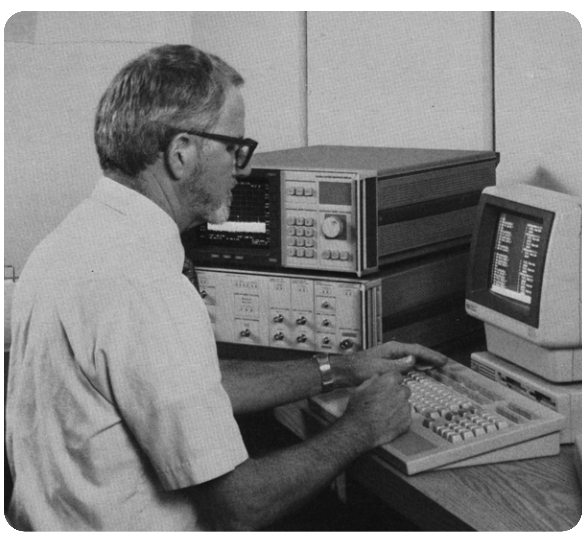
Agilent
Spectrum Analysis Basics

Application Note 150



Agilent Technologies

Innovating the HP Way

Contents

Agilent Technologies Signal Analysis Division would like to acknowledge the author, Blake Peterson, for more than 40 years of outstanding service in engineering applications and technical education for Agilent and our customers.

Chapter 1
Introduction

This application note is intended to serve as a primer on superheterodyne spectrum analyzers. Such analyzers can also be described as frequency-selective, peak-responding voltmeters calibrated to display the rms value of a sine wave. It is important to understand that the spectrum analyzer is not a power meter, although we normally use it to display power directly. But as long as we know some value of a sine wave (for example, peak or average) and know the resistance across which we measure this value, we can calibrate our voltmeter to indicate power.

What is a spectrum?

Before we get into the details of describing a spectrum analyzer, we might first ask ourselves: just what is a spectrum and why would we want to analyze it?

Our normal frame of reference is time. We note when certain events occur. This holds for electrical events, and we can use an oscilloscope to view the instantaneous value of a particular electrical event (or some other event converted to volts through an appropriate transducer) as a function of time; that is, to view the waveform of a signal in the time domain.

Enter Fourier.[1] He tells us that any time-domain electrical phenomenon is made up of one or more sine waves of appropriate frequency, amplitude, and phase. Thus with proper filtering we can decompose the waveform of figure 1 into separate sine waves, or spectral components, that we can then evaluate independently. Each sine wave is characterized by an amplitude and a phase. In other words, we can transform a time-domain signal into its frequency-domain equivalent. In general, for RF and microwave signals, preserving the phase information complicates this transformation process without adding significantly to the value of the analysis. Therefore, we are willing to do without the phase information. If the signal that we wish to analyze is periodic, as in our case here, Fourier says that the constituent sine waves are separated in the frequency domain by 1/T, where T is the period of the signal.[2]

To properly make the transformation from the time to the frequency domain, the signal must be evaluated over all time, that is, over ± infinity. However, we normally take a shorter, more practical view and assume that signal behavior over several seconds or minutes is indicative of the overall characteristics of the signal. The transformation can also be made from the frequency to the time domain, according to Fourier. This case requires the evaluation of all spectral components over frequencies to ± infinity, and the phase of the individual components is indeed critical. For example, a square wave transformed to the frequency domain and back again could turn into a saw tooth wave if phase were not preserved.

