B_0 \cdot \cos(n \cdot \omega_0 \cdot t) \quad \textbf{\textit{(Equation 2-4)}}$$

The Fourier coefficients A_0, A_n and B_n depend on the waveform of signal $x(t)$ and can be calculated as follows:

$$A_0 = \frac{2}{T_0} \int_0^{T_0} x(t)\,\mathrm{d}t \quad \textbf{\textit{(Equation 2-5)}}$$

$$A_n = \frac{2}{T_0} \int_0^{T_0} x(t) \cdot \sin(n \cdot \omega_0 \cdot t)\,\mathrm{d}t \quad \textbf{\textit{(Equation 2-6)}}$$

$$B_n = \frac{2}{T_0} \int_0^{T_0} x(t) \cdot \cos(n \cdot \omega_0 \cdot t)\,\mathrm{d}t \quad \textbf{\textit{(Equation 2-7)}}$$

where

$\frac{A_0}{2}$	DC component
$x(t)$	signal in time domain
n	order of harmonic oscillation
T_0	period
ω_0	angular frequency

Fig. 2-3b shows a rectangular signal approximated by a Fourier series. The individual components are shown in Fig. 2-3a. The greater the number of these components, the closer the signal approaches the ideal rectangular pulse.

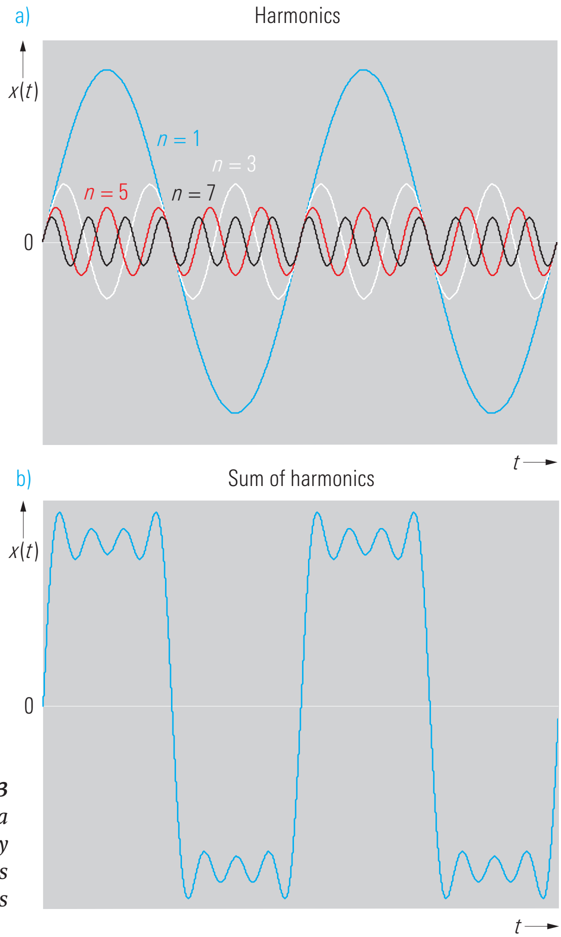


Fig. 2-3
Approximation of a rectangular signal by summation of various sinusoidal oscillations

In the case of a sine or cosine signal a closed-form solution can be found for Equation 2-2 so that the following relationships are obtained for the complex spectrum display:

$$F\left\{\sin\left(2\cdot\pi\cdot f_0\cdot t\right)\right\} = \frac{1}{j}\cdot\delta\left(f-f_0\right) = -j\cdot\delta\left(f-f_0\right)$$ ***(Equation 2-8)***

and

$$F\left\{\cos\left(2\cdot\pi\cdot f_0\cdot t\right)\right\} = \delta\left(f-f_0\right)$$ ***(Equation 2-9)***

where $\delta(f-f_0)$ is a Dirac function $\delta(f-f_0)=\infty$ if $f-f_0=0$, and $f=f_0$

$\delta(f-f_0)=0$, otherwise

$$\int_{-\infty}^{+\infty}\delta(f-f_0)\,\mathrm{d}f = 1$$

It can be seen that the frequency spectrum both of the sine signal and cosine signal is a Dirac function at f_0 (see also Fig. 2-5a). The Fourier transforms of sine and cosine signal are identical in magnitude, so that the two signals exhibit an identical magnitude spectrum at the same frequency f_0.

To calculate the frequency spectrum of a periodic signal whose time characteristic is described by a Fourier series in accordance with Equation 2-4, each component of the series has to be transformed. Each of these elements leads to a Dirac function, that is a discrete component in the frequency domain. Periodic signals therefore always exhibit discrete spectra which are also referred to as line spectra. Accordingly, the spectrum shown in Fig. 2-4 is obtained for the approximated rectangular signal of Fig. 2-3.

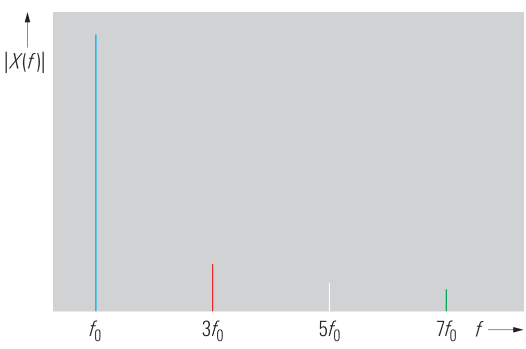


Fig. 2-4
Magnitude spectrum of approximated rectangular signal shown in Fig. 2-3

Fig. 2-5 shows some further examples of periodic signals in the time and frequency domain.

Non-periodic signals

Signals with a non-periodic characteristic in the time domain cannot be described by a Fourier series. Therefore the frequency spectrum of such signals is not composed of discrete spectral components. Non-periodic signals exhibit a continuous frequency spectrum with a frequency-dependent spectral density. The signal in the frequency domain is calculated by means of a Fourier transform (Equation 2-2).

Similar to the sine and cosine signals, a closed-form solution can be found for Equation 2-2 for many signals. Tables with such transform pairs can be found in [2-1].

For signals with random characteristics in the time domain, such as noise or random bit sequences, a closed-form solution is rarely found.

The frequency spectrum can in this case be determined more easily by a numeric solution of Equation 2-2.

Fig. 2-6 shows some non-periodic signals in the time and frequency domain.

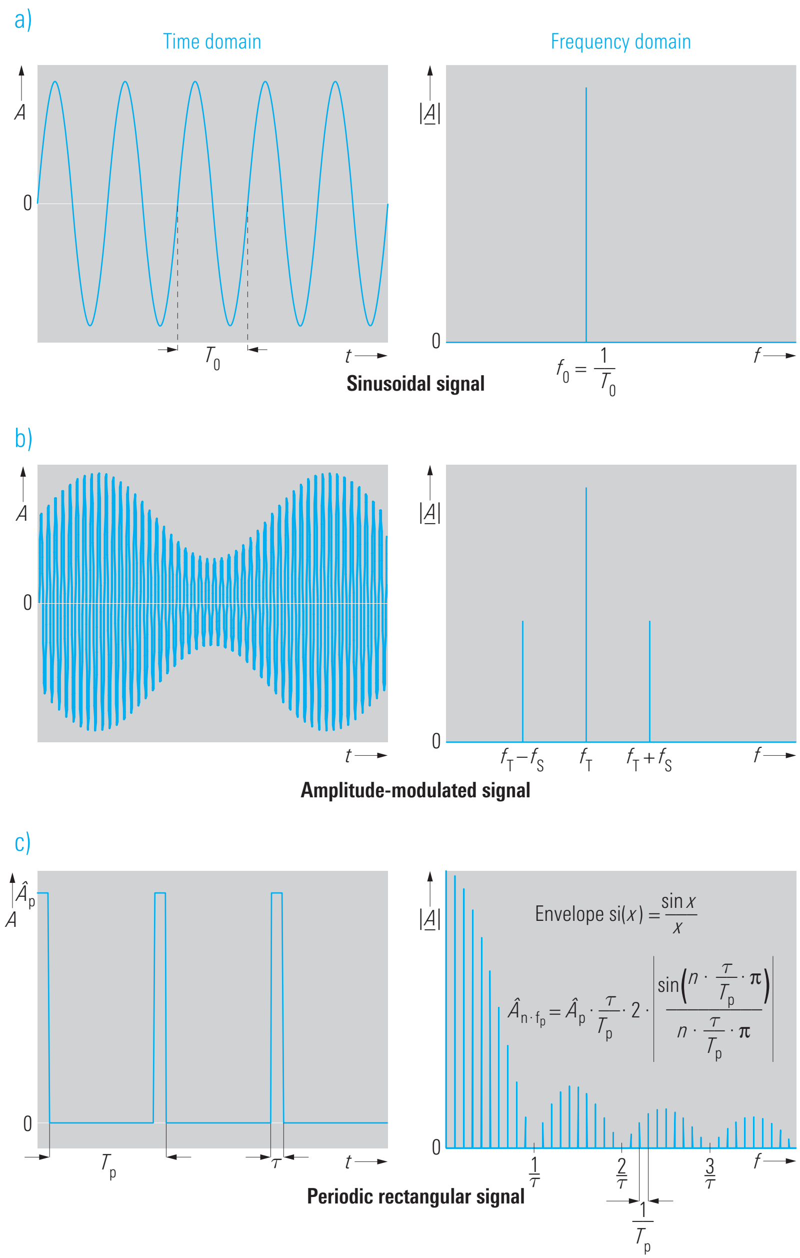


Fig. 2-5 Periodic signals in time and frequency domain (magnitude spectra)

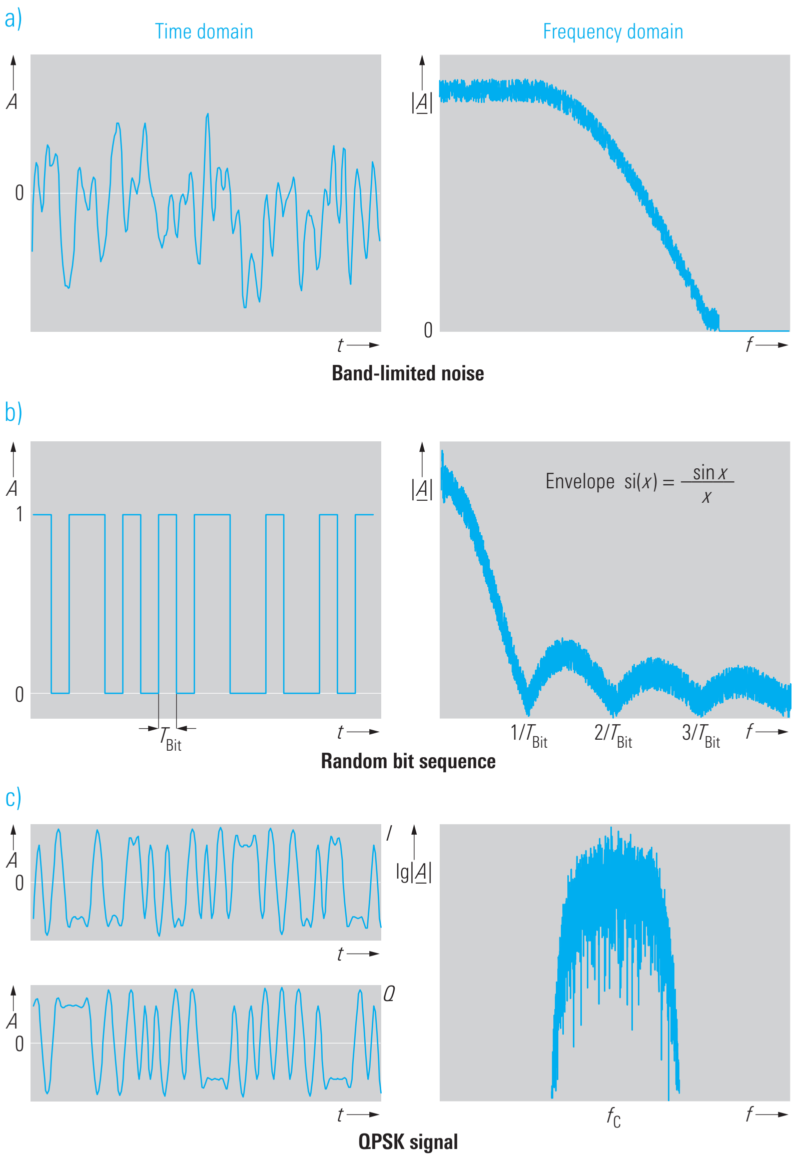


Fig. 2-6 Non-periodic signals in time and frequency domain

Depending on the measurement to be performed, examination may be useful either in the time or in the frequency domain. Digital data transmission jitter measurements, for example, require an oscilloscope. For determining the harmonic content, it is more useful to examine the signal in the frequency domain:

The signal shown in Fig. 2-7 seems to be a purely sinusoidal signal with a frequency of 20 MHz. Based on the above considerations one would expect the frequency spectrum to consist of a single component at 20 MHz.

On examining the signal in the frequency domain with the aid of a spectrum analyzer, however, it becomes evident that the fundamental (1st order harmonic) is superimposed by several higher-order harmonics i.e.multiples of 20 MHz (Fig. 2-8). This information cannot be easily obtained by examining the signal in the time domain. A practical quantitative assessment of the higher-order harmonics is not feasible. It is much easier to examine the short-term stability of frequency and amplitude of a sinusoidal signal in the frequency domain compared to the time domain (see also chapter 6.1 Phase noise measurement).

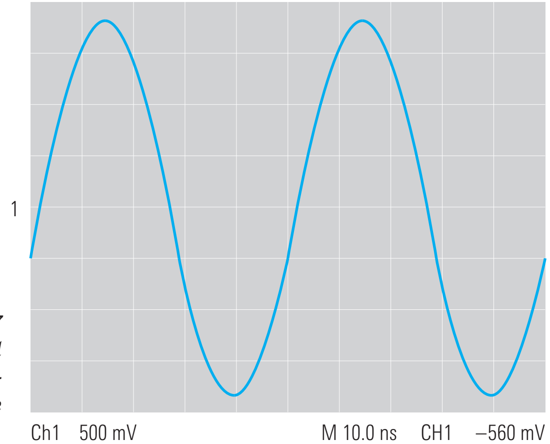


Fig. 2-7
Sinusoidal signal (f = 20 MHz) examined on oscilloscope

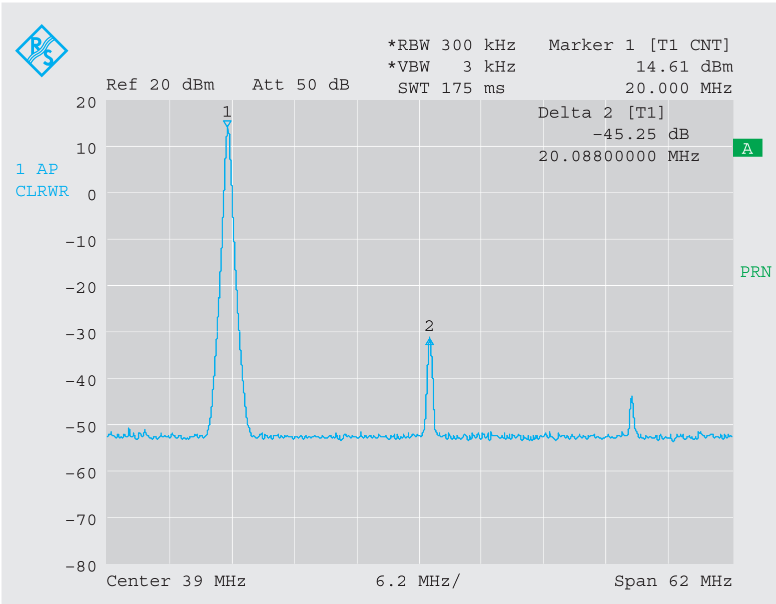


Fig. 2-8 *The sinusoidal signal of Fig. 2-7 examined in the frequency domain with the aid of a spectrum analyzer*

3 Configuration and Control Elements of a Spectrum Analyzer

Depending on the kind of measurement, different requirements are placed on the maximum input frequency of a spectrum analyzer. In view of the various possible configurations of spectrum analyzers, the input frequency range can be subdivided as follows:

- AF range — up to approx. 1 MHz
- RF range — up to approx. 3 GHz
- microwave range — up to approx. 40 GHz
- millimeter-wave range — above 40 GHz

The AF range up to approx. 1 MHz covers low-frequency electronics as well as acoustics and mechanics. In the RF range, wireless communication applications are mainly found, such as mobile communications and sound and TV broadcasting, while frequency bands in the microwave or millimeter-wave range are utilized to an increasing extent for broadband applications such as digital radio links.

Various analyzer concepts can be implemented to suit the frequency range. The two main concepts are described in detail in the following sections.

3.1 Fourier analyzer (FFT analyzer)

As explained in chapter 2, the frequency spectrum of a signal is clearly defined by the signal's time characteristic. Time and frequency domain are linked to each other by means of the Fourier transform. Equation 2-2 can therefore be used to calculate the spectrum of a signal recorded in the time domain. For an exact calculation of the frequency spectrum of an input signal, an infinite period of observation would be required. Another prerequisite of Equation 2-2 is that the signal amplitude should be known at every point in time. The result of this calculation would be a continuous spectrum, so the frequency resolution would be unlimited.

It is obvious that such exact calculations are not possible in practice. Given certain prerequisites, the spectrum can nevertheless be determined with sufficient accuracy.

In practice, the Fourier transform is made with the aid of digital signal processing, so the signal to be analyzed has to be sampled by an analog-digital converter and quantized in amplitude. By way of sampling the continuous input signal is converted into a time-discrete signal and the information about the time characteristic is lost. The bandwidth of the input signal must therefore be limited or else the higher signal frequencies will cause aliasing effects due to sampling (see Fig. 3-1). According to Shannon's law of sampling, the sampling frequency f_S must be at least twice as high as the bandwidth B_{in} of the input signal. The following applies:

$$f_S \geq 2 \cdot B_{in} \quad \text{and} \quad f_S = \frac{1}{T_S} \qquad \textit{(Equation 3-1)}$$

where f_S sampling rate
B_{in} signal bandwidth
T_S sampling period

For sampling lowpass-filtered signals (referred to as lowpass signals) the minimum sampling rate required is determined by the maximum signal frequency $f_{in,max}$. Equation 3-1 then becomes:

$$f_S \geq 2 \cdot f_{in,max} \qquad \textit{(Equation 3-2)}$$

If $f_S = 2 \cdot f_{in,max}$, it may not be possible to reconstruct the signal from the sampled values due to unfavorable sampling conditions. Moreover, a lowpass filter with infinite skirt selectivity would be required for band limitation. Sampling rates that are much greater than $2 \cdot f_{in,max}$ are therefore used in practice.

A section of the signal is considered for the Fourier transform. That is, only a limited number N of samples is used for calculation. This process is called windowing. The input signal (see Fig. 3-2a) is multiplied with a specific window function before or after sampling in the time domain. In the example shown in Fig. 3-2, a rectangular window is used (Fig. 3-2b). The result of multiplication is shown in Fig. 3-2c.