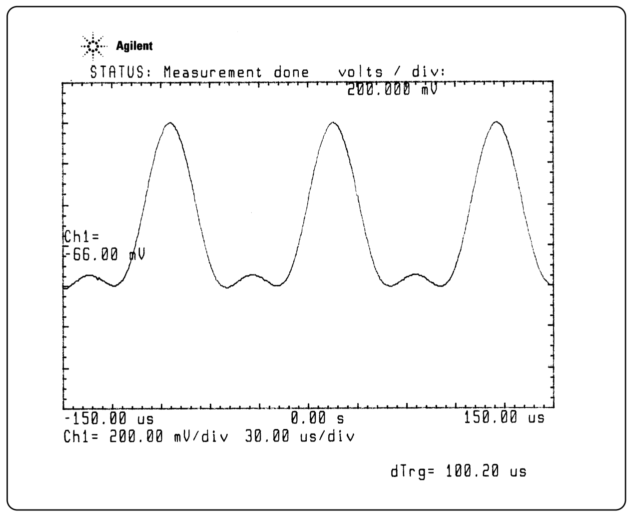


Figure 1. Complex time-domain signal

1 Jean Baptiste Joseph Fourier, 1768-1830, French mathematician and physicist.
2 If the time signal occurs only once, then T is infinite, and the frequency representation is a continuum of sine waves.

So what is a spectrum in the context of this discussion? A collection of sine waves that, when combined properly, produce the time-domain signal under examination. Figure 1 shows the waveform of a complex signal. Suppose that we were hoping to see a sine wave. Although the waveform certainly shows us that the signal is not a pure sinusoid, it does not give us a definitive indication of the reason why.

Figure 2 shows our complex signal in both the time and frequency domains. The frequency-domain display plots the amplitude versus the frequency of each sine wave in the spectrum. As shown, the spectrum in this case comprises just two sine waves. We now know why our original waveform was not a pure sine wave. It contained a second sine wave, the second harmonic in this case.

Are time-domain measurements out? Not at all. The time domain is better for many measurements, and some can be made only in the time domain. For example, pure time-domain measurements include pulse rise and fall times, overshoot, and ringing.

Why measure spectra?

The frequency domain has its measurement strengths as well. We have already seen in figures 1 and 2 that the frequency domain is better for determining the harmonic content of a signal. Communications people are extremely interested in harmonic distortion. For example, cellular radio systems must be checked for harmonics of the carrier signal that might interfere with other systems operating at the same frequencies as the harmonics. Communications people are also interested in distortion of the message modulated onto a carrier. Third-order intermodulation (two tones of a complex signal modulating each other) can be particularly troublesome because the distortion components can fall within the band of interest and so not be filtered away.

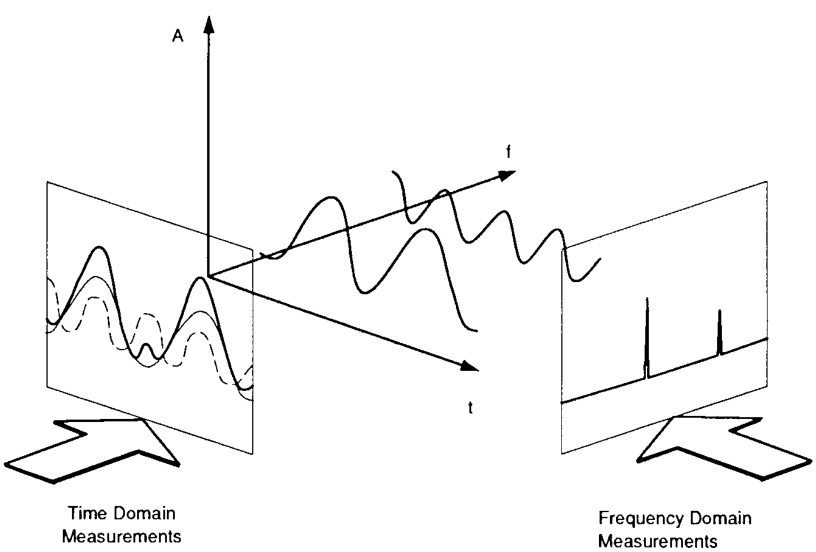


Figure 2. Relationship between time and frequency domain

Spectral occupancy is another important frequency-domain measurement. Modulation on a signal spreads its spectrum, and to prevent interference with adjacent signals, regulatory agencies restrict the spectral bandwidth of various transmissions. Electromagnetic interference (EMI) might also be considered a form of spectral occupancy. Here the concern is that unwanted emissions, either radiated or conducted (through the power lines or other interconnecting wires), might impair the operation of other systems. Almost anyone designing or manufacturing electrical or electronic products must test for emission levels versus frequency according to one regulation or another.

So frequency-domain measurements do indeed have their place. Figures 3 through 6 illustrate some of these measurements.

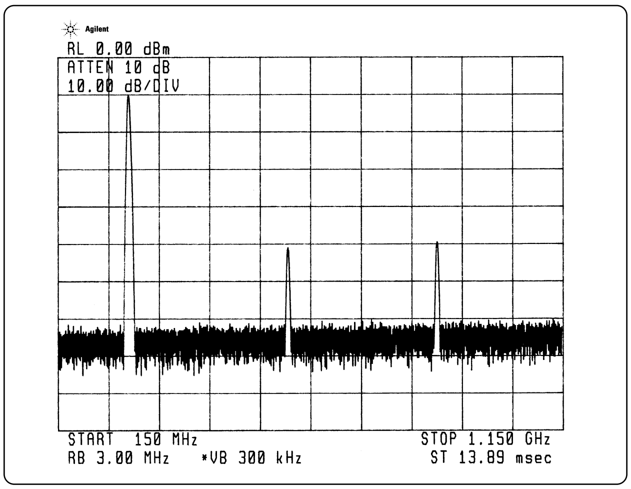


Figure 3. Harmonic distortion test

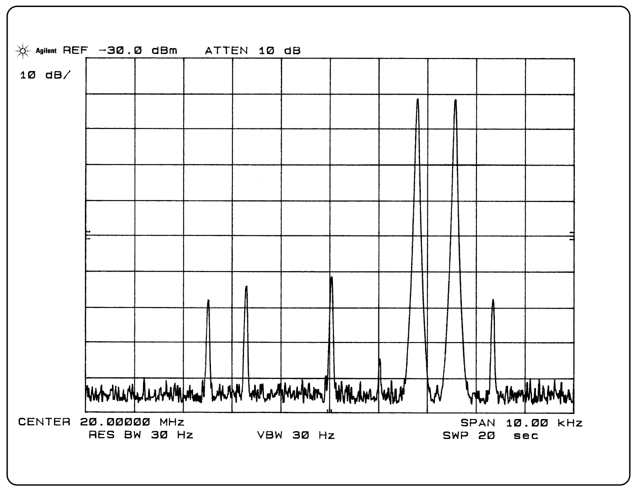


Figure 4. Two-tone test on SSB transmitter

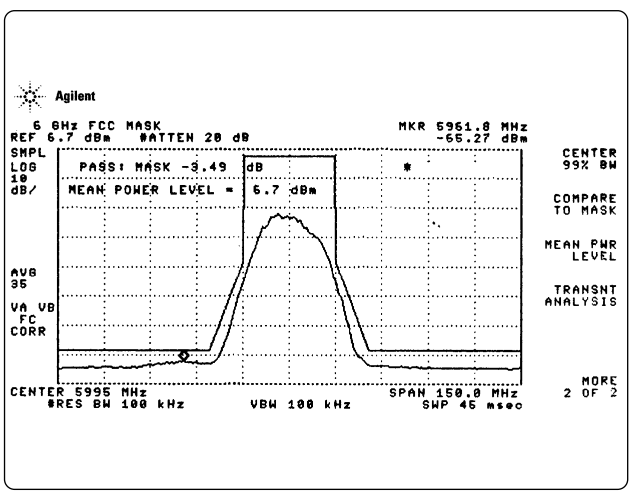


Figure 5. Digital radio signal and mask showing limits of spectral occupancy

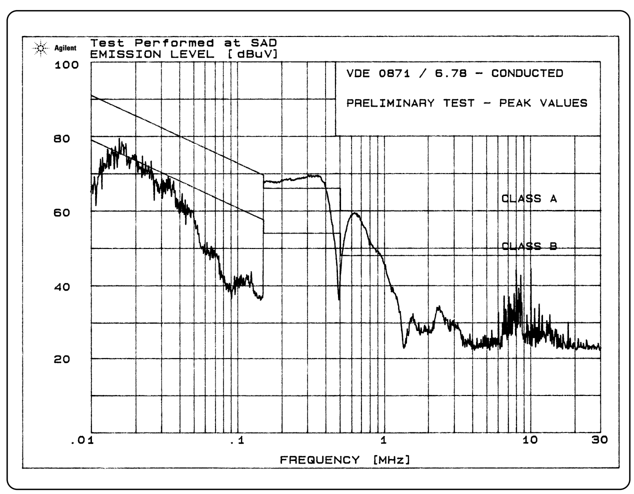


Figure 6. Conducted emissions plotted against VDE limits as part of EMI test

Chapter 2
The superheterodyne spectrum analyzer

While we shall concentrate on the superheterodyne spectrum analyzer in this note, there are several other spectrum analyzer architectures. Perhaps the most important non-superheterodyne type is that which digitizes the time-domain signal and then performs a Fast Fourier Transform (FFT) to display the signal in the frequency domain. One advantage of the FFT approach is its ability to characterize single-shot phenomena. Another is that phase as well as magnitude can be measured. However, at the present state of technology, FFT machines do have some limitations relative to the superheterodyne spectrum analyzer, particularly in the areas of frequency range, sensitivity, and dynamic range.