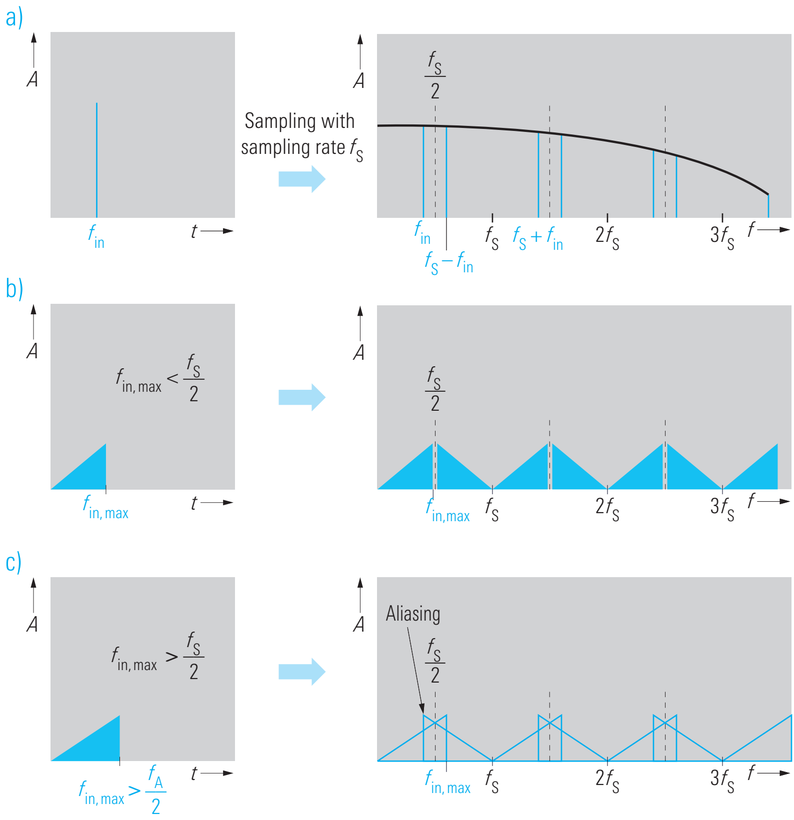


Fig. 3-1 Sampling a lowpass signal with sampling rate f_S a), b) $f_{in, max} < f_S/2$, c) $f_{in, max} > f_S/2$, therefore ambiguity exists due to aliasing

The calculation of the signal spectrum from the samples of the signal in the time domain is referred to as a discrete Fourier transform (DFT). Equation 2-2 then becomes:

$$\underline{X}(k) = \sum_{n=0}^{N-1} \underline{x}(nT_S) \cdot e^{-j2\pi kn/N} \qquad \textbf{\textit{(Equation 3-3)}}$$

where k index of discrete frequency bins, where $k = 0, 1, 2, \ldots$

n index of samples

$\underline{x}(nT_S)$ samples at the point $n \cdot T_S$, where $n = 0, 1, 2, \ldots$

N length of DFT, i.e. total number of samples used for calculation of Fourier transform

The result of a discrete Fourier transform is again a discrete frequency spectrum (see Fig. 3-2d). The calculated spectrum is made up of individual components at the frequency bins which are expressed as:

$$f(k) = k \cdot \frac{f_S}{N} = k \cdot \frac{1}{N \cdot T_S} \qquad \textit{(Equation 3-4)}$$

where $f(k)$ discrete frequency bin
k index of discrete frequency bins, where $k = 0, 1, 2 \ldots$
f_S sampling frequency
N length of DFT

It can be seen that the resolution (the minimum spacing required between two spectral components of the input signal for the latter being displayed at two different frequency bins $f(k)$ and $f(k+1)$ depends on the observation time $N \cdot T_S$. The required observation time increases with the desired resolution.

The spectrum of the signal is periodicized with the period f_S through sampling (see Fig. 3-1). Therefore, a component is shown at the frequency bin $f(k = 6)$ in the discrete frequency spectrum display in Fig. 3-2d. On examining the frequency range from 0 to f_S in Fig. 3-1a, it becomes evident that this is the component at $f_S - f_{in}$.

In the example shown in Fig. 3-2, an exact calculation of the signal spectrum was possible. There is a frequency bin in the discrete frequency spectrum that exactly corresponds to the signal frequency. The following requirements have to be fulfilled:

- the signal must be periodic (period T_0)
- the observation time $N \cdot T_S$ must be an integer multiple of the period T_0 of the signal.

These requirements are usually not fulfilled in practice so that the result of the Fourier transform deviates from the expected result. This deviation is characterized by a wider signal spectrum and an amplitude error. Both effects are described in the following.

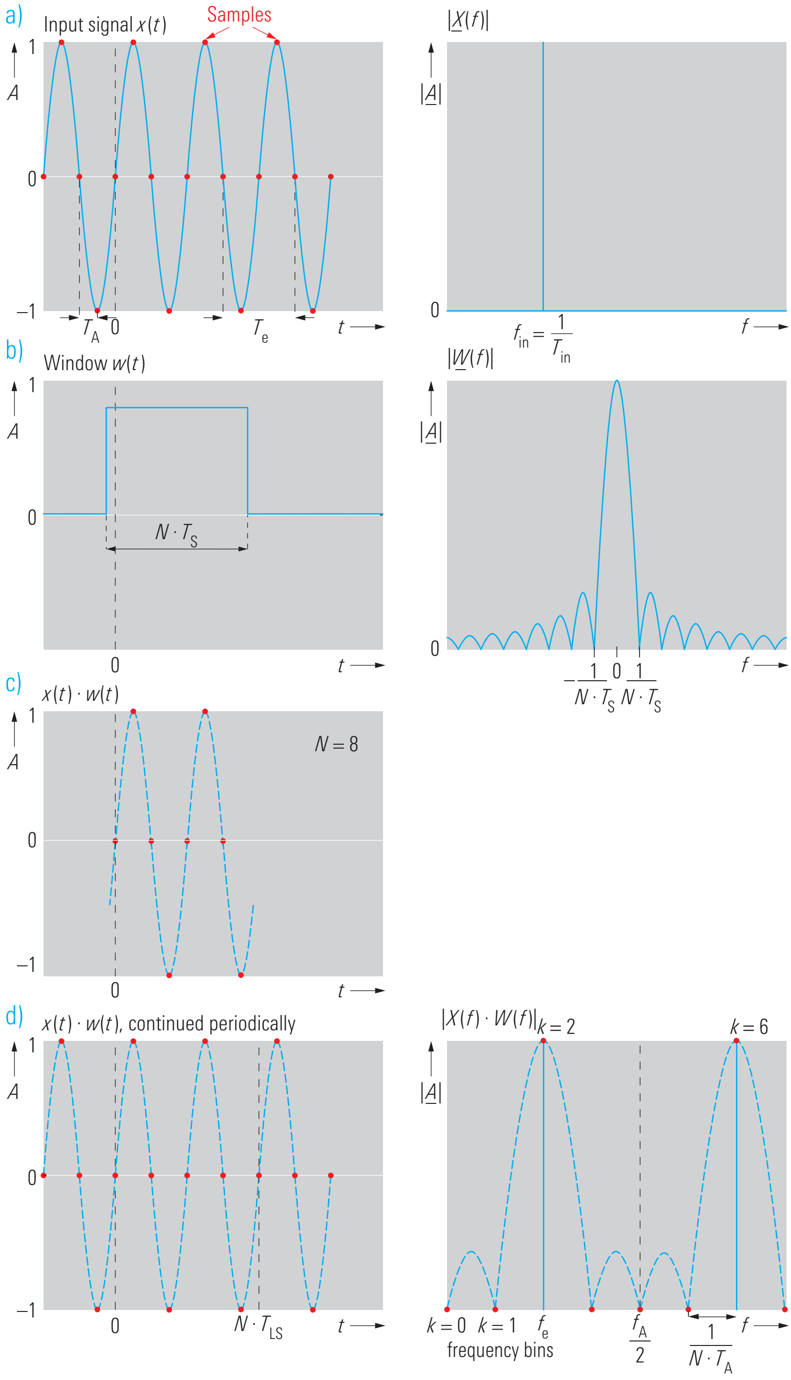


Fig. 3-2 DFT with periodic input signal. Observation time is an integer multiple of the period of the input signal

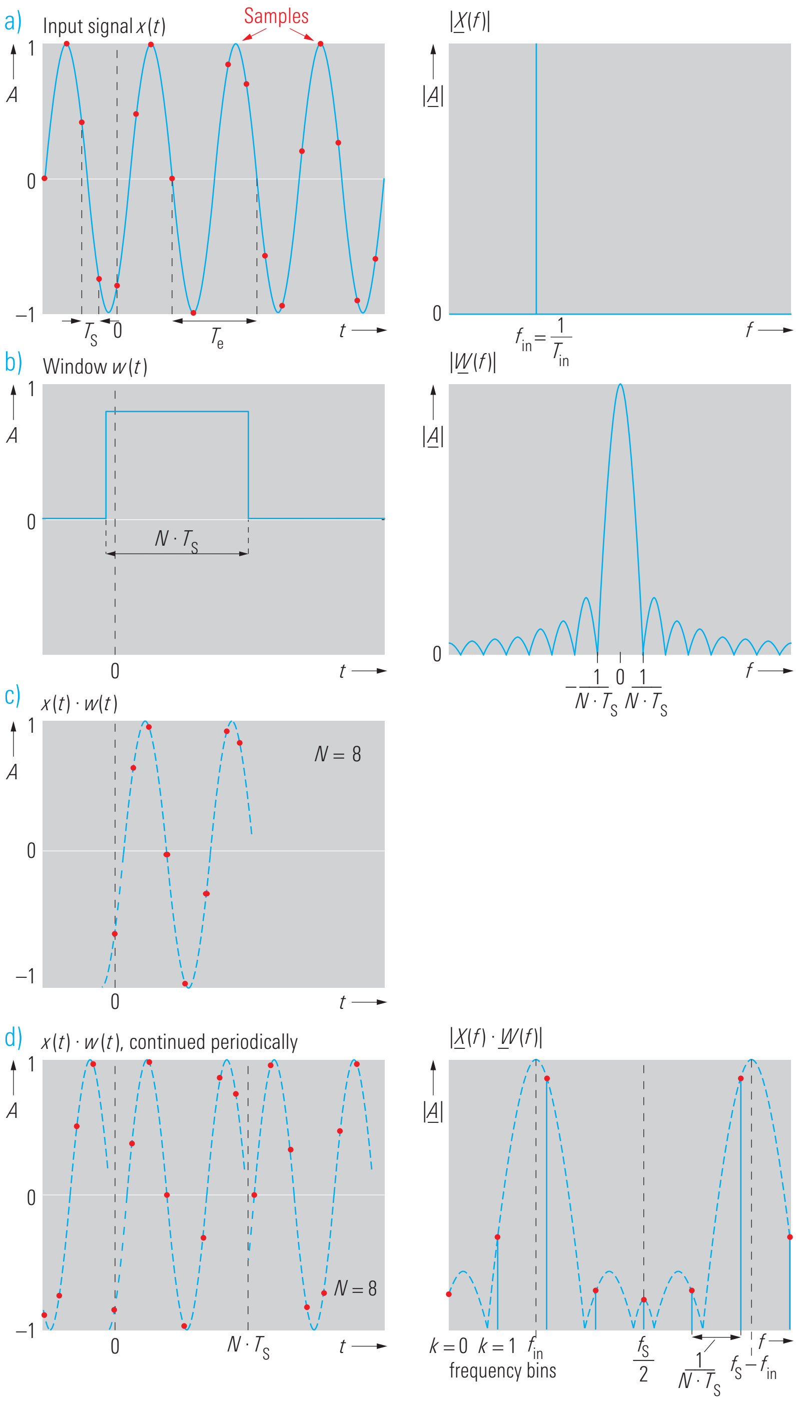


Fig. 3-3 DFT with periodic input signal. Observation time is not an integer multiple of the period of the input signal

The multiplication of input signal and window function in the time domain corresponds to a convolution in the frequency domain (see [2-1]). In the frequency domain the magnitude of the transfer function of the rectangular window used in Fig. 3-2 follows a sine function:

$$\left|W(f)\right| = N \cdot T_S \cdot \operatorname{si}\left(2\pi f \cdot N \cdot T_S/2\right) = N \cdot T_S \cdot \frac{\sin\left(2\pi f \cdot N \cdot T_S/2\right)}{2\pi f \cdot N \cdot T_S/2}$$

(Equation 3-5)

where $W(f)$ windowing function in frequency domain
$N \cdot T_S$ window width

In addition to the distinct secondary maxima, nulls are obtained at multiples of $1/(N \cdot T_S)$. Due to the convolution by means of the window function the resulting signal spectrum is smeared, so it becomes distinctly wider. This is referred to as leakage effect.

If the input signal is periodic and the observation time $N \cdot T_S$ is an integer multiple of the period, there is no leakage effect of the rectangular window since, with the exception of the signal frequency, nulls always fall within the neighboring frequency bins (see Fig. 3-2d).

If these conditions are not satisfied, which is the normal case, there is no frequency bin that corresponds to the signal frequency. This case is shown in Fig. 3-3. The spectrum resulting from the DFT is distinctly wider since the actual signal frequency lies between two frequency bins and the nulls of the windowing function no longer fall within the neighboring frequency bins.

As shown in Fig. 3.3d, an amplitude error is also obtained in this case. At constant observation time the magnitude of this amplitude error depends on the signal frequency of the input signal (see Fig. 3-4). The error is at its maximum if the signal frequency is exactly between two frequency bins.

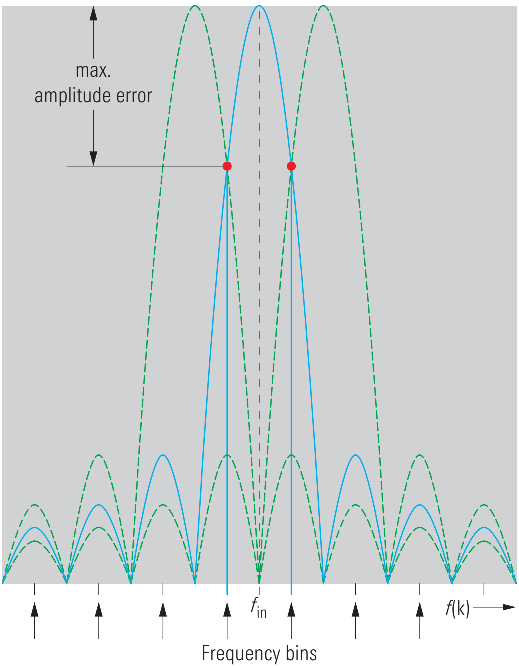


Fig. 3-4
Amplitude error caused by rectangular windowing as a function of signal frequency

By increasing the observation time it is possible to reduce the absolute widening of the spectrum through the higher resolution obtained, but the maximum possible amplitude error remains unchanged. The two effects can, however, be reduced by using optimized windowing instead of the rectangular window. Such windowing functions exhibit lower secondary maxima in the frequency domain so that the leakage effect is reduced as shown in Fig. 3-5. Further details of the windowing functions can be found in [3-1] and [3-2].

To obtain the high level accuracy required for spectrum analysis a flat-top window is usually used. The maximum level error of this windowing function is as small as 0.05 dB. A disadvantage is its relatively wide main lobe which reduces the frequency resolution.

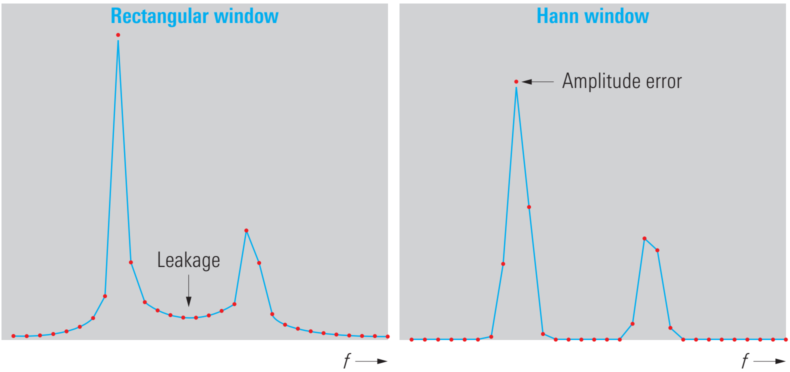


Fig. 3-5 *Leakage effect when using rectangular window or Hann window (MatLab® simulation)*

The number of computing operations required for the Fourier transform can be reduced by using optimized algorithms. The most widely used method is the fast Fourier transform (FFT). Spectrum analyzers operating on this principle are designated as FFT analyzers. The configuration of such an analyzer is shown in Fig. 3-6.

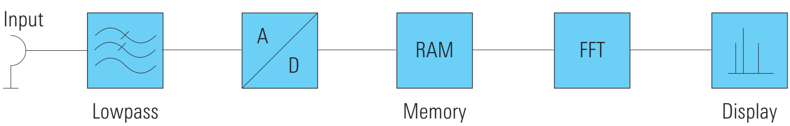


Fig. 3-6 *Configuration of FFT analyzer*

To adhere to the sampling theorem, the bandwidth of the input signal is limited by an analog lowpass filter (cutoff frequency $f_c = f_{in,max}$) ahead of the A/D converter. After sampling the quantized values are saved in a memory and then used for calculating the signal in the frequency domain. Finally, the frequency spectrum is displayed.

Quantization of the samples causes the quantization noise which causes a limitation of the dynamic range towards its lower end. The higher the resolution (number of bits) of the A/D converter used, the lower the quantization noise.

Due to the limited bandwidth of the available high-resolution A/D converters, a compromise between dynamic range and maximum input frequency has to be found for FFT analyzers. At present, a wide dynamic range of about 100 dB can be achieved with FFT analyzers only for low-frequency applications up to 100 kHz. Higher bandwidths inevitably lead to a smaller dynamic range.

In contrast to other analyzer concepts, the phase information is not lost during the complex Fourier transform. FFT analyzers are therefore able to determine the complex spectrum by magnitude and phase. If they feature sufficiently high computing speed, they even allow realtime analysis.

FFT analyzers are not suitable for the analysis of pulsed signals (see Fig. 3-7). The result of the FFT depends on the selected section of the time function. For correct analysis it is therefore necessary to know certain parameters of the analyzed signal, such as the triggering a specific measurement.

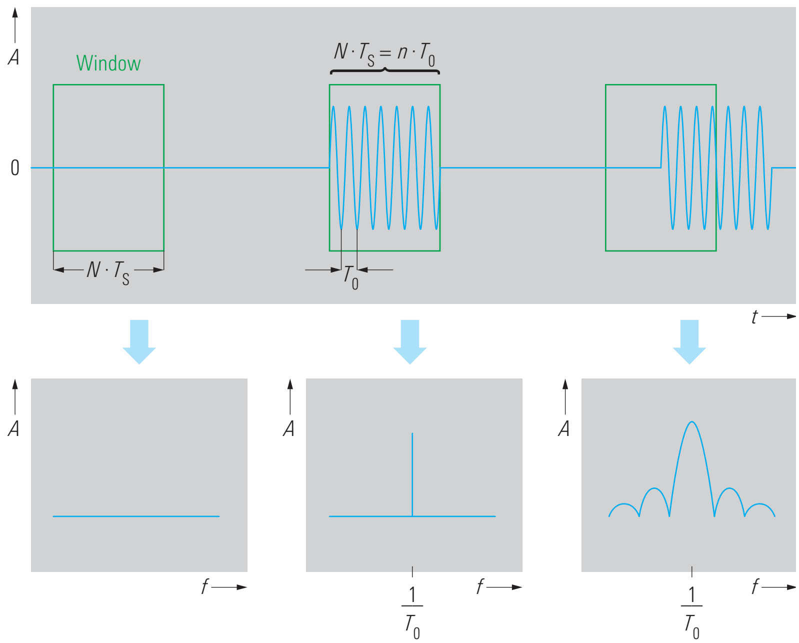


Fig. 3-7 FFT of pulsed signals. The result depends on the time of the measurement

3.2 Analyzers operating in accordance with the heterodyne principle

Due to the limited bandwidth of the available A/D converters, FFT analyzers are only suitable for measurements on low-frequency signals. To display the spectra of high-frequency signals up to the microwave or millimeter-wave range, analyzers with frequency conversion are used. In this case the spectrum of the input signal is not calculated from the time characteristic, but determined directly by analysis in the frequency domain. For such an analysis it is necessary to break down the input spectrum into its individual components. A tunable bandpass filter as shown in Fig. 3-8 could be used for this purpose.

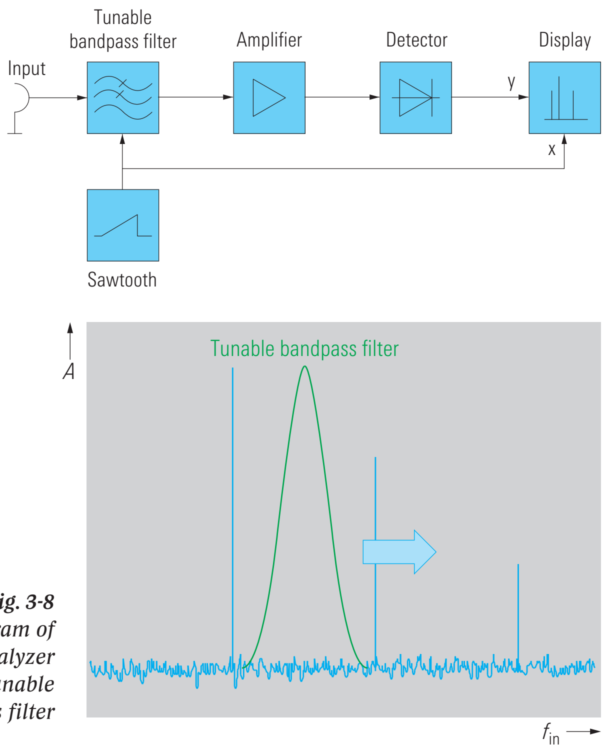


Fig. 3-8 Block diagram of spectrum analyzer with tunable bandpass filter

The filter bandwidth corresponds to the resolution bandwidth (RBW) of the analyzer. The smaller the resolution bandwidth, the higher the spectral resolution of the analyzer.

Narrowband filters tunable throughout the input frequency range of modern spectrum analyzers are however technically hardly feasible. Moreover, tunable filters have a constant relative bandwidth with respect to the center frequency. The absolute bandwidth therefore increases with increasing center frequency so that this concept is not suitable for spectrum analysis.

Spectrum analyzers for high input frequency ranges therefore usually operate in accordance with the principle of a heterodyne receiver. The block diagram of such a analyzer is shown in Fig. 3-9.

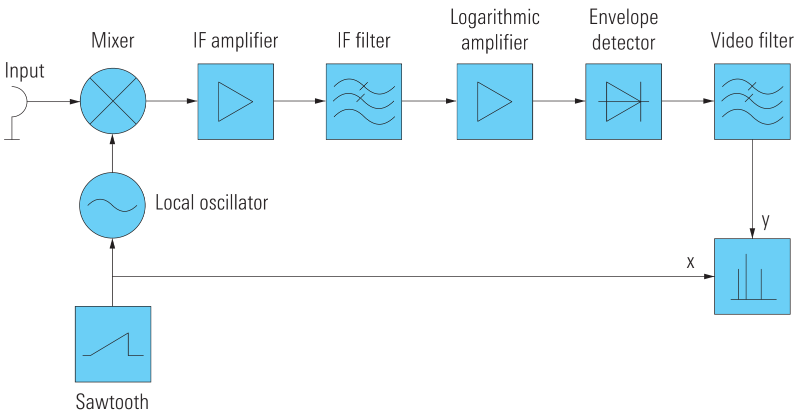


Fig. 3-9 Block diagram of spectrum analyzer operating on heterodyne principle

The heterodyne receiver converts the input signal with the aid of a mixer and a local oscillator (LO) to an intermediate frequency (IF). If the local oscillator frequency is tunable (a requirement that is technically feasible), the complete input frequency range can be converted to a constant intermediate frequency by varying the LO frequency. The resolution of the analyzer is then given by a filter at the IF with fixed center frequency.

In contrast to the concept described above, where the resolution filter as a dynamic component is swept over the spectrum of the input signal, the input signal is now swept past a fixed-tuned filter.

The converted signal is amplified before it is applied to the IF filter which determines the resolution bandwidth. This IF filter has a constant center frequency so that problems associated with tunable filters can be avoided.

To allow signals in a wide level range to be simultaneously displayed on the screen, the IF signal is compressed using of a logarithmic amplifier and the envelope determined. The resulting signal is referred to as the video signal. This signal can be averaged with the aid of an adjustable lowpass filter called a video filter. The signal is thus freed from noise and smoothed for display. The video signal is applied to the vertical deflection of a cathode-ray tube. Since it is to be displayed as a function of frequency, a sawtooth signal is used for the horizontal deflection of the electron beam as well as for tuning the local oscillator. Both the IF and the LO frequency are known. The input signal can thus be clearly assigned to the displayed spectrum.

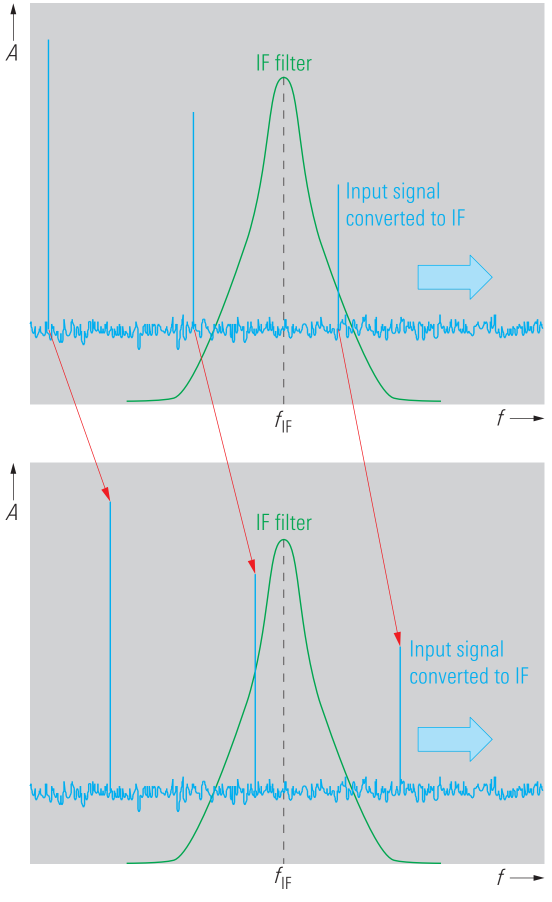


Fig. 3-10
Signal "swept past" resolution filter in heterodyne receiver

In modern spectrum analyzers practically all processes are controlled by one or several microprocessors, giving a large variety of new functions which otherwise would not be feasible. One application in this respect is the remote control of the spectrum analyzer via interfaces such as the IEEE bus.

Modern analyzers use fast digital signal processing where the input signal is sampled at a suitable point with the aid of an A/D converter and further processed by a digital signal processor. With the rapid advances made in digital signal processing, sampling modules are moved further ahead in the signal path. Previously, the video signal was sampled after the analog envelope detector and video filter, whereas with modern spectrum analyzers the signal is often digitized at the last low IF. The envelope of the IF signal is then determined from the samples.

Likewise, the first LO is no longer tuned with the aid of an analog sawtooth signal as with previous heterodyne receivers. Instead, the LO is locked to a reference frequency via a phase-locked loop (PLL) and tuned by varying the division factors. The benefit of the PLL technique is a considerably higher frequency accuracy than achievable with analog tuning.

An LC display can be used instead of the cathode-ray tube, which leads to more compact designs.

3.3 Main setting parameters

Spectrum analyzers usually provide the following elementary setting parameters (see Fig. 3-11):

Frequency display range

The frequency range to be displayed can be set by the start and stop frequency (that is the minimum and maximum frequency to be displayed), or by the center frequency and the span centered about the center frequency. The latter setting mode is shown in Fig. 3-11. Modern spectrum analyzers feature both setting modes.

Level display range

This range is set with the aid of the maximum level to be displayed (the reference level), and the span. In the example shown in Fig. 3-11, a reference level of 0 dBm and a span of 100 dB is set. As will be described later, the attenuation of an input RF attenuator also depends on this setting.

Frequency resolution
For analyzers operating on the heterodyne principle, the frequency resolution is set via the bandwidth of the IF filter. The frequency resolution is therefore referred to as the resolution bandwidth (RBW).

Sweep time (only for analyzers operating on the heterodyne principle)
The time required to record the whole frequency spectrum that is of interest is described as sweep time.

Some of these parameters are dependent on each other. Very small resolution bandwidths, for instance, call for a correspondingly long sweep time. The precise relationships are described in detail in chapter 4.6.

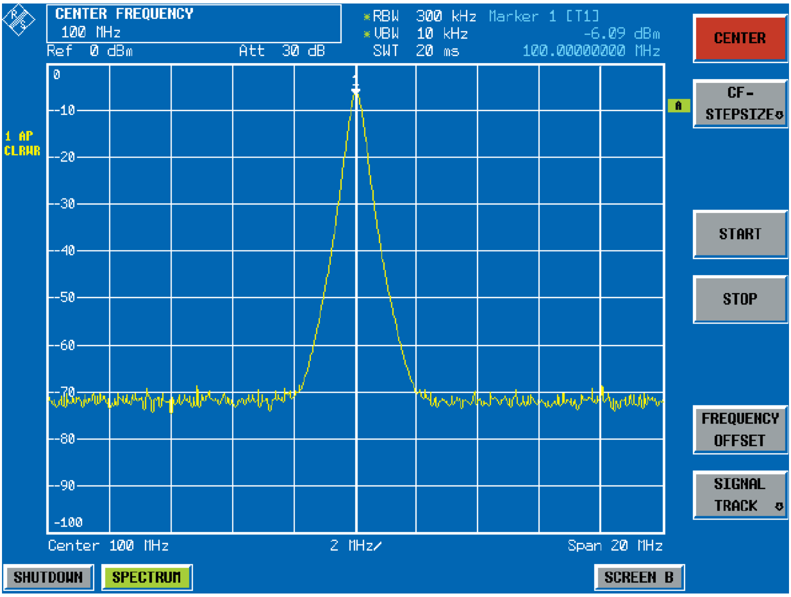


Fig. 3-11 Graphic display of recorded spectrum

4 Practical Realization of an Analyzer Operating on the Heterodyne Principle

This chapter provides a detailed description of the individual components of an analyzer operating on the heterodyne principle as well as of the practical implementation of a modern spectrum analyzer for the frequency range 9 kHz to 3 GHz/7 GHz. A detailed block diagram can be found on the fold-out page at the end of the book. The individual blocks are numbered and combined in functional units.

4.1 RF input section (frontend)

Like most measuring instruments used in modern telecommunications, spectrum analyzers usually feature an RF input impedance of 50 Ω. To enable measurements in 75 Ω systems such as cable television (CATV), some analyzers are alternatively provided with a 75 Ω input impedance. With the aid of impedance transformers, analyzers with 50 Ω input may also be used (see measurement tip: Measurements in 75 Ω system).

A quality criterion of the spectrum analyzer is the input VSWR, which is highly influenced by the frontend components, such as the attenuator, input filter and first mixer. These components form the RF input section whose functionality and realization will be examined in detail in the following.

A step attenuator (2)* is provided at the input of the spectrum analyzer for the measurement of high-level signals. Using this attenuator, the signal level at the input of the first mixer can be set.

The RF attenuation of this attenuator is normally adjustable in 10 dB steps. For measurement applications calling for a wide dynamic range, attenuators with finer step adjustment of 5 dB or 1 dB are used in some analyzers (see chapter 5.5: Dynamic range).

* The colored code numbers in parentheses refer to the block diagram at the end of the book.

Measurements in 75 Ω system

In sound and TV broadcasting, an impedance of 75 Ω is more common than the widely used 50 Ω. To carry out measurements in such systems with the aid of spectrum analyzers that usually feature an input impedance of 50 Ω, appropriate matching pads are required. Otherwise, measurement errors would occur due to mismatch between the device under test and spectrum analyzer.

The simplest way of transforming 50 Ω to 75 Ω is by means of a 25 Ω series resistor. While the latter renders for low insertion loss (approx. 1.8 dB), only the 75 Ω input is matched, however, the output that is connected to the RF input of the spectrum analyzer is mismatched (see Fig. 4-1a). Since the input impedance of the spectrum analyzer deviates from the ideal 50 Ω value, measurement errors due to multiple reflection may occur especially with mismatched DUTs.

Therefore it is recommendable to use matching pads that are matched at both ends (e.g. Π or L pads). The insertion loss through the attenuator may be higher in this case.

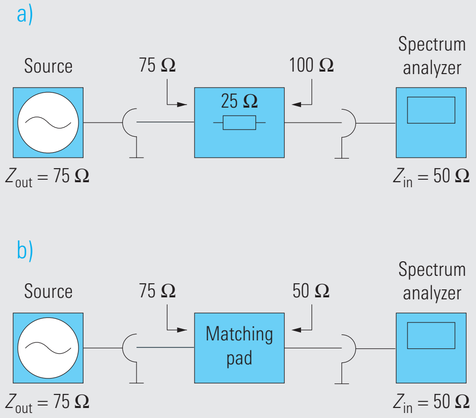


Fig. 4-1 *Input matching to 75 Ω using external matching pads*

The heterodyne receiver converts the input signal with the aid of a mixer (4) and a local oscillator (5) to an intermediate frequency (IF). This type of frequency conversion can generally be expressed as:

$$\left| m \cdot f_{LO} \pm n \cdot f_{in} \right| = f_{IF} \qquad \textit{(Equation 4-1)}$$

where
m, n — $1, 2, \ldots$
f_{LO} — frequency of local oscillator
f_{in} — frequency of input signal to be converted
f_{IF} — intermediate frequency

If the fundamentals of the input and LO signal are considered ($m, n = 1$), Equation 4-1 is simplified to:

$$\left| f_{LO} \pm f_{in} \right| = f_{IF} \qquad \textit{(Equation 4-2)}$$

or solved for f_{in}

$$f_{in} = \left| f_{LO} \pm f_{IF} \right| \qquad \textit{(Equation 4-3)}$$

With a continuously tunable local oscillator, a further input frequency range can be implemented at constant frequency. For specific LO and intermediate frequencies, Equation 4-3 shows that there are always two receive frequencies for which the criterion set by Equation 4-2 is fulfilled (see Fig. 4-2). This means that in addition to the wanted receive frequency there are also image frequencies. To ensure unambiguity of this concept, input signals at such unwanted image frequencies have to be rejected with the aid of suitable filters ahead of the RF input of the mixer.

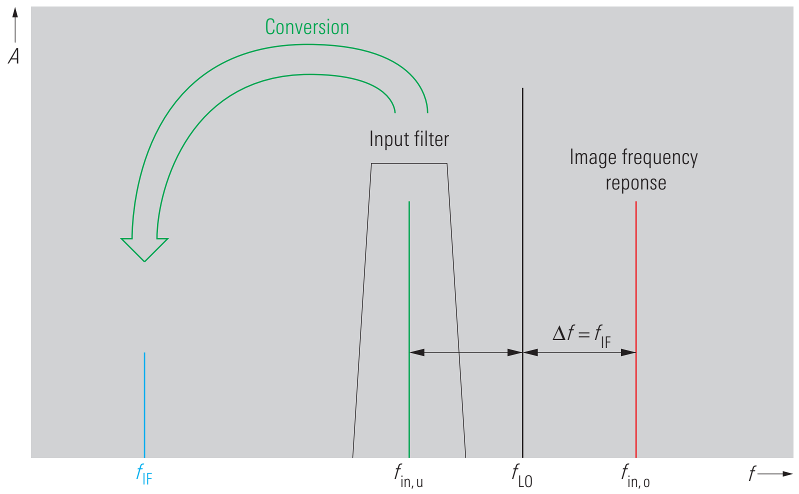


Fig. 4-2 Ambiguity of heterodyne principle

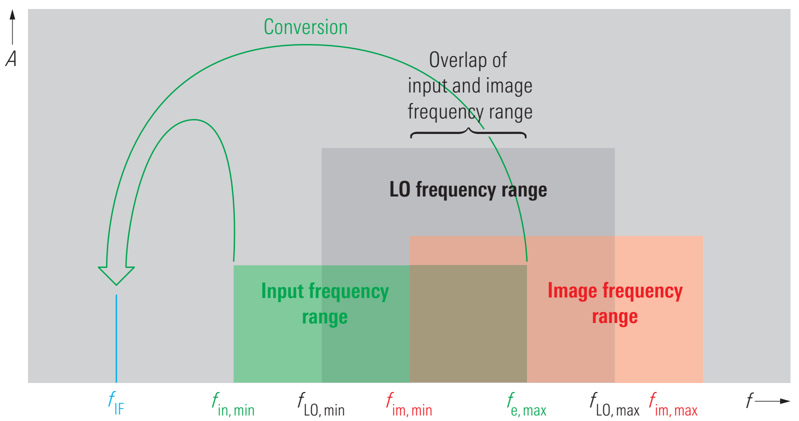


Fig. 4-3 Input and image frequency ranges (overlapping)

Fig. 4-3 illustrates the input and image frequency ranges for a tunable receiver with low first IF. If the input frequency range is greater than $2 \cdot f_{IF}$, the two ranges are overlapping, so an input filter must be implemented as a tunable bandpass for image frequency rejection without affecting the wanted input signal.

To cover the frequency range from 9 kHz to 3 GHz, which is typical of modern spectrum analyzers, this filter concept would be extremely complex because of the wide tuning range (several decades). Much less complex is the principle of a high first IF (see Fig. 4-4).

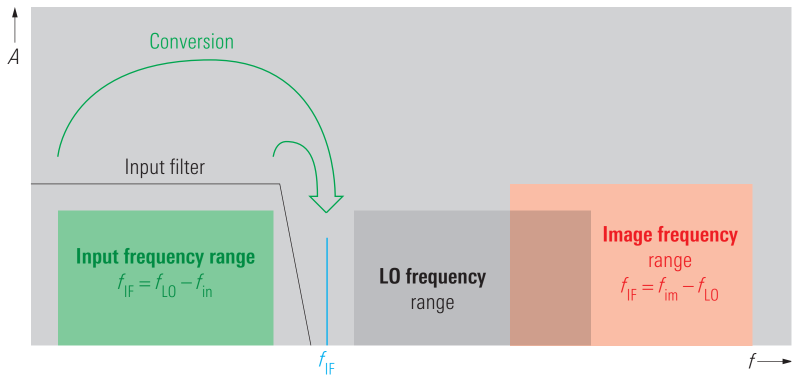


Fig. 4-4 *Principle of high intermediate frequency*

In this configuration, image frequency range lies above the input frequency range. Since the two frequency ranges do not overlap, the image frequency can be rejected by a fixed-tuned lowpass filter. The following relationships hold for the conversion of the input signal:

$$f_{IF} = f_{LO} - f_{in} \quad \textit{(Equation 4-4)}$$

and for the image frequency response:

$$f_{IF} = f_{im} - f_{LO} \quad \textit{(Equation 4-5)}$$

Frontend for frequencies up to 3 GHz

The analyzer described here uses the principle of high intermediate frequency to cover the frequency range from 9 kHz to 3 GHz. The input attenuator (2) is therefore followed by a lowpass filter (3) for rejection of the image frequencies. Due to the limited isolation between RF and IF port as well as between LO and RF port of the first mixer, this lowpass filter also serves for minimizing the IF feedthrough and LO reradiation at the RF input.