Figure 7 is a simplified block diagram of a superheterodyne spectrum analyzer. Heterodyne means to mix - that is, to translate frequency - and super refers to super-audio frequencies, or frequencies above the audio range. Referring to the block diagram in figure 7, we see that an input signal passes through a low-pass filter (later we shall see why the filter is here) to a mixer, where it mixes with a signal from the local oscillator (LO). Because the mixer is a non-linear device, its output includes not only the two original signals but also their harmonics and the sums and differences of the original frequencies and their harmonics. If any of the mixed signals falls within the passband of the intermediate-frequency (IF) filter, it is further processed (amplified and perhaps logged), essentially rectified by the envelope detector, digitized (in most current analyzers), and applied to the vertical plates of a cathode-ray tube (CRT) to produce a, vertical deflection on the CRT screen (the display). A ramp generator deflects the CRT beam horizontally across the screen from left to right.[1] The ramp also tunes the LO so that its frequency changes in proportion to the ramp voltage.

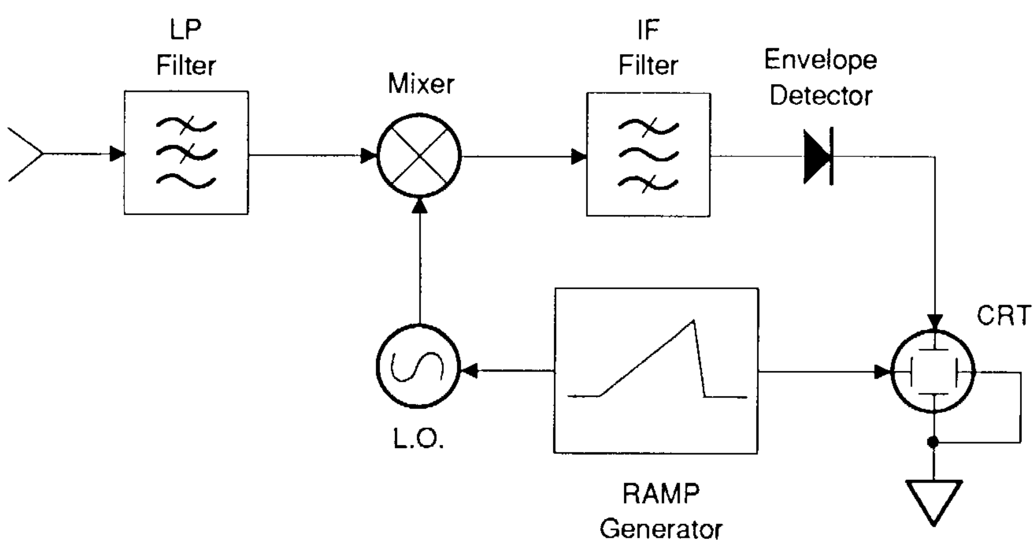


Figure 7. Superheterodyne spectrum analyzer

1 Not exactly true for analyzers with digital displays. However, describing the ramp as if it did directly control the CRT beam simplifies the discussion, so we shall continue to do so. See CRT Displays.