In our example the first IF is 3476.4 MHz. For converting the input frequency range from 9 kHz to 3 GHz to an upper frequency of 3476.4 MHz, the LO signal (5) must be tunable in the frequency range from 3476.40 MHz to 6476.4 MHz. According to Equation 4-5, an image frequency range from 6952.809 MHz to 9952.8 MHz is then obtained.

Measurement on signals with DC component

Many spectrum analyzers, in particular those featuring a very low input frequency at their lower end (such as 20 Hz), are DC-coupled, so there are no coupling capacitors in the signal path between RF input and first mixer.

A DC voltage may not be applied to the input of a mixer because it usually damages the mixer diodes. For measurements of signals with DC components, an external coupling capacitor (DC block) is used with DC-coupled spectrum analyzers. It should be noted that the input signal is attenuated by the insertion loss of this DC block. This insertion loss has to be taken into account in absolute level measurements.

Some spectrum analyzers have an integrated coupling capacitor to prevent damage to the first mixer. The lower end of the frequency range is thus raised. AC-coupled analyzers therefore have a higher input frequency at the lower end, such as 9 kHz.

Due to the wide tuning range and low phase noise far from the carrier (see chapter 5.3: Phase noise) a YIG oscillator is often used as local oscillator. This technology uses a magnetic field for tuning the frequency of a resonator.

Some spectrum analyzers use voltage-controlled oscillators (VCO) as local oscillators. Although such oscillators feature a smaller tuning range than the YIG oscillators, they can be tuned much faster than YIG oscillators.

To increase the frequency accuracy of the recorded spectrum, the LO signal is synthesized. That is, the local oscillator is locked to a reference signal (26) via a phase-locked loop (6). In contrast to analog spectrum analyzers, the LO frequency is not tuned continuously, but in many small steps. The step size depends on the resolution bandwidth. Small resolution bandwidths call for small tuning steps. Otherwise, the input signal may not be fully recorded or level errors could occur. To illustrate this effect, a filter tuned in steps throughout the input frequency range is shown in Fig. 4-5. To avoid such errors, a step size that is much lower than the resolution bandwidth (such as $0.1 \cdot B_N$) is selected in practice.

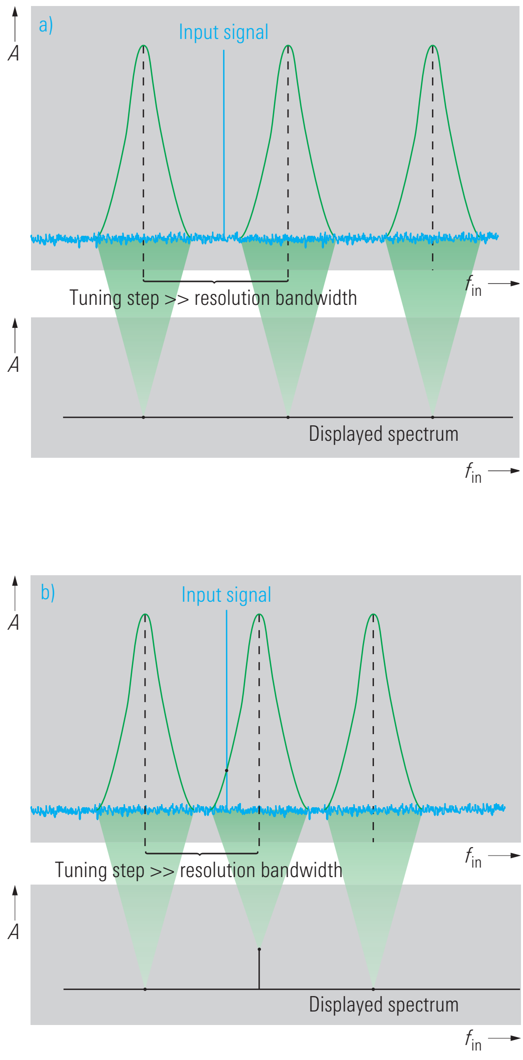


Fig. 4-5
Effects of too large tuning steps
a) *input signal is completely lost*
b) *level error in display of input signal*

The reference signal is usually generated by a temperature-controlled crystal oscillator (TCXO). To increase the frequency accuracy and long-term stability (see also chapter 5.9: Frequency accuracy), an oven-controlled crystal oscillator (OCXO) is optionally available for most spectrum analyzers. For synchronization with other measuring instruments, the reference signal (usually 10 MHz) is made available at an output connector (28). The spectrum analyzer may also be synchronized to an externally applied reference signal (27). If only one connector is available for coupling a reference signal in or out, the function of such connector usually depends on a setting internal to the spectrum analyzer.

As shown in Fig. 3-9, the first conversion is followed by IF signal processing and detection of the IF signal. With such a high IF, narrowband IF filters can hardly be implemented, which means that the IF signal in the concept described here has to be converted to a lower IF (such as 20.4 MHz in our example).

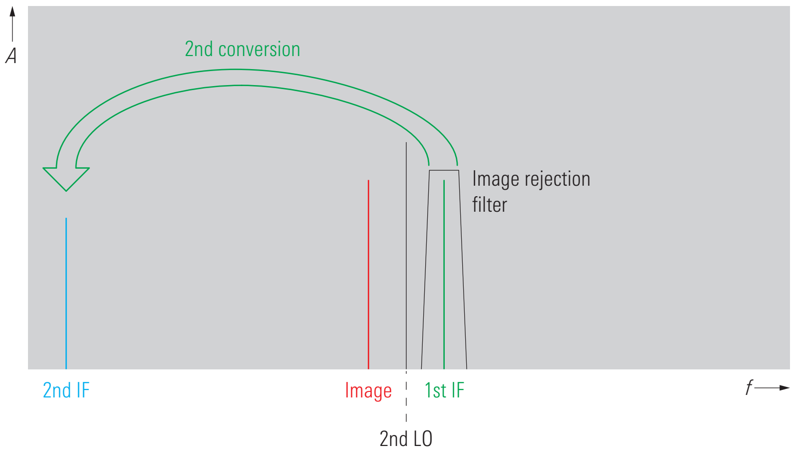


Fig. 4-6 Conversion of high 1st IF to low 2nd IF

With direct conversion to 20.4 MHz, the image frequency would only be offset 2 · 20.4 MHz = 40.8 MHz from the signal to be converted at 3476.4 MHz (Fig. 4-6). Rejection of this image frequency is important since the limited isolation between the RF and IF port of the mixers signals may be passed to the first IF without conversion. This effect is referred to as IF feedthrough (see chapter 5.6: Immunity to interference). If the frequency of the input signal corresponds to the image frequency of the second conversion, this effect is shown in the image frequency response of the second IF. Under certain conditions, input signals may also be converted to the image frequency of the second conversion. Since the conversion loss of mixers is usually much smaller than the isolation between RF and IF port of the mixers, this kind of image frequency response is far more critical.

Due to the high signal frequency, an extremely complex filter with high skirt selectivity would be required for image rejection at a low IF of 20.4 MHz. It is therefore advisable to convert the input signal from the first IF to a medium IF such as 404.4 MHz as in our example. A fixed LO signal (10) of 3072 MHz is required for this purpose since the image frequency for this conversion is at 2667.6 MHz. Image rejection is then simple to realize with the aid of a suitable bandpass filter (8). The bandwidth of this bandpass filter must be sufficiently large so that the signal will not be impaired even for maximum resolution bandwidths. To reduce the total noise figure of the analyzer, the input signal is amplified (7) prior to the second conversion.

The input signal converted to the second IF is amplified again, filtered by an image rejection bandpass filter for the third conversion and converted to the low IF of 20.4 MHz with the aid of a mixer. The signal thus obtained can be subjected to IF signal processing.

Frontend for frequencies above 3 GHz

The principle of a high first IF calls for a high LO frequency range ($f_{\mathrm{LO,\,max}} = f_{\mathrm{in,\,max}} + f_{\mathrm{1st\,IF}}$). In addition to a broadband RF input, the first mixer must also feature an extremely broadband LO input and IF output – requirements that are increasingly difficult to satisfy if the upper input frequency limit is raised. Therefore this concept is only suitable for input frequency ranges up to 7 GHz.

To cover the microwave range, other concepts have to be implemented by taking the following criteria into consideration:

- The frequency range from 3 GHz to 40 GHz extends over more than a decade, whereas 9 kHz to 3 GHz corresponds to approx. 5.5 decades.
- In the microwave range, filters tunable in a wide range and with narrow relative bandwidth can be implemented with the aid of YIG technology [4-1]. Tuning ranges from 3 GHz to 50 GHz are fully realizable.

Direct conversion of the input signal to a low IF calls for a tracking bandpass filter for image rejection. In contrast to the frequency range up to 3 GHz, such preselection can be implemented for the range above 3 GHz due to the previously mentioned criteria. Accordingly, the local oscillator need only be tunable in a frequency range that corresponds to the input frequency range.

In our example the frequency range of the spectrum analyzer is thus enhanced from 3 GHz to 7 GHz. After the attenuator, the input signal is split by a diplexer (19) into the frequency ranges 9 kHz to 3 GHz and 3 GHz to 7 GHz and applied to corresponding RF frontends.

In the high-frequency input section, the signal passes a tracking YIG filter (20) to the mixer. The center frequency of the bandpass filter corresponds to the input signal frequency to be converted to the IF. Direct conversion to a low IF (20.4 MHz, in our example) is difficult with this concept due to the bandwidth of the YIG filter. It is therefore best to convert the signal first to a medium IF (404.4 MHz) as was performed with the low-frequency input section.

In our example, a LO frequency range from 2595.6 MHz to 6595.6 MHz would be required for converting the input signal as upper sideband, (that is for $f_{IF} = f_{in} - f_{LO}$). For the conversion as lower sideband ($f_{IF} = f_{LO} - f_{in}$), the local oscillator would have to be tunable from 3404.4 MHz to 7404.4 MHz.

If one combines the two conversions by switching between the upper and lower sideband at the center of the input frequency band, this concept can be implemented even with a limited LO frequency range of 3404.4 MHz to 6595.6 MHz (see Fig. 4-7).

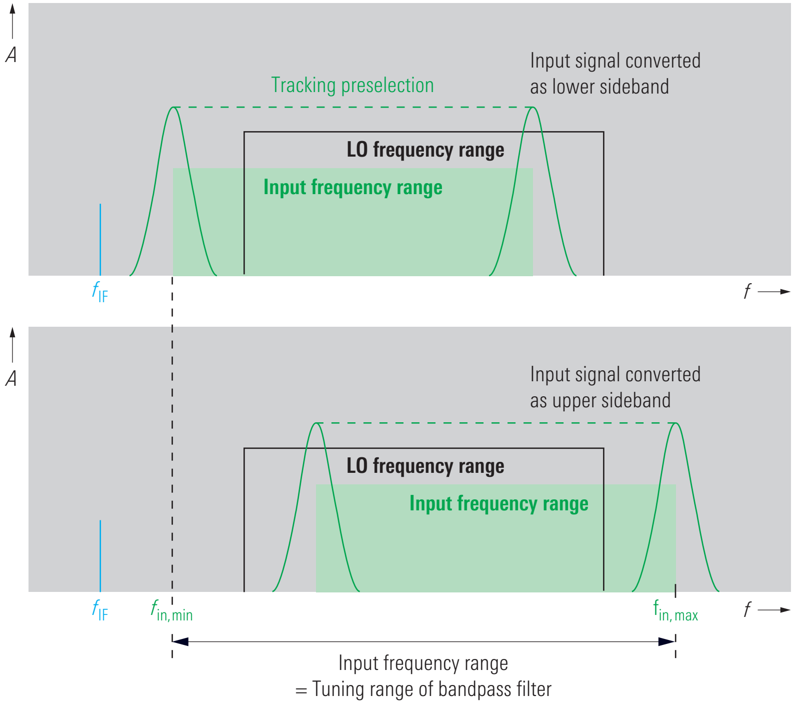


Fig. 4-7 Conversion to a low IF; image rejection by tracking preselection

The signal converted to an IF of 404.4 MHz is amplified (23) and coupled into the IF signal path of the low-frequency input section through a switch (13).

Upper and lower frequency limits of this implementation are determined by the technological constraints of the YIG filter. A maximum frequency of about 50 GHz is feasible.

In our example, the upper limit of 7 GHz is determined by the tuning range of the local oscillator. There are again various possibilities for converting input signals above 7 GHz with the specified LO frequency range:

Fundamental mixing
The input signal is converted by means of the fundamental of the LO signal. For covering a higher frequency range with the specified LO frequency range it is necessary to double, for instance, the LO signal frequency by means of a multiplier before the mixer.

Harmonic mixing
The input signal is converted by a means of a harmonic of the LO signal produced in the mixer due to the mixer's nonlinearities.

Fundamental mixing is preferred to obtain minimal conversion loss, thereby maintaining a low noise figure for the spectrum analyzer. The superior characteristics attained in this way, however, require complex processing of the LO signal. In addition to multipliers (22), filters are required for rejecting subharmonics after multiplying. The amplifiers required for a sufficiently high LO level must be highly broadband since they must be designed for a frequency range that roughly corresponds to the input frequency range of the high-frequency input section.

Conversion by means of harmonic mixing is easier to implement but implies a higher conversion loss. A LO signal in a comparatively low frequency range is required which has to be applied at a high level to the mixer. Due to the nonlinearities of the mixer and the high LO level, harmonics of higher order with sufficient level are used for the conversion. Depending on the order m of the LO harmonic, the conversion loss of the mixer compared to that in fundamental mixing mode is increased by:

$$\Delta a_{\mathrm{M}} = 20\ \mathrm{dB} \cdot \lg m \qquad \textit{(Equation 4-6)}$$

where Δa_{M} increase of conversion loss compared to that in fundamental mixing mode

m order of LO harmonic used for conversion

The two concepts are employed in practice depending on the price class of the analyzer. A combination of the two methods is possible. For example, a conversion using the harmonic of the LO signal doubled by a multiplier would strike a compromise between complexity and sensitivity at an acceptable expense.

External mixers

For measurements in the millimeter-wave range (above 40 GHz), the frequency range of the spectrum analyzer can be enhanced by using external harmonic mixers. These mixers also operate on the principle of harmonic mixing, so that a LO signal in a frequency range that is low compared to the input signal frequency range is required.

The input signal is converted to a low IF by means of a LO harmonic and an IF input inserted at a suitable point into the IF signal path of the low-frequency input section of the analyzer.

In the millimeter-wave range, waveguides are normally used for conducted signal transmission. Therefore, external mixers available for enhancing the frequency range of spectrum analyzers are usually waveguides. These mixers do not normally have a preselection filter and therefore do not provide for image rejection. Unwanted mixture products have to be identified with the aid of suitable algorithms. Further details about frequency range extension with the aid of external harmonic mixers can be found in [4-2].

4.2 IF signal processing

IF signal processing is performed at the last intermediate frequency (20.4 MHz in our example).

Here the signal is amplified again and the resolution bandwidth defined by the IF filter.

The gain at this last IF can be adjusted in defined steps (0.1 dB steps in our example), so the maximum signal level can be kept constant in the subsequent signal processing regardless of the attenuator setting and mixer level. With high attenuator settings, the IF gain has to be increased so that the dynamic range of the subsequent envelope detector and A/D converter will be fully utilized (see chapter 4.6: Parameter dependencies).

The IF filter is used to define that section of the IF-converted input signal that is to be displayed at a certain point on the frequency axis. Due to the high skirt selectivity and resulting selectivity characteristics, a rectangular filter would be desirable. The transient response, however, of such rectangular filters is unsuitable for spectrum analysis. Since such a filter has a long transient time, the input signal spectrum could be converted to the IF only by varying the LO frequency very slowly to avoid level errors from occurring. Short measurement times can be achieved through the use of Gaussian filters optimized for transients. The transfer function of such a filter is shown in Fig. 4-8.

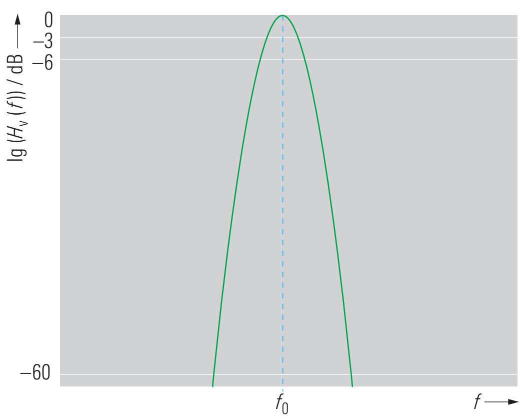


Fig. 4-8
Voltage transfer function of Gaussian filter

In contrast to rectangular filters featuring an abrupt transition from passband to stopband, the bandwidth of Gaussian filters must be specified for filters with limited skirt selectivity. In spectrum analysis it is common practice to specify the 3 dB bandwidth (the frequency spacing between two points of the transfer function at which the insertion loss of the filter has increased by 3 dB relative to the center frequency).

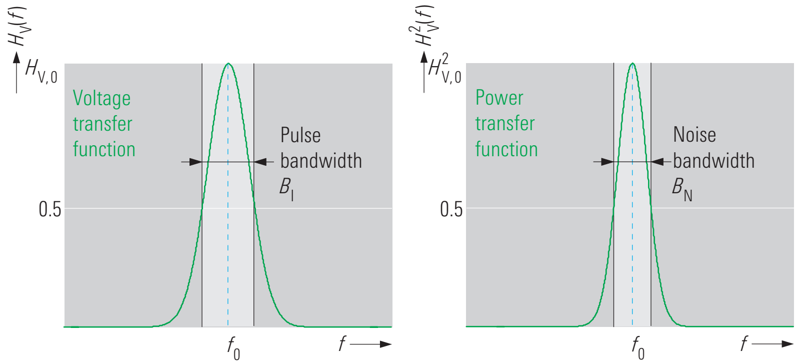


***Fig. 4-9** Voltage and power transfer function of Gaussian filter*

For many measurements on noise or noise-like signals (e.g. digitally modulated signals) the measured levels have to be referenced to the measurement bandwidth, in our example the resolution bandwidth. To this end the equivalent noise bandwidth B_N of the IF filter must be known which can be calculated from the transfer function as follows:

$$B_N = \frac{1}{H_{V,0}^2} \cdot \int_0^{+\infty} H_V^2(f) \cdot \mathrm{d}f \qquad \textbf{(Equation 4-7)}$$

where B_N noise bandwidth
$H_V(f)$ voltage transfer function
$H_{V,0}$ value of voltage transfer function at center of band (at f_0)

This can best be illustrated by looking at the power transfer function (see Fig. 4-9). The noise bandwidth corresponds to the width of a rectangle with the same area as the area of the transfer function $H_V^2(f)$. The effects of the noise bandwidth of the IF filter are dealt with in detail in chapter 5.1 Inherent noise.

For measurements on correlated signals, as can typically be found in the field of radar, the pulse bandwidth is also of interest. In contrast to the noise bandwidth, the pulse bandwidth is calculated by integration of the voltage transfer function. The following applies:

$$B_I = \frac{1}{H_{V,0}} \cdot \int_0^{+\infty} H_V(f) \cdot df \qquad \textit{(Equation 4-8)}$$

where B_I pulse bandwidth
$H_V(f)$ voltage transfer function
$H_{V,0}$ value of voltage transfer function at center of band (at f_0)

The pulse bandwidth of Gaussian or Gaussian-like filters corresponds approximately to the 6 dB bandwidth. In the field of interference measurements, where spectral measurements on pulses are frequently carried out, 6 dB bandwidths are exclusively specified. Further details of measurements on pulsed signals can be found in chapter 6.2.