If you are familiar with superheterodyne AM radios, the type that receive ordinary AM broadcast signals, you will note a strong similarity between them and the block diagram of figure 7. The differences are that the output of a spectrum analyzer is the screen of a CRT instead of a speaker, and the local oscillator is tuned electronically rather than purely by a front-panel knob.

Since the output of a spectrum analyzer is an X-Y display on a CRT screen, let's see what information we get from it. The display is mapped on a grid (graticule) with ten major horizontal divisions and generally eight or ten major vertical divisions. The horizontal axis is calibrated in frequency that increases linearly from left to right. Setting the frequency is usually a two-step process. First we adjust the frequency at the centerline of the graticule with the center frequency control. Then we adjust the frequency range (span) across the full ten divisions with the Frequency Span control. These controls are independent, so if we change the center frequency, we do not alter the frequency span. Some spectrum analyzers allow us to set the start and stop frequencies as an alternative to setting center frequency and span. In either case, we can determine the absolute frequency of any signal displayed and the frequency difference between any two signals.

The vertical axis is calibrated in amplitude. Virtually all analyzers offer the choice of a linear scale calibrated in volts or a logarithmic scale calibrated in dB. (Some analyzers also offer a linear scale calibrated in units of power.) The log scale is used far more often than the linear scale because the log scale has a much wider usable range. The log scale allows signals as far apart in amplitude as 70 to 100 dB (voltage ratios of 3100 to 100,000 and power ratios of 10,000,000 to 10,000,000,000) to be displayed simultaneously. On the other hand, the linear scale is usable for signals differing by no more than 20 to 30 dB (voltage ratios of 10 to 30). In either case, we give the top line of the graticule, the reference level, an absolute value through calibration techniques[1] and use the scaling per division to assign values to other locations on the graticule. So we can measure either the absolute value of a signal or the amplitude difference between any two signals.

In older spectrum analyzers, the reference level in the log mode could be calibrated in only one set of units. The standard set was usually dBm (dB relative to 1 mW). Only by special request could we get our analyzer calibrated in dBmV or dBuV (dB relative to a millivolt or a microvolt, respectively). The linear scale was always calibrated in volts. Today's analyzers have internal microprocessors, and they usually allow us to select any amplitude units (dBm, dBuV, dBmV, or volts) on either the log or the linear scale.

Scale calibration, both frequency and amplitude, is shown either by the settings of physical switches on the front panel or by annotation written onto the display by a microprocessor. Figure 8 shows the display of a typical microprocessor-controlled analyzer.

But now let's turn our attention back to figure 7.

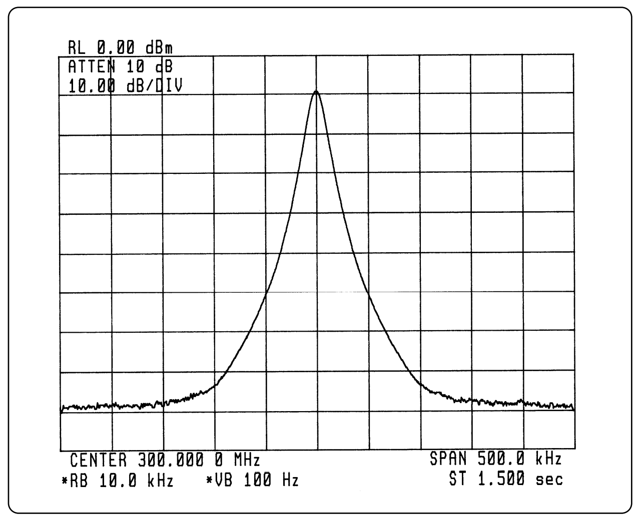


Figure 8. Typical spectrum analyzer display with control settings

1 See amplitude accuracy.

Tuning equation

To what frequency is the spectrum analyzer of figure 7 tuned? That depends. Tuning is a function of the center frequency of the IF filter, the frequency range of the LO, and the range of frequencies allowed to reach the mixer from the outside world (allowed to pass through the low-pass filter). Of all the products emerging from the mixer, the two with the greatest amplitudes and therefore the most desirable are those created from the sum of the LO and input signal and from the difference between the LO and input signal. If we can arrange things so that the signal we wish to examine is either above or below the LO frequency by the IF, one of the desired mixing products will fall within the pass-band of the IF filter and be detected to create a vertical deflection on the display.

How do we pick the LO frequency and the IF to create an analyzer with the desired frequency range? Let us assume that we want a tuning range from 0 to 2.9 GHz. What IF should we choose? Suppose we choose 1 GHz. Since this frequency is within our desired tuning range, we could have an input signal at 1 GHz. And since the output of a mixer also includes the original input signals, an input signal at 1 GHz would give us a constant output from the mixer at the IF. The 1 GHz signal would thus pass through the system and give us a constant vertical deflection on the display regardless of the tuning of the LO. The result would be a hole in the frequency range at which we could not properly examine signals because the display deflection would be independent of the LO.

So we shall choose instead an IF above the highest frequency to which we wish to tune. In Agilent spectrum analyzers that tune to 2.9 GHz, the IF chosen is about 3.6 (or 3.9) GHz. Now if we wish to tune from 0 Hz (actually from some low frequency because we cannot view a to 0-Hz signal with this architecture) to 2.9 GHz, over what range must the LO tune? If we start the LO at the IF (LO - IF = 0) and tune it upward from there to 2.9 GHz above the IF, we can cover the tuning range with the LO-minus-IF mixing product. Using this information, we can generate a tuning equation:

$$f_{sig} = f_{LO} - f_{IF}$$

where f_{sig} = signal frequency,
f_{LO} = local oscillator frequency, and
f_{IF} = intermediate frequency (IF).

If we wanted to determine the LO frequency needed to tune the analyzer to a low-, mid-, or high-frequency signal (say, 1 kHz, 1.5 GHz, and 2.9 GHz), we would first restate the tuning equation in terms of f_{LO}:

$$f_{LO} = f_{sig} + f_{IF}.$$

Then we would plug in the numbers for the signal and IF:

$$f_{LO} = 1 \text{ kHz} + 3.6 \text{ GHz} = 3.600001 \text{ GHz},$$
$$f_{LO} = 1.5 \text{ GHz} + 3.6 \text{ GHz} = 5.1 \text{ GHz, and}$$
$$f_{LO} = 2.9 \text{ GHz}; + 3.6 \text{ GHz} = 6.5 \text{ GHz}.$$

Figure 9 illustrates analyzer tuning. In the figure, f_{LO} is not quite high enough to cause the $f_{LO} - f_{sig}$ mixing product to fall in the IF passband, so there is no response on the display. If we adjust the ramp generator to tune the LO higher, however, this mixing product will fall in the IF passband at some point on the ramp (sweep), and we shall see a response on the display.

Since the ramp generator controls both the horizontal position of the trace on the display and the LO frequency, we can now calibrate the horizontal axis of the display in terms of input-signal frequency.

We are not quite through with the tuning yet. What happens if the frequency of the input signal is 8.2 GHz? As the LO tunes through its 3.6-to-6.5-GHz range, it reaches a frequency (4.6 GHz) at which it is the IF away from the 8.2-GHz signal, and once again we have a mixing product equal to the IF, creating a deflection on the display. In other words, the tuning equation could just as easily have been

$$F_{sig} = f_{LO} + f_{IF}.$$

This equation says that the architecture of figure 7 could also result in a tuning range from 7.2 to 10.1 GHz. But only if we allow signals in that range to reach the mixer. The job of the low-pass filter in figure 7 is to prevent these higher frequencies from getting to the mixer. We also want to keep signals at the intermediate frequency itself from reaching the mixer, as described above, so the low-pass filter must do a good job of attenuating signals at 3.6 GHz as well as in the range from 7.2 to 10.1 GHz.