Chapter 6 concentrates on pulse and phase noise measurements. For these and other measurement applications the exact relationships between 3 dB, 6 dB, noise and pulse bandwidth are of particular interest. Table 4-1 provides conversion factors for various filters that are described in detail further below.

Initial value is 3 dB bandwidth	**4 filter circuits (analog)**	**5 filter circuits (analog)**	**Gaussian filter (digital)**
6 dB bandwidth (B_{6dB})	$1.480 \cdot B_{3dB}$	$1.464 \cdot B_{3dB}$	$1.415 \cdot B_{3dB}$
Noise bandwidth (B_N)	$1.129 \cdot B_{3dB}$	$1.114 \cdot B_{3dB}$	$1.065 \cdot B_{3dB}$
Pulse bandwidth (B_I)	$1.806 \cdot B_{3dB}$	$1.727 \cdot B_{3dB}$	$1.506 \cdot B_{3dB}$
Initial value is 6 dB bandwidth			
3 dB bandwidth (B_{3dB})	$0.676 \cdot B_{6dB}$	$0.683 \cdot B_{6dB}$	$0.707 \cdot B_{6dB}$
Noise bandwidth (B_N)	$0.763 \cdot B_{6dB}$	$0.761 \cdot B_{6dB}$	$0.753 \cdot B_{6dB}$
Pulse bandwidth (B_I)	$1.220 \cdot B_{6dB}$	$1.179 \cdot B_{6dB}$	$1.065 \cdot B_{6dB}$

Table 4-1 *Relationship between 3 dB / 6 dB bandwidths and noise and pulse bandwidths*

If one uses an analyzer operating on the heterodyne principle to record a purely sinusoidal signal, one would expect a single spectral line in accordance with the Fourier theorem even when a small frequency span about the signal frequency is taken. In fact, the display shown in Fig. 4-10 is obtained.

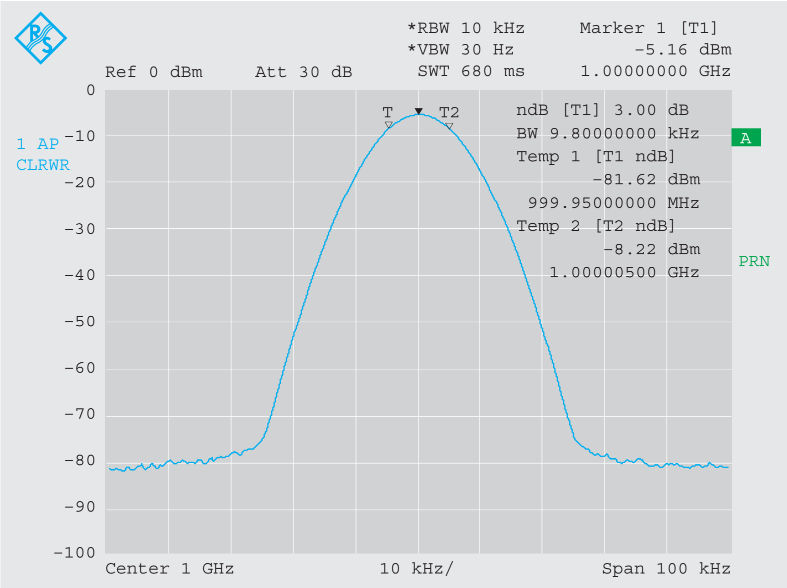


Fig. 4-10 *IF filter imaged by a sinusoidal input signal*

The display shows the image of the IF filter. During the sweep, the input signal converted to the IF is "swept past" the IF filter and multiplied with the transfer function of the filter.

A schematic diagram of this process is shown in Fig. 4-11. For reasons of simplification the filter is "swept past" a fixed-tuned signal, both kinds of representations being equivalent.

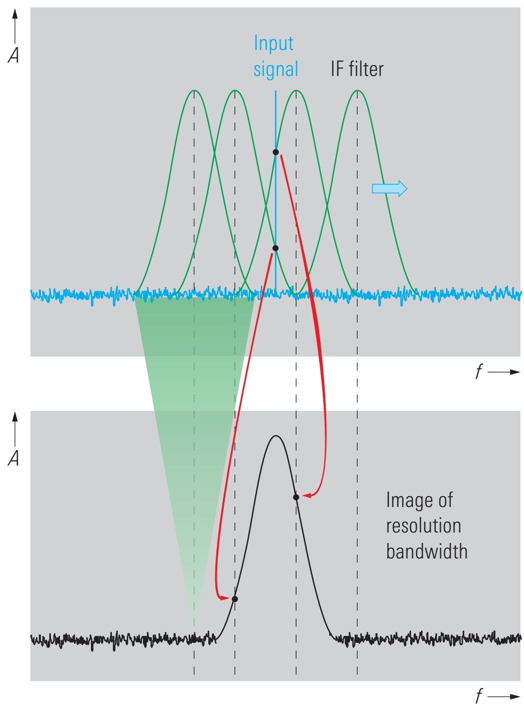


Fig. 4-11
IF filter imaged by an input signal "swept past" the filter (schematic representation of imaging process)

As pointed out before, the spectral resolution of the analyzer is mainly determined by the resolution bandwidth, that is, the bandwidth of the IF filter. The IF bandwidth (3 dB bandwidth) corresponds to the minimum frequency offset required between two signals of equal level to make the signals distinguishable by a dip of about 3 dB in the display when using a sample or peak detector (see chapter 4.4.). This case is shown in Fig. 4-12a. The red trace was recorded with a resolution bandwidth of 30 kHz. By reducing the resolution bandwidth, the two signals are clearly distinguishable (Fig. 4-12a, blue trace).

If two neighboring signals have distinctly different levels, the weaker signal will not be shown in the displayed spectrum at a too high resolution bandwidth setting (see Fig. 4-12b, red trace). By reducing the resolution bandwidth, the weak signal can be displayed.

In such cases, the skirt selectivity of the IF filter is also important and is referred to as the selectivity of a filter. The skirt selectivity is specified in form of the shape factor which is calculated as follows:

$$SF_{60/3} = \frac{B_{60\,\mathrm{dB}}}{B_{3\,\mathrm{dB}}} \qquad \textit{(Equation 4-9)}$$

where $B_{3\,\mathrm{dB}}$ 3 dB bandwidth
$B_{60\,\mathrm{dB}}$ 60 dB bandwidth

For 6 dB bandwidths, as is customary in EMC measurements, the shape factor is derived from the ratio of the 60 dB bandwidth to the 6 dB bandwidth.

The effects of the skirt selectivity can clearly be seen in Fig. 4-13. One Kilohertz IF filters with different shape factors were used for the two traces. In the blue trace ($SF = 4.6$), the weaker signal can still be recognized by the dip, but a separation of the two signals is not possible in the red trace ($SF = 9.5$) where the weaker signal does not appear at all.

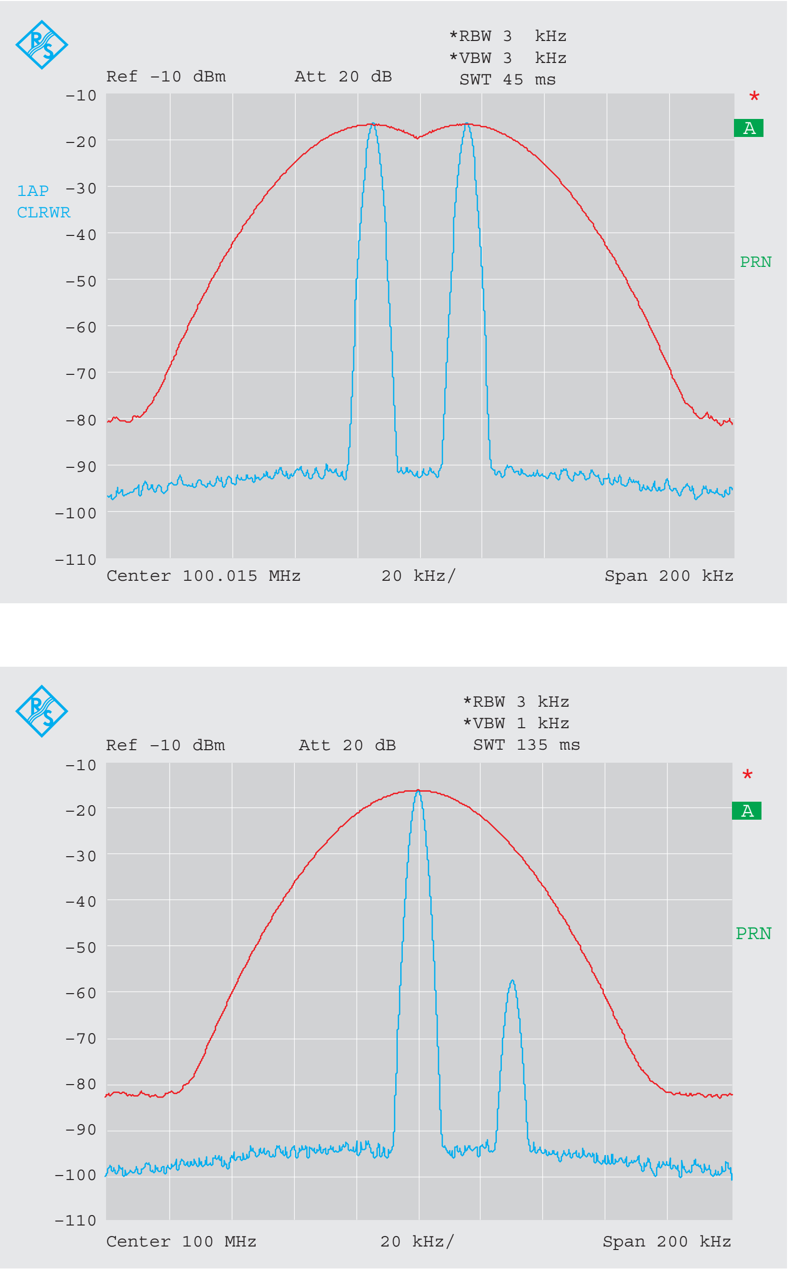


Fig. 4-12 *Spectrum of input signal consisting of two sinusoidal carriers with same and with different level, recorded with different resolution bandwidths (blue traces RBW = 3 kHz, red traces RBW = 30 kHz)*

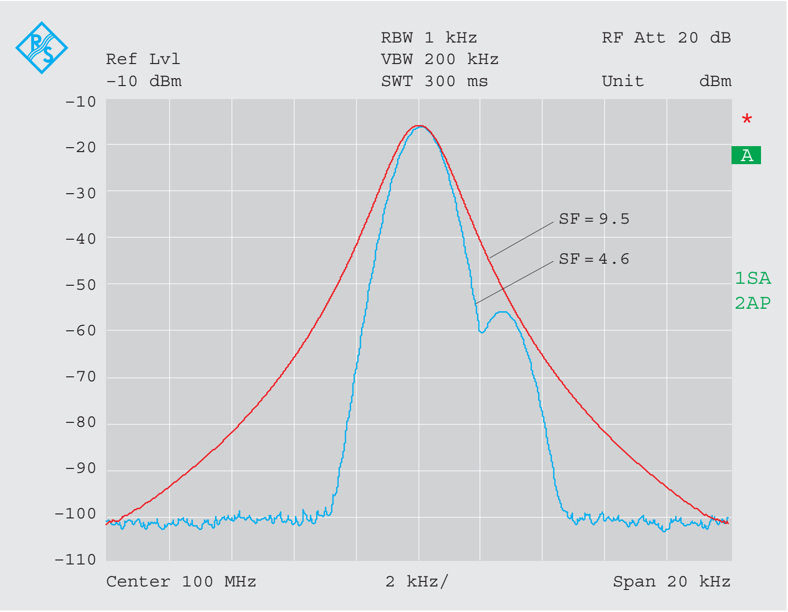


Fig. 4-13 *Two neighboring sinusoidal signals with different levels recorded with a resolution bandwidth of 1 kHz and a shape factor of 9.5 and 4.6*

If the weaker signal is to be distinguished by a filter with a lower skirt selectivity, the resolution bandwidth has to be reduced. Due to the longer transient time of narrowband IF filters, the minimum sweep time must be increased. For certain measurement applications, shorter sweep times are therefore feasible with filters of high skirt selectivity.

As mentioned earlier, the highest resolution is attained with narrowband IF filters. These filters, however, always have a longer transient time than broadband filters, so contemporary spectrum analyzers provide a large number of resolution bandwidths to allow resolution and measurement speed to be adapted to specific applications. The setting range is usually large (from 10 Hz to 10 MHz). The individual filters are implemented in different ways. There are three different types of filters:

- analog filters
- digital filters
- FFT

Analog IF filters

Analog filters are used to realize very large resolution bandwidths. In the spectrum analyzer described in our example, these are bandwidths from 100 kHz to 10 MHz. Ideal Gaussian filters cannot be implemented using analog filters. A very good approximation, however, is possible at least within the 20 dB bandwidth so that the transient response is almost identical to that of a Gaussian filter. The selectivity characteristics depend on the number of filter circuits. Spectrum analyzers typically have four filter circuits, but models with five filter circuits can be found, too. Shape factors of about 14 and 10 can thus be attained, whereas an ideal Gaussian filter exhibits a shape factor of 4.6.

The spectrum analyzer described in our example uses IF filters that are made up of four individual circuits. Filtering is distributed so that two filter circuits each (29 and 31) are arranged before and after the IF amplifier (30). This configuration offers the following benefits:

- The filter circuits ahead of the IF amplifier provide for rejection of mixture products outside the passband of the IF filter. Intermodulation products that may be caused by such signals in the last IF amplifier without prefiltering can thus be avoided (see chapter 5.2: Nonlinearities).
- The filter circuits after the IF amplifier are used to reduce the noise bandwidth. If they were arranged ahead of the IF amplifier, the total noise power in the subsequent envelope detection would be distinctly higher due to the broadband noise of the IF amplifier.

Digital IF filters

Narrow bandwidths can best be implemented with the aid of digital signal processing. In contrast to analog filters, ideal Gaussian filters can be realized. Much better selectivity can be achieved using digital filters instead of analog filters at an acceptable circuit cost. Analog filters consisting of five individual circuits, for instance, have a shape factor of about 10, whereas a digitally implemented ideal Gaussian filter exhibits a shape factor of 4.6. Moreover, digital filters feature temperature stability, are free of aging effects and do not require adjustment. Therefore they feature a higher accuracy regarding bandwidth.

The transient response of digital filters is defined and known. Using suitable correction factors, digital filters allow shorter sweep times than analog filters of the same bandwidth (see chapter 4.6: Parameter dependencies).

In contrast to that shown in the block diagram, the IF signal after the IF amplifier must first be sampled by an A/D converter. To comply with the sampling theorem, the bandwidth of the IF signal must be limited by analog prefilters prior to sampling. This band limiting takes place before the IF amplifier so that intermodulation products can be avoided, as was the case for analog filters. The bandwidth of the prefilter is variable, so depending on the set digital resolution bandwidth, the smallest possible bandwidth can be selected. The digital IF filter provides for limiting the noise bandwidth prior to envelope detection.

The digital IF filter can be implemented by configurations as described in [3-1] or [3-2]. In our example, the resolution bandwidths from 10 Hz to 30 kHz of the spectrum analyzer are realized by digital filters.

FFT

Very narrow IF bandwidths lead to long transient times which considerably reduce the permissible sweep speed. With very high resolution it is therefore advisable to calculate the spectrum from the time characteristic – similar to the FFT analyzer described in chapter 3.1. Since very high frequency signals (up to several GHz) cannot directly be sampled by an A/D converter, the frequency range of interest is converted to the IF as a block, using a fixed-tuned LO signal, and the bandpass signal is sampled in the time domain (see Fig. 4-14). To ensure unambiguity, an analog prefilter is required in this case.

For an IF signal with the center frequency f_{IF} and a bandwidth B, one would expect a minimum sampling rate of $2 \cdot (f_{\mathrm{IF}} + 0.5 \cdot B)$ in accordance with the sampling theorem (Equation 3-1). If the relative bandwidth, however, is small ($B/f_{\mathrm{IF}} \ll 1$), then undersampling is permissible to a certain extent. That is, the sampling frequency may be lower than that resulting from the sampling theorem for baseband signals. To ensure unambiguity, adherance to the sampling theorem for bandpass signals must be maintained. The permissible sampling frequencies are determined by:

$$\frac{2 \cdot f_{\mathrm{IF}} + B}{k+1} \le f_{\mathrm{S}} \le \frac{2 \cdot f_{\mathrm{IF}} - B}{k} \qquad \textit{(Equation 4-10)}$$

where
f_{S} sampling frequency
f_{IF} intermediate frequency
B bandwidth of IF signal
k $1, 2, \ldots$

The spectrum can be determined from the sampled values with the aid of the Fourier transform.

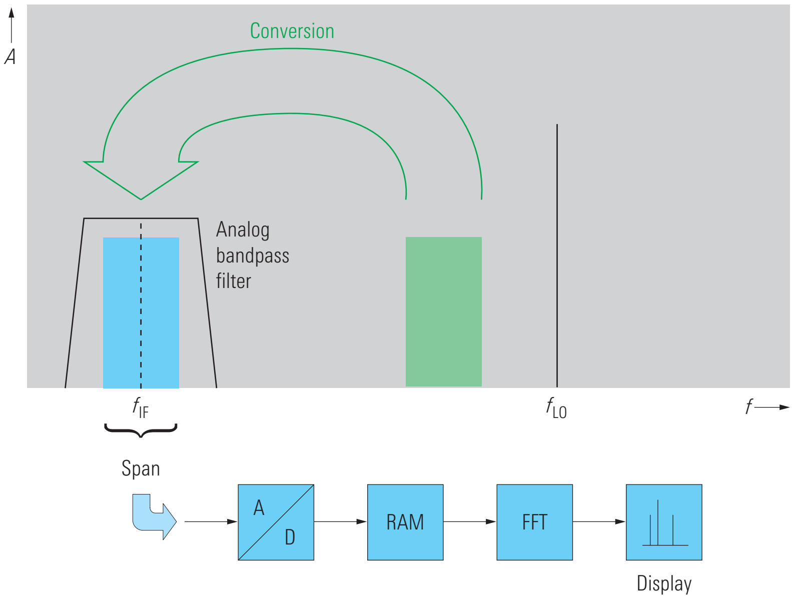


***Fig. 4-14** Spectrum analysis using FFT*

The maximum span that can be analyzed at a specific resolution by means of an FFT is limited by the sampling rate of the A/D converter and by the memory available for saving the sampled values. Large spans must therefore be subdivided into individual segments which are then converted to the IF in blocks and sampled.

While analog or digital filter sweep times increase directly proportional to the span, the observation time required for FFT depends on the desired frequency resolution as described in chapter 3.1. To comply with sampling principles, more samples have to be recorded for the FFT with increasing span so that the computing time for the FFT also increases. At sufficiently high computing speed of digital signal processing, distinctly shorter measurement times than that of conventional filters can be attained with FFT, especially with high span/B_N ratios (see chapter 4.6 Parameter dependencies).

The far-off selectivity of FFT filters is limited by the leakage effect, depending on the windowing function used. The Hann window described in chapter 3.1 is not suitable for spectrum analysis because of the amplitude loss and the resulting level error. A flat-top window is therefore often used to allow the leakage effect to be reduced so that a negligible amplitude error may be maintained. This is at the expense of an observation time that is by a factor of 3.8 longer than that of a rectangular window. The flat-top window causes a wider representation of the windowing function in the frequency domain (corresponding to the convolution with a Dirac function in the frequency domain). When the flat-top window is implemented, a shape factor of about 2.6 can be attained, which means that selectivity is clearly better than when analog or digital IF filters are used.