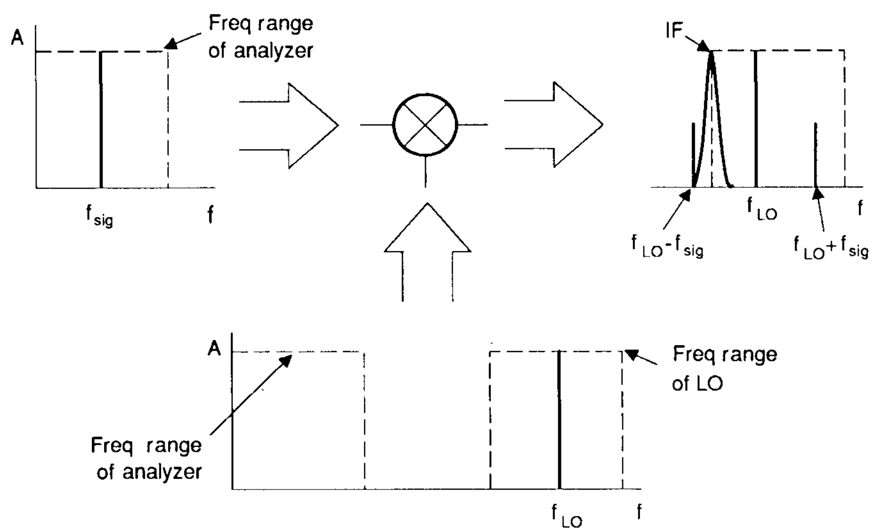


Figure 9. The LO must be tuned to $f_{IF} + f_{sig}$ to produce a response on the display.

In summary, we can say that for a single-band RF spectrum analyzer, we would choose an IF above the highest frequency of the tuning range, make the LO tunable from the IF to the IF plus the upper limit of the tuning range, and include a low-pass filter in front of the mixer that cuts off below the IF.

To separate closely spaced signals (see Resolution below), some spectrum analyzers have IF bandwidths as narrow as 1 kHz; others, 100 Hz; still others, 10 Hz. Such narrow filters are difficult to achieve at a center frequency of 3.6 GHz. So we must add additional mixing stages, typically two to four, to down-convert from the first to the final IF. Figure 10 shows a possible IF chain based on the architecture of the Agilent 71100. The full tuning equation for the 71100 is:

$$f_{sig} = f_{LO1} - (f_{LO2} + f_{LO3} + f_{LO4} + f_{final\ IF}).$$

However,

$$f_{LO2} + f_{LO3} + f_{LO4} + f_{final\ IF}$$

$$= 3.3\text{ GHz} + 300\text{ MHz} + 18.4\text{ MHz} + 3\text{ MHz}$$
$$= 3.6214\text{ GHz, the first IF.}$$

So simplifying the tuning equation by using just the first IF leads us to the same answers. Although only passive filters are shown in the diagram, the actual implementation includes amplification in the narrower IF stages, and a logarithmic amplifier is part of the final IF section.[1]

Most RF analyzers allow an LO frequency as low as and even below the first IF. Because there is not infinite isolation between the LO and IF ports of the mixer, the LO appears at the mixer output. When the LO equals the IF, the LO signal itself is processed by the system and appears as a response on the display. This response is called LO feed through. LO feed through actually can be used as a 0-Hz marker.

An interesting fact is that the LO feed through marks 0 Hz with no error. When we use an analyzer with non-synthesized LOs, frequency uncertainty can be ±5 MHz or more, and we can have a tuning uncertainty of well over 100% at low frequencies. However, if we use the LO feed through to indicate 0 Hz and the calibrated frequency span to indicate frequencies relative to the LO feed through, we can improve low-frequency accuracy considerably. For example, suppose we wish to tune to a 10-kHz signal on an analyzer with 5-MHz tuning uncertainty and 3% span accuracy. If we rely on the tuning accuracy, we might find the signal with the analyzer tuned anywhere from -4.99 to 5.01 MHz. On the other hand, if we set our analyzer to a 20-kHz span and adjust tuning to set the LO feed through at the left edge of the display, the 10-kHz signal appears within ±