FFT filters are unsuitable for the analysis of pulsed signals (see chapter 3.1). Therefore it is important for spectrum analyzers to be provided with both FFT and conventional filters.

4.3 Determination of video voltage and video filters

Information about the level of the input signal is contained in the level of the IF signal, such as amplitude-modulated signals in the envelope of the IF signal. With the use of analog and digital IF filters, the envelope of the IF signal is detected after filtering the last intermediate frequency (see Fig. 4-15).

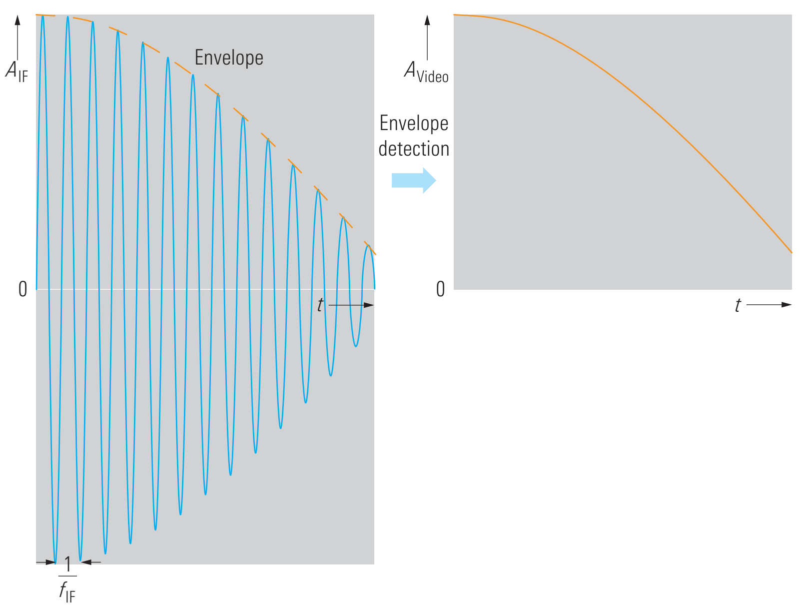


Fig. 4-15 Detection of IF signal envelope

This functional configuration is similar to analog envelope detector circuitry used to demodulate AM signals (see Fig. 4-16). The IF signal is detected and the high-frequency signal component eliminated by a lowpass filter and the video voltage is available at the output of this circuit.

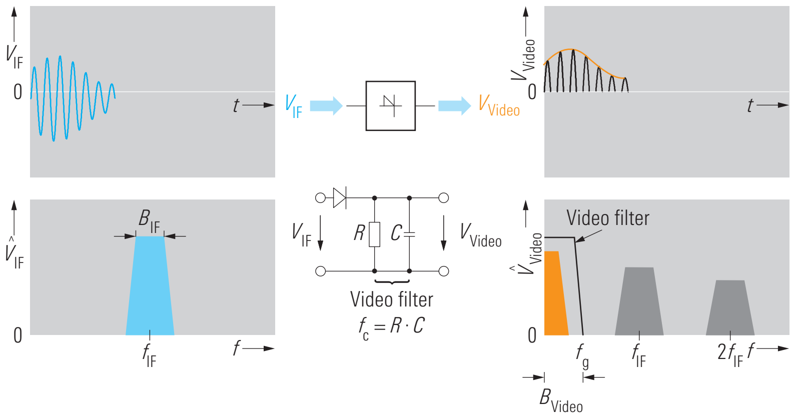


***Fig. 4-16** Detection of IF signal envelope by means of envelope detector*

For digital bandwidths, the IF signal itself is sampled, i.e. the envelope is determined from the samples after the digital IF filter. If one looks at the IF signal represented by a complex rotating vector (cf. chapter 2.1), the envelope corresponds to the length of the vector rotating at an angular velocity of ω_{IF} (see Fig. 4-17). The envelope can be determined by forming the magnitude using the Cordic algorithm [4-3].

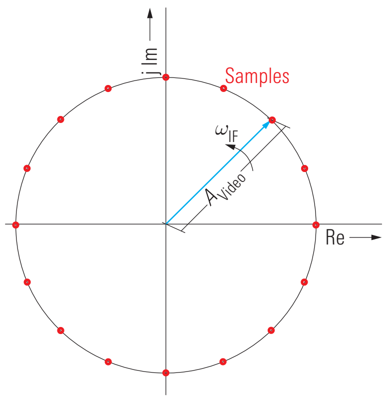


Fig. 4-17
IF signal with sinusoidal input signal, represented by complex rotating vector

Due to envelope detection, the phase information of the input signal gets lost, so that only the magnitude can be indicated in the display. This is one of the primary differences between the envelope detector and the FFT analyzer as described in chapter 3.1.

The dynamic range of the envelope detector determines the dynamic range of a spectrum analyzer. Modern analyzers feature a dynamic range of about 100 dB. It has no sense to simultaneously display so much different values in a linear scale. The level is usually displayed in a logarithmic scale on the spectrum analyzer. The IF signal can therefore be amplified with the aid of a log amplifier (32) ahead of the envelope detector (33), thereby increasing the dynamic range of the display.

The resulting video voltage depends on the input signal and the selected resolution bandwidth. Fig. 4-18 shows some examples. The spectrum analyzer is tuned to a fixed frequency in these examples, so the displayed span is 0 Hz (zero span).

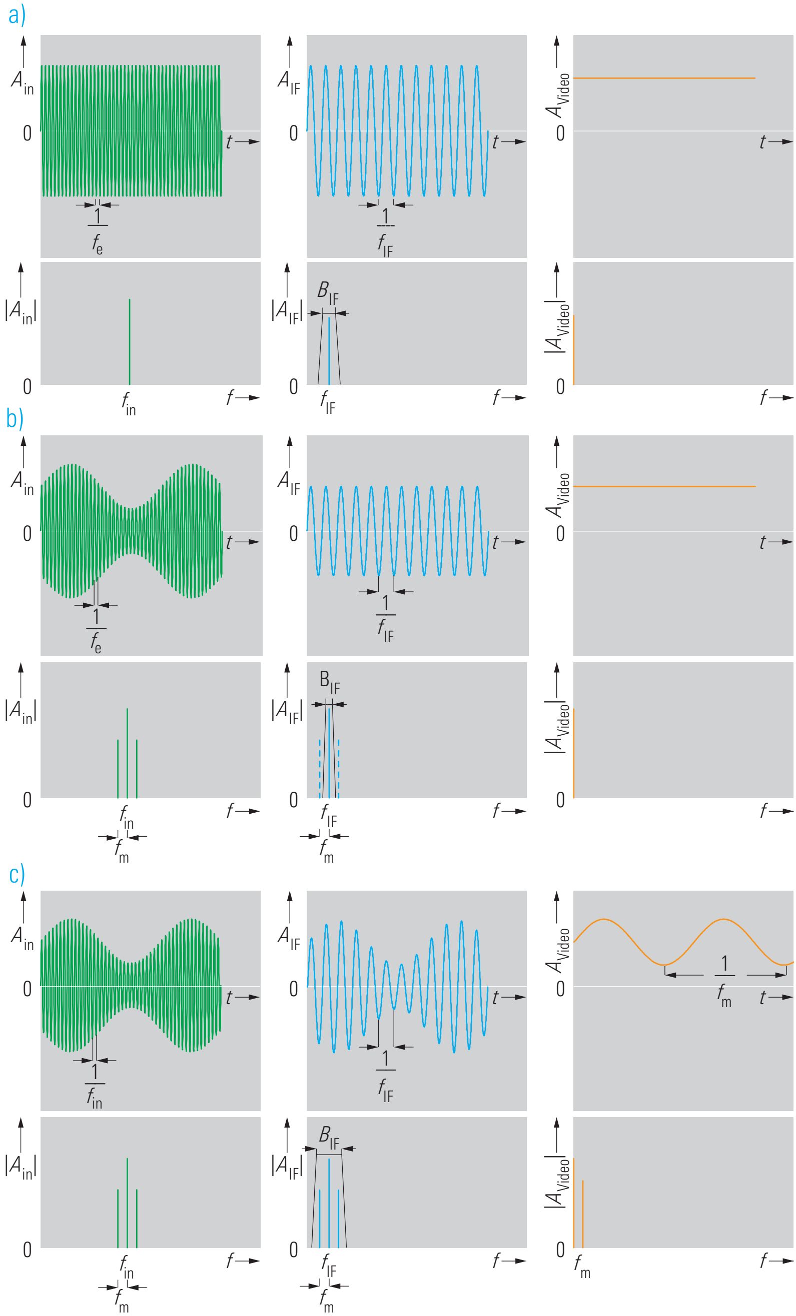


Fig. 4-18 *Video signal (orange traces) and IF signal after IF filter (blue traces) for various input signals (green traces) and resolution bandwidths a) sinusoidal signal b) AM signal, resolution bandwidth smaller than twice the modulation bandwidth c) AM signal, resolution bandwidth greater than twice the modulation bandwidth*

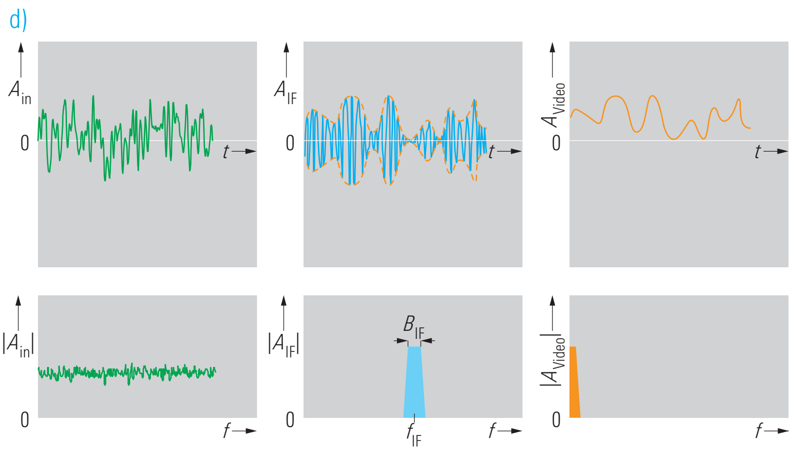


Fig. 4-18 (continued) *Video signal (orange traces) and IF signal after IF filter (blue traces) for various input signals (green traces) and resolution bandwidths d) noise*

The envelope detector is followed by the video filter (35) which defines the video bandwidth (B_V). The video filter is a first order lowpass configuration used to free the video signal from noise, and to smooth the trace that is subsequently displayed so that the display is stabilized. In the analyzer described, the video filter is implemented digitally. Therefore, the video signal is sampled at the output of the envelope detector with the aid of an A/D converter (34) and its amplitude is quantized.

Similar for the resolution bandwidth, the video bandwidth also limits the maximum permissible sweep speed. The minimum sweep time required increases with decreasing video bandwidth (chapter 4.6.1).

The examples in Fig. 4-18 show that the video bandwidth has to be set as a function of the resolution bandwidth and the specific measurement application. The detector used also has be taken into account in the video bandwidth setting (chapter 4.5). The subsequent considerations do not hold true for RMS detectors (chapter 4.4 Detectors).

For measurements on sinusoidal signals with sufficiently high signal-to-noise ratio a video bandwidth that is equal to the resolution bandwidth is usually selected. With a low S/N ratio the display can however be stabilized by reducing the video bandwidth. Signals with weak level are thus shown more distinctly in the spectrum (Fig. 4-19) and the measured level values are stabilized and reproducible. In the case of a sinusoidal signal the displayed level is not influenced by a reduction of the video bandwidth. This becomes quite clear when looking at the video voltage resulting from the sinusoidal input signal in Fig. 4-18a. The video signal is a pure DC voltage, so the video filter has no effect on the overall level of the video signal.

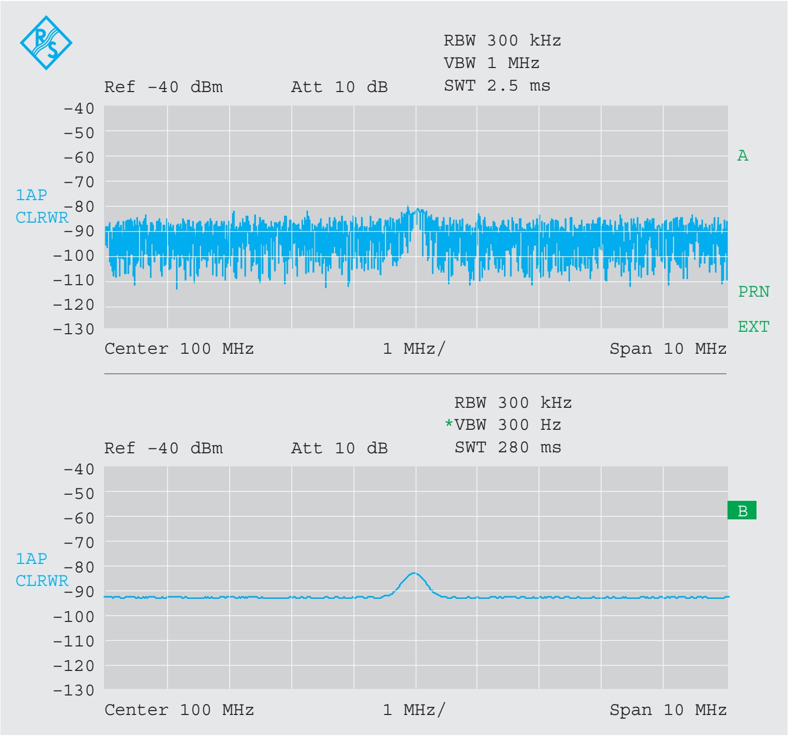


Fig. 4-19 *Sinusoidal signal with low S/N ratio shown for large (top) and small (bottom half of screen) video bandwidth*

To obtain stable and reproducible results of noise measurements, a narrow video bandwidth should be selected. The noise bandwidth is thus reduced and high noise peaks are averaged. As described in greater detail in chapter 4.4, the displayed average noise level will be 2.5 dB below the signal´s RMS value.

Averaging should be avoided when making measurements on pulsed signals. Pulses have a high peak and a low average value (depending on mark-to-space ratio). In order to avoid too low display levels, the video bandwidth should be selected much greater than the resolution bandwidth (Fig. 4-20). This is further discussed in chapter 6.2.

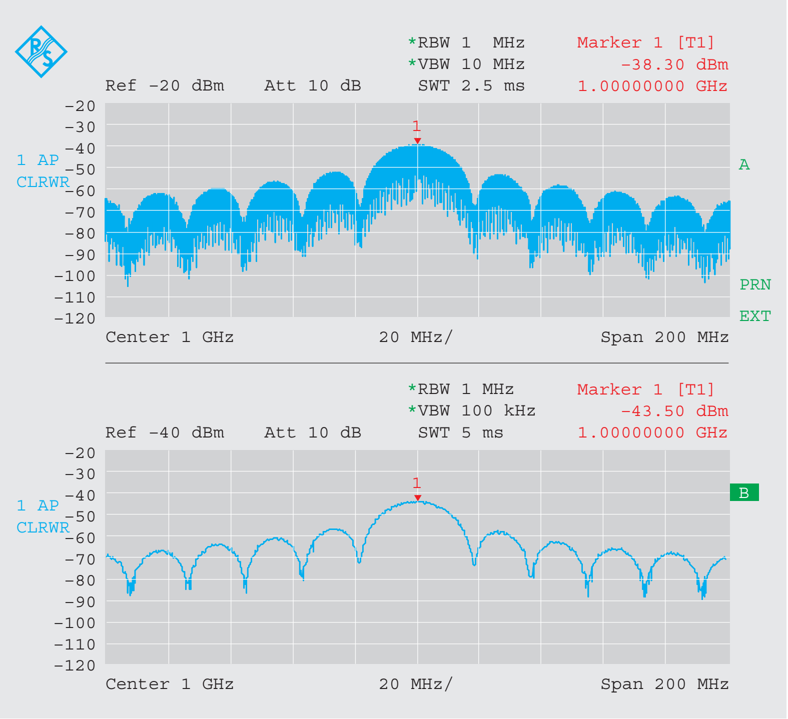


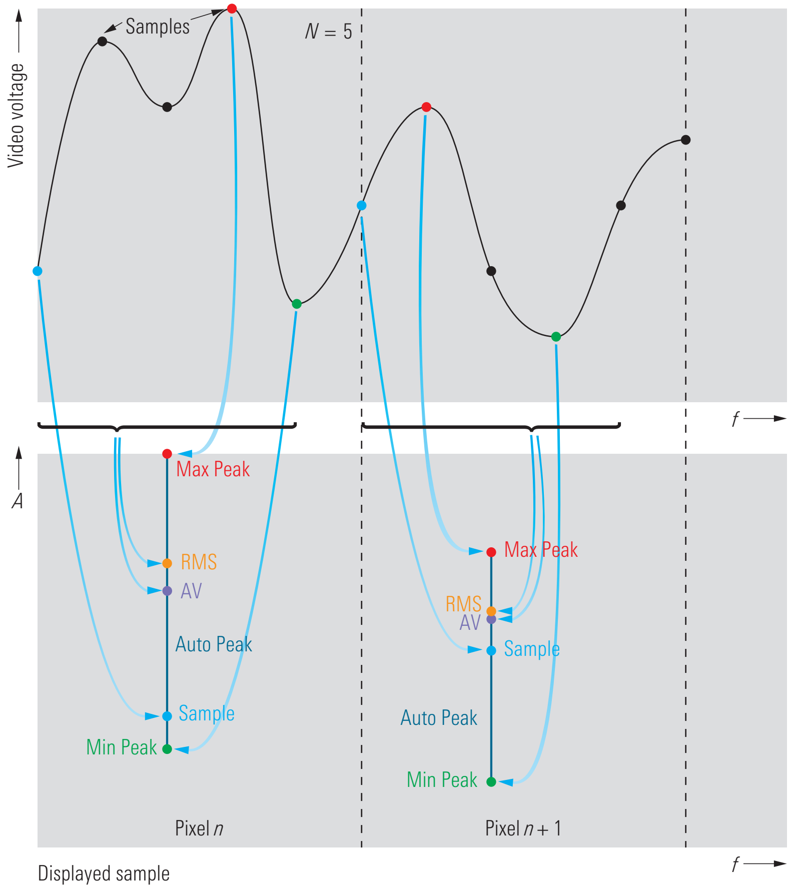
Fig. 4-20 *Pulsed signal recorded with large and small video bandwidth (top and bottom half of screen); note amplitude loss with small video bandwidth (see marker)*

4.4 Detectors

Modern spectrum analyzers use LC displays instead of cathode ray tubes for the display of the recorded spectra. Accordingly, the resolution of both the level and the frequency display is limited.

The limited resolution of the level display range can be remedied by using marker functions (see chapter 4.5: Trace processing). Results can then be determined with considerably high resolution.

Particularily when large spans are displayed, one pixel contains the spectral information of a relatively large subrange. As explained in chapter 4.1, the tuning steps of the 1st local oscillator depend on the resolution bandwidth so that several measured values, referred to as samples or as bins, fall on one pixel. Which of the samples will be represented by the pixel depends on the selected weighting which is determined by the detector. Most of the spectrum analyzers feature min peak, max peak, auto peak and sample detectors. The principles of the detectors is shown in Fig. 4-21.



Figs 4-21 *Selection of sample to be displayed as a function of detector used*

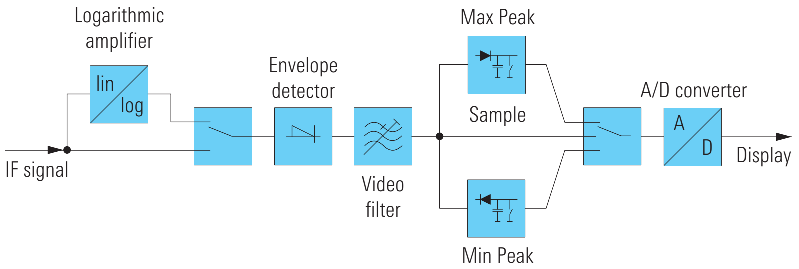


Fig 4-22 Analog realization of detectors

These detectors can be implemented by analog circuits as shown in Fig. 4-22. In this figure, the weighted video signal is sampled at the output of the detector. In the spectrum analyzer described, the detectors (36 to 39) are implemented digitally, so that the video signal is sampled ahead of the detectors (in this case even ahead of the video filter). In addition to the above detectors, average and RMS detectors may also be realized. Quasi-peak detectors for interference measurements are implemented in this way.

Max peak detector

The max peak detector displays the maximum value. From the samples allocated to a pixel the one with the highest level is selected and displayed. Even if wide spans are displayed with very narrow resolution bandwidth (span/RBW >> number of pixels on frequency axis), no input signals are lost. Therefore this type of detector is particularly useful for EMC measurements.

Min peak detector

The min peak detector selects from the samples allocated to a pixel the one with the minimum value for display.

Auto peak detector

The auto peak detector provides for simultaneous display of maximum and minimum value. The two values are measured and their levels displayed, connected by a vertical line (see Fig. 4-21).

Sample detector

The sample detector samples the IF envelope for each pixel of the trace to be displayed only once. That is, it selects only one value from the samples allocated to a pixel as shown in Fig. 4-21 to be displayed. If the span to be displayed is much greater than the resolution bandwidth (span/RBW >> number of pixels on frequency axis), input signals are no longer reliably detected. The same unreliability applies when too large tuning steps of the local oscillator are chosen (see Fig. 4-5). In this case, signals may not be displayed at the correct level or may be completely lost.

RMS detector

The RMS (root mean square) detector calculates the power for each pixel of the displayed trace from the samples allocated to a pixel. The result corresponds to the signal power within the span represented by the pixel. For the RMS calculation, the samples of the envelope are required on a linear level scale. The following applies:

$$V_{\text{RMS}} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} v_i^2} \qquad \textit{(Equation 4-11)}$$

where V_{RMS} RMS value of voltage
N number of samples allocated to the pixel concerned
v_i samples of envelope

The reference resistance R can be used to calculate the power:

$$P = \frac{V_{\text{RMS}}^2}{R} \qquad \textit{(Equation 4-12)}$$

AV detector

The AV (average) detector calculates the linear average for each pixel of the displayed trace from the samples allocated to a pixel. For this calculation the samples of the envelope are required on a linear level scale. The following applies:

$$V_{\text{AV}} = \frac{1}{N} \cdot \sum_{i=1}^{N} v_i \qquad \textit{(Equation 4-13)}$$

where V_{AV} average voltage
N number of samples allocated to the pixel concerned
v_i samples of envelope

Like with the RMS detector, the reference resistance R can be used to calculate the power (Equation 4-12).

Quasi peak detector
This is a peak detector for interference measurement applications with defined charge and discharge times. These times are laid down by CISPR 16-1 [4-4] for instruments measuring spurious emission. A detailed description of this type of detector can be found in chapter 6.2.5.1.

With a constant sampling rate of the A/D converter, the number of samples allocated to a certain pixel increases at longer sweep times. The effect on the displayed trace depends on the type of the input signal and the selected detector. They are described in the following section.

Effects of detectors on the display of different types of input signals
Depending on the type of input signal, the different detectors partly provide different measurement results. Assuming that the spectrum analyzer is tuned to the frequency of the input signal (span = 0 Hz), the envelope of the IF signal and thus the video voltage of a sinusoidal input signal with sufficiently high signal-to-noise ratio are constant. Therefore, the level of the displayed signal is independent of the selected detector since all samples exhibit the same level and since the derived average value (AV detector) and RMS value (RMS detector) correspond to the level of the individual samples.

This is different however with random signals such as noise or noise-like signals in which the instantaneous power varies with time. Maximum and minimum instantaneous value as well as average and RMS value of the IF signal envelope are different in this case.

The power of a random signal is calculated as follows:

$$P = \frac{1}{R} \cdot \lim_{T \to \infty} \left(\frac{1}{T} \cdot \int_{-\frac{T}{2}}^{+\frac{T}{2}} v^2 \left(t \right) \mathrm{d}t \right) \qquad \textit{(Equation 4-14)}$$

or for a certain limited observation time T

$$P = \frac{1}{R} \cdot \frac{1}{T} \cdot \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} v^2 \left(t \right) \mathrm{d}t \qquad \textit{(Equation 4-15)}$$

In the specified observation time T, a peak value can also be found for the instantaneous power. The relationship between the peak value and power can be expressed by the crest factor as follows:

$$CF = 10\ \mathrm{dB} \cdot \lg\left(\frac{P_\mathrm{S}}{P}\right) \qquad \textit{(Equation 4-16)}$$

where CF crest factor

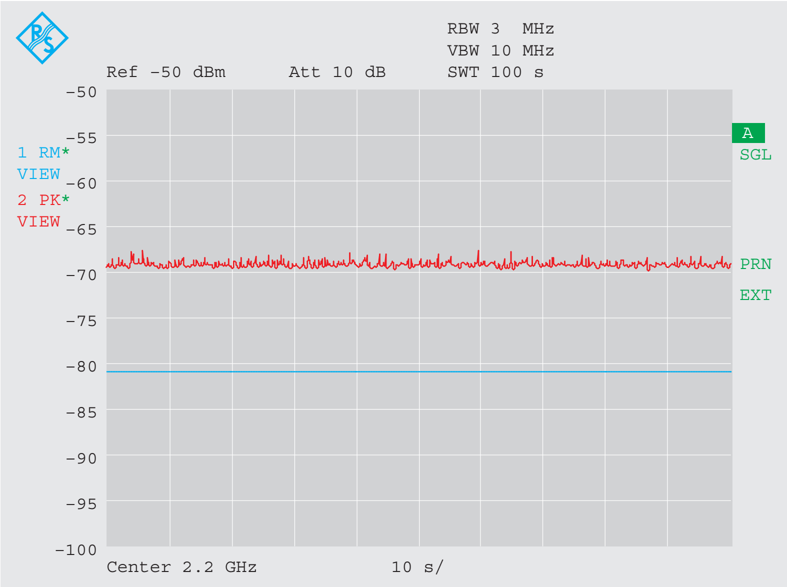
P_S peak value of instantaneous power in observation time T

P power

With noise, any voltage values may occur theoretically, so the crest factor would be arbitrarily high. Since the probability for very high or very low voltage values is low, a crest factor of about 12 dB is usually obtained in practice for Gaussian noise observed over a sufficiently long period.

Digitally modulated signals often exhibit a spectrum similar to noise. However, the crest factor usually differs from that for Gaussian noise. Fig. 4-23 shows the peak and RMS values of Gaussian noise and of a IS-95 CDMA signal (forward channel).

a) Crest factor 12 dB



b) Crest factor 13.8 dB

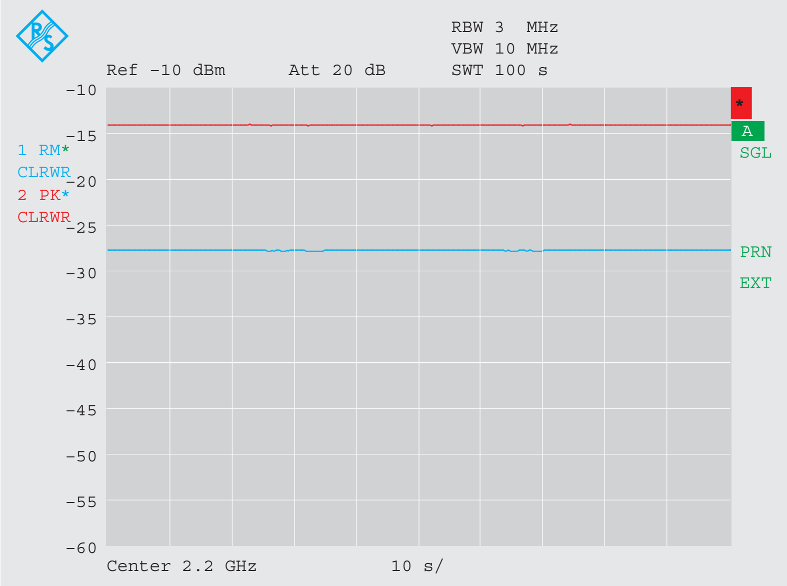


Fig. 4-23 *Peak (red traces) and RMS values (blue traces) of Gaussian noise (a) and of a IS-95 CDMA signal (b), recorded with max peak and RMS detectors*

The effects of the selected detector and of the sweep time on the results of measurements on stochastic signals are described in the following.

Max peak detector

When using the max peak detector, stochastic signals are overweighted so that the maximum level is displayed. With increasing sweep time the dwell time in a frequency range allocated to a pixel is also increased. In the case of Gaussian noise the probability that higher instantaneous values will occur also rises. This means that the levels of the displayed pixels also become higher (see Fig. 4-24a).

With a small ratio between span and resolution bandwidth, the noise displayed for short sweep times is equal to that displayed with a sample detector, since only one sample is recorded per pixel.

Min peak detector

When using the min peak detector, stochastic signals are underweighted so that the minimum level is displayed. The noise displayed on the spectrum analyzer is strongly suppressed. In the case of Gaussian noise the probability that lower instantaneous values will occur increases with increasing sweep time. This means that the levels of the displayed pixels also become lower (see Fig. 4-24a).

If measurements are carried out on sinusoidal signals with low signal-to-noise ratio, the minimum of the noise superimposed on the signal will also be displayed so that the level measurements yield too low values.

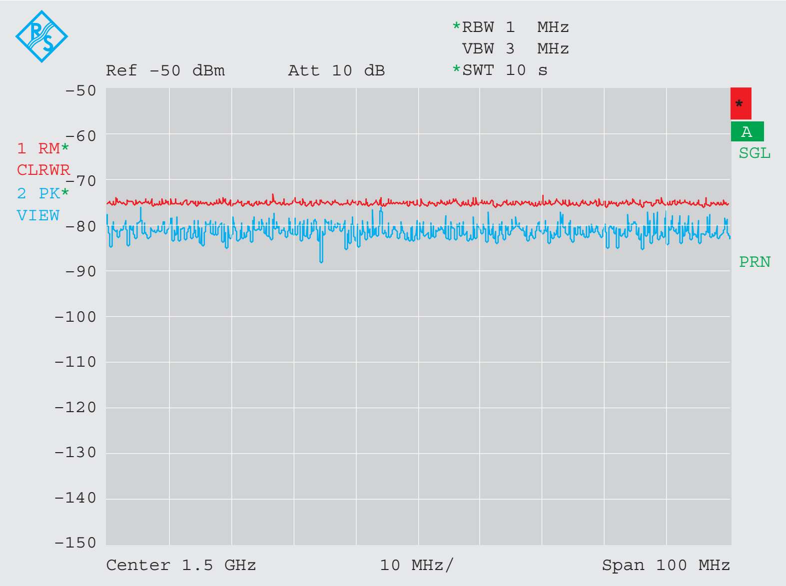
With a small ratio between span and resolution bandwidth, the noise displayed for short sweep times is equal to that displayed with a sample detector, since only one sample is recorded per pixel.

Auto peak detector

When using the auto peak detector, the results of the max peak and min peak detectors are displayed simultaneously, the two values being connected by a line. With increasing sweep time the displayed noise band becomes distinctly wider.

With a small ratio between span and resolution bandwidth, the noise displayed for short sweep times is equal to that displayed with a sample detector, since only one sample is recorded per pixel.

a)



b)

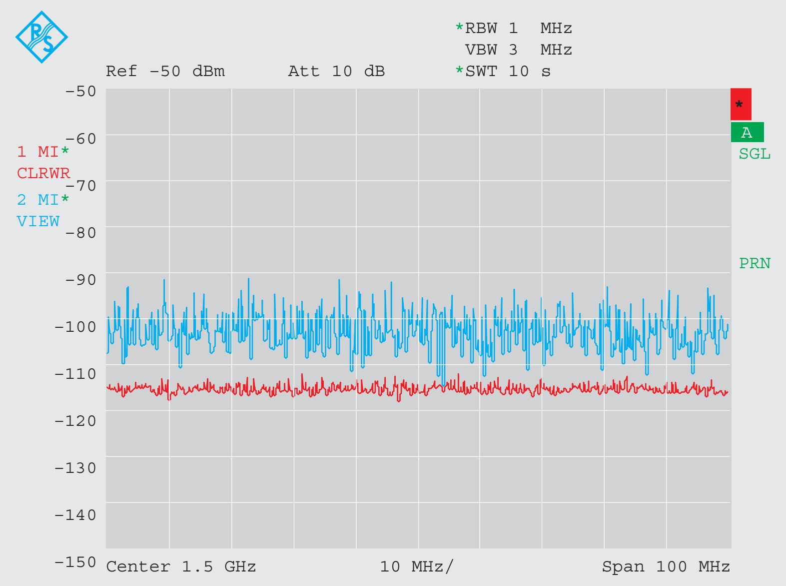


Fig. 4-24 *Displayed noise varying as a function of sweep time, with max peak detector (a) and min peak detector (b), sweep time 2.5 ms (blue trace) and 10 s (red trace)*

Sample detector

As shown in Fig. 4-21, the sample detector always displays a sample recorded at a defined point in time. Due to the distribution of the instantaneous values, the trace displayed in the case of Gaussian noise therefore varies about the average value of the IF signal envelope resulting from noise. This average value is 1.05 dB below the RMS value. If noise is averaged over a narrow video bandwidth (VBW $<$ RBW) using the logarithmic level scale, the displayed average value is an additional 1.45 dB too low. The displayed noise is then 2.5 dB below the RMS value.

In contrast to the other detectors the sweep time has no effect on the displayed trace since the number of the recorded samples is independent of the sweep time.

RMS detector

The RMS detector allows measurement of the actual power of an input signal irrespective of its temporal characteristic. When using a sample or max peak detector, the relationship between RMS value and peak value must be precisely known for determining the power of signals with random instantaneous value. This knowledge is not required when using an RMS detector.

The RMS value displayed by a specific pixel is calculated from all samples pertaining to this pixel. By increasing the sweep time, the number of samples available for the calculation is increased, thus allowing smoothing of the displayed trace. Smoothing by reducing the video bandwidth or by averaging over several traces (see chapter 4.5) is neither permissible nor necessary with the RMS detector. The measurement results would be falsified, since the displayed values would be too low (max. 2.51 dB). To avoid any falsification of results, the video bandwidth should be at least three times the resolution bandwidth when using the RMS detector.

AV detector

The AV detector determines the average value from the samples using the linear level scale. The actual average value is thus obtained irrespective of the type of input signal. Averaging of logarithmic samples (log average) would yield results that were too low since higher signal levels are subject to greater compression by logarithmation. By increasing the sweep time, several samples are available for calculating the average value that is displayed by a specific pixel. The displayed trace can thus be smoothed.

A narrow video bandwidth causes averaging of the video signal. If samples of the linear level scale are applied to the input of the video filter, the linear average of the samples is formed when reducing the video bandwidth. This corresponds to the function of the AV detector so that smoothing by means of narrow video bandwidths is permissible in this case.

The same holds true for the analyzer described here, since samples with linear level scale are applied to the input of the video filter when the AV detector is used (see block diagram).

If the video bandwidth is reduced, the displayed noise converges for max peak, min peak, auto peak and sample detectors since the samples are averaged by the video filter before they are weighted by the detector. If a linear envelope detector is used to determine the IF signal envelope, samples with linear scale are averaged by the video filter. The resulting display corresponds to the true average value and hence to the displayed noise when using an AV detector. If the IF signal is log-amplified before the video voltage is formed, the resulting averaged samples are lower than the true average value. In the case of Gaussian noise the difference is 1.45 dB (see Fig. 4-25a). Since the linear average of the video voltage resulting from Gaussian noise is already 1.05 dB below the RMS value, the samples obtained are all 2.5 dB lower than those obtained with the RMS detector (see Fig. 4-25a). Due to this known relationship an RMS detector is not required to determine the Gaussian noise power. The power can be calculated from the samples collected by the sample detector, taking into account a correction factor of 2.5 dB.

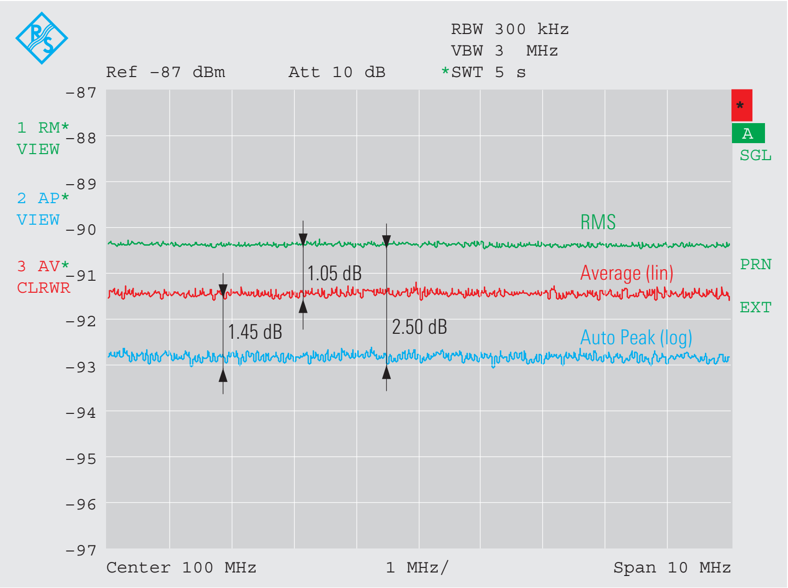
This relationship does not apply to other random signals whose instantaneous values are not in line with the Gaussian distribution (for example, digitally modulated signals, see Fig. 4-25b). If the crest factor is unknown, the power of such signals can only be determined using an RMS detector.

Averaging over several measurements

As described in the following chapter, modern analyzers feature the possibility of averaging traces over several measurements (trace average). This method of averaging partly leads to results different from those when using narrowband video filters.

Depending on whether the recorded trace is displayed on a linear or logarithmic level scale, linear or logarithmic samples are used for averaging. Whether the trace is falsified by averaging depends on the display mode.

a)



b)

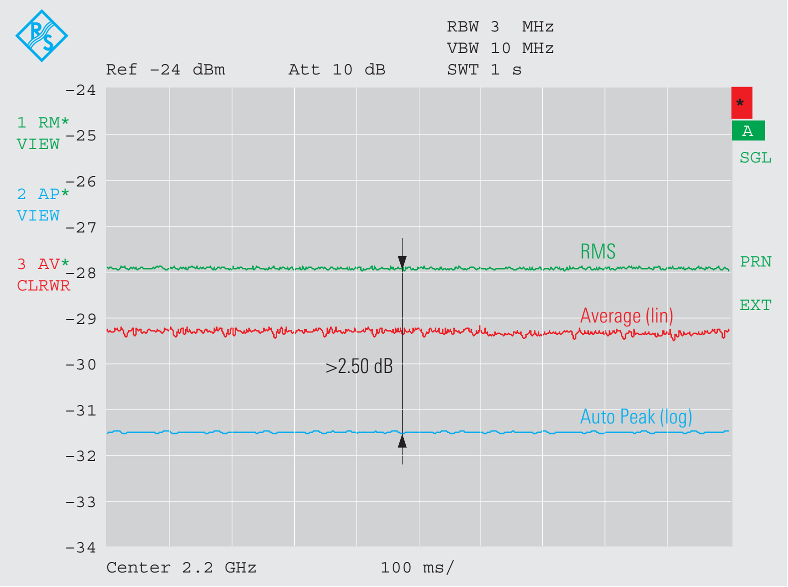


Fig. 4-25 *Measurement of Gaussian noise (a) and IS-95 CDMA signal (b) using RMS and AV detectors (green and red traces) as well as auto peak detector with averaging over narrow video bandwidth (blue trace)*

In the case of averaging over several measurements, the displayed noise levels do not converge for max peak, min peak and sample detectors. The average is derived from the maximum and minimum values, whereas with the use of the video filter, the samples are averaged prior to weighting and therefore converge.

The sample detector yields the average noise level. With logarithmic level display, the displayed average value is 1.45 dB too low, as already explained above. With linear level display and large video bandwidth (VBW $\geq$ 10 · RBW) the true average is obtained, as with the AV detector.

When using the auto peak detector, averaging over several traces is not recommended since the maximum and minimum value is displayed. When the trace average function is activated, automatic switchover to sample detector is often made.

For the RMS detector, trace averaging is permitted neither in the linear nor in the logarithmic level mode.

4.5 Trace processing

As was explained in chapter 4.4, linear samples are required for AV and RMS detectors. For displaying the traces on a logarithmic level scale when these detectors are used, the detectors are followed by a log amplifier (40) which may be optionally activated.

In modern spectrum analyzers, the measurement results are digitized before they are displayed. This allows many different methods of trace evaluation (41).

Measured data memory
Several traces can be stored in modern analyzers and simultaneously displayed. This function is particularly useful for comparative measurements.

Trace average
With the aid of this function a displayed trace can be smoothed by averaging over several measurements (sweeps). The user can enter the number of sweeps to be averaged.

Depending on the input signal and the detector used, this way of averaging may lead to other results than averaging by reducing the video bandwidth.

Marker functions

Marker functions are particularly useful for the evaluation of recorded traces. They allow frequency and level to be displayed at any point of the trace. The limited display accuracy due to the constrained screen resolution can thus be remedied. In addition to functions which set the marker automatically to a signal with maximum level, level differences between signals can also be directly displayed using the delta marker feature.

Modern spectrum analyzers feature enhanced marker functions allowing, for instance, direct noise or phase noise measurements, without manual setting of bandwidth or correction factors (see Fig. 4-26).

The precise frequency of a displayed signal can also be determined with the aid of a marker and a count function (signal count). In many cases the spectrum analyzer can thus replace a frequency counter.

Tolerance masks (limit lines)

Limit values to be adhered to by the device under test can easily be checked with the aid of tolerance masks. To simplify use in production, recorded traces are automatically checked for violation of the specified limit values and the result is output in form of a “pass” or “fail” message (see Fig. 4-27).

Channel power measurement

In the case of digitally modulated signals, power often has to be measured within one channel or within a specific frequency range. Channel power is calculated from the recorded trace, with special functions being provided for this purpose by modern spectrum analyzers. Adjacent-channel power measurement with the aid of a spectrum analyzer is described in detail in chapter 6.3.

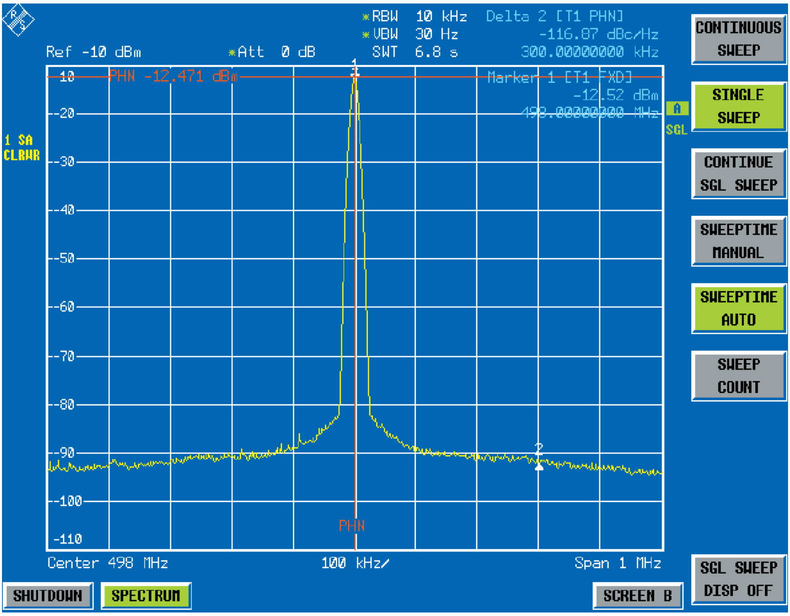


Fig. 4-26 *Marker functions for easy phase noise measurement of an input signal*

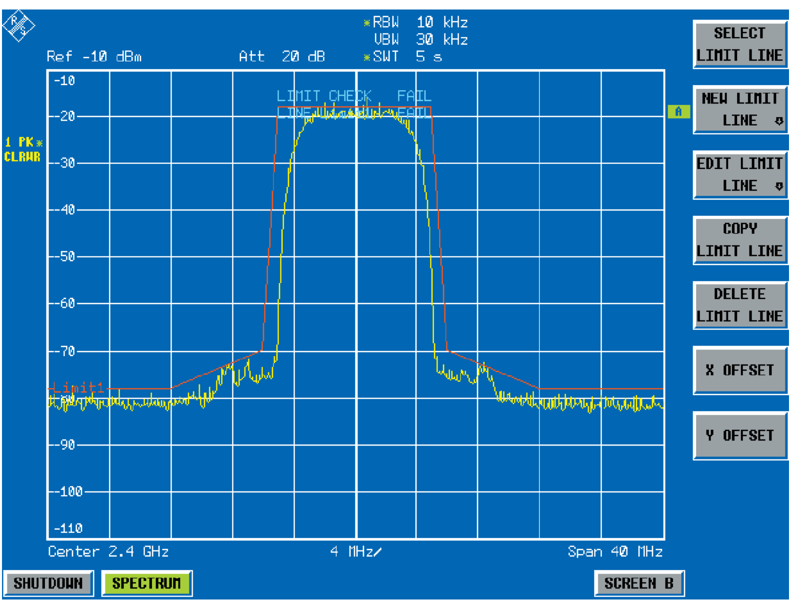


Fig. 4-27 *Evaluation of traces with the aid of limit lines*

4.6 Parameter dependencies

Some of the analyzer settings are interdependent. To avoid measurement errors, these parameters are coupled to one another in normal operating mode of modern spectrum analyzers. That is, upon varying one setting all other dependent parameters will be adapted automatically. The parameters can, however, also be set individually by the user. In such a case it is especially important to know relationships and effects of various settings.

4.6.1 Sweep time, span, resolution and video bandwidths

Through the use of analog or digital IF filters, the maximum permissible sweep speed is limited by the transient time of the IF filter and video filter. The transient time has no effect if the video bandwidth is larger than the resolution bandwidth. In this case, the required transient time increases inversely with the square of the resolution bandwidth, so with a decrease of the resolution bandwidth by the factor n the required minimum sweep time becomes n^2 longer. The following applies:

$$T_{\text{Sweep}} = k \cdot \frac{\Delta f}{B_{\text{IF}}^2} \qquad \textbf{\textit{(Equation 4-17)}}$$

where T_{Sweep} minimum sweep time required (with specified span and resolution bandwidth)
B_{IF} resolution bandwidth
Δf span
k proportionality factor

The proportionality factor k depends on the type of filter and the permissible transient response error. For analog filters made up of four or five individual circuits, the proportionality factor k is 2.5 (maximum transient response error approx. 0.15 dB). With digitally implemented Gaussian filters, the transient response is known and exactly reproducible. Compared to analog filters, higher sweep speeds without amplitude loss can be obtained through appropriate correction factors independent of the type of input signal. A k factor of 1 can thus be attained. Fig. 4-28 shows the required sweep time for a span of 1 MHz as a function of the resolution bandwidth.

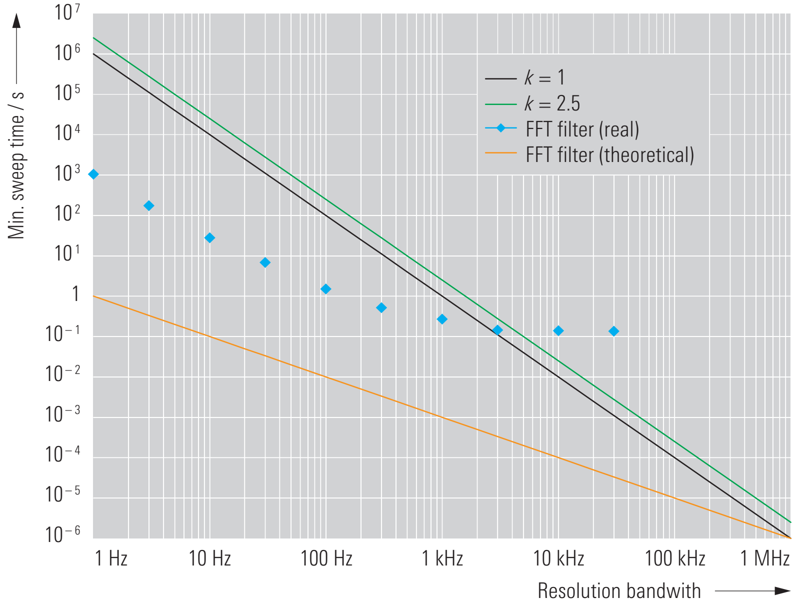


Fig. 4-28 Theoretically required sweep time as a function of resolution bandwidth at a span of 1 MHz. Example of sweep times that can be attained with FFT filters in a modern spectrum analyzer

If the video bandwidth is smaller than the resolution bandwidth, the required minimum sweep time is influenced by the transient time of the video filter. Similar to the IF filter, the transient time of the video filter increases with decreasing bandwidth. The video filter is usually a 1st order lowpass, or a simple RC section if implemented in analog form. Therefore there is a linear relationship between video bandwidth and sweep time. Reducing the video bandwidth by a factor n results in an n times longer sweep time.

Upon failure to attain the minimum sweep time, the IF filter or video filter cannot reach steady state, causing an amplitude loss and distorted signal display (frequency offset). A sinusoidal signal, for instance, would be displayed neither at the correct level nor correct frequency (see Fig. 4-29). Moreover, the effective resolution would be degraded due to the widened signal display.

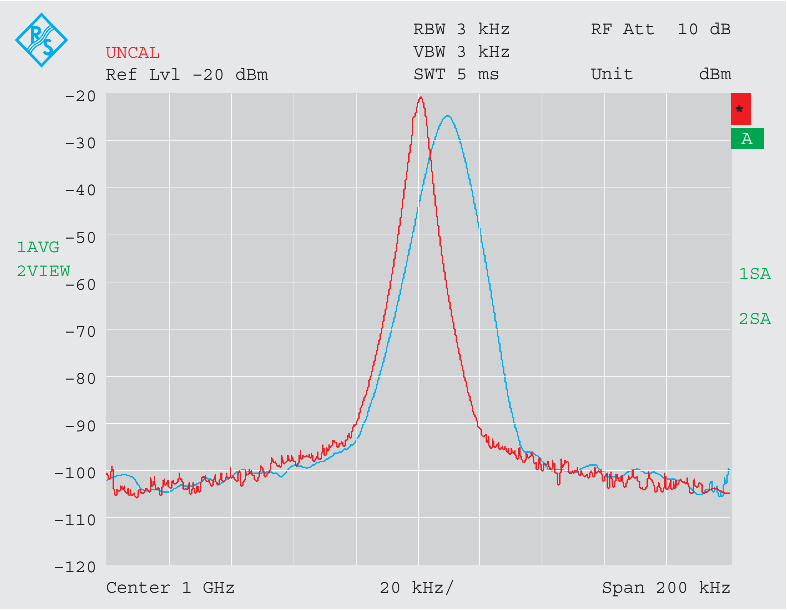


Fig. 4-29 *Amplitude loss if minimum sweep time required is not attained (blue trace)*

To avoid measurement errors due to short sweep times, resolution bandwidth, video bandwidth, sweep time and span are coupled in normal operating mode of modern spectrum analyzers.

Resolution bandwidth is automatically adapted to the selected span. Long sweep times due to narrow resolution bandwidths at large spans or poor resolution due to high resolution bandwidths at small spans are thus avoided. Handling of a spectrum analyzer becomes much easier. The coupling ratio between span and resolution bandwidth can often be set by the user.

Partial coupling of the parameters is also possible. With manual setting of the resolution and video bandwidths, the sweep time can, for instance, be adapted automatically.

When using manual settings, if the minimum sweep time is not adhered to, a warning is usually displayed (UNCAL in Fig. 4-29 upper left corner).

With FFT filters, the transient time is replaced by the observation time required for a specific resolution (Equation 3-4). In contrast to the sweep time with analog or digital filters, the observation time is independent of the span, so even if the span were increased, the observation time would not be increased for constant resolution. The observation time as a function of the resolution (yellow trace) shown in Fig. 4-28 is therefore independent of the span.

In practice, larger spans are made up of several subranges. At a specific resolution, the resulting observation time is required for each subrange. The total observation time is directly proportional to the number of subranges. The attainable measurement time therefore is distinctly longer than the theoretically expected one. Fig. 4-28 shows sweep times that can be attained with a modern spectrum analyzer using FFT filters. It is clearly shown that high span-to-resolution bandwidth ratios allow greatly reduced sweep times with FFT filters, especially when using very narrow resolution bandwidths.

In modern spectrum analyzers, the video bandwidth can be coupled to the resolution bandwidth. When varying the IF bandwidth, the video bandwidth is automatically adapted. The coupling ratio (the ratio between resolution and video bandwidth) depends on the application mode and therefore has to be set by the user (see chapter 4.3). In addition to the user-defined entry of a numeric value, the following options are often available:

- Sine $B_N/B_V = 0.3$ to 1
- Pulse $B_N/B_V = 0.1$
- Noise $B_N/B_V = 10$

In the default setting, the video bandwidth is usually selected so that maximum averaging is achieved without increasing the required sweep time with the video filter. With a proportionality factor $k = 2.5$ (Equation 4-17), the video bandwidth must be at least equal to the resolution bandwidth ($B_N/B_V = 1$). If the IF filter is implemented digitally, a proportionality factor $k = 1$ can be attained through appropriate compensation as described above, and the minimum sweep time required can be reduced by a factor of 2.5. To ensure steady state of the video filter despite the reduced sweep time, the video bandwidth selected should be about three times greater than the resolution bandwidth ($B_N/B_V = 0.3$).

4.6.2 Reference level and RF attenuation

Spectrum analyzers allow measurements in a very wide level range that is limited by the inherent noise and the maximum permissible input level (see chapter 5.1 and chapter 5.4). With modern analyzers this level range may extend from −147 dBm to +30 dBm (with a resolution bandwidth of 10 Hz), thus covering almost 180 dB. It is not possible however to reach the two range limits at a time since they require different settings and the dynamic range of log amplifiers, envelope detectors and A/D converters is much smaller anyway. Within the total level range only a certain window can be used which must be adapted by the user to the specific measurement application by selecting the reference level (maximum signal level to be displayed). The RF attenuation a_{RF} and the IF gain g_{IF} are to be adjusted as a function of the reference level.

To avoid overdriving or even damaging of the first mixer and subsequent processing stages, the high-level input signals must be attenuated by the analyzer's attenuator (see Fig. 4-30). The attenuation required for a specific reference level depends on the dynamic range of the first mixer and subsequent stages. The level at the input of the first mixer (i. e. the mixer level) should be distinctly below the 1 dB compression point. Due to nonlinearities, products are generated in the spectrum analyzer whose levels increase over-proportionally with increasing mixer level. If the mixer level is too high, these products may cause interference in the displayed spectrum so that the socalled intermodulation-free range will be reduced.

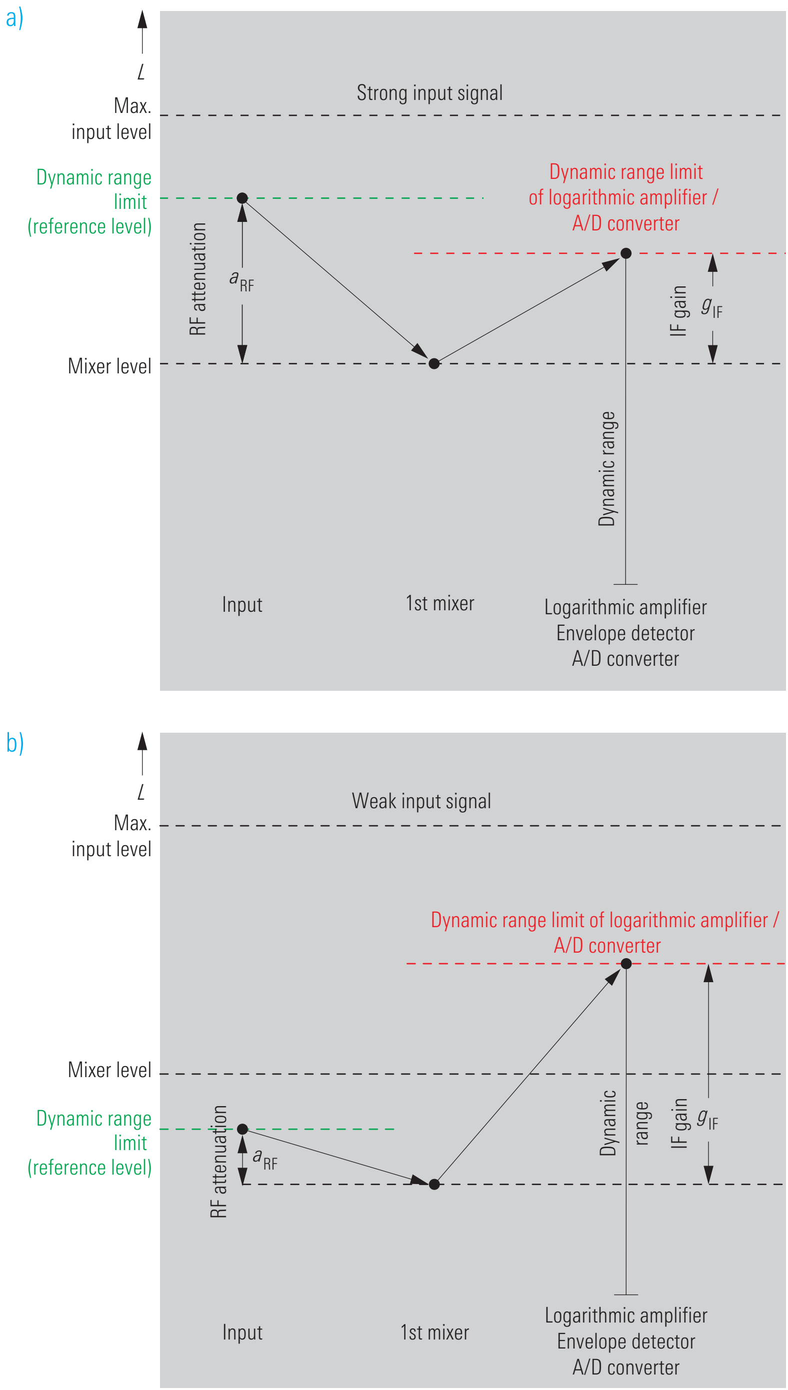


Fig. 4-30 *Adaptation of RF attenuation and IF gain to maximum signal level level to be displayed (max. signal level = reference level)*

Block diagram of spectrum analyzer described in this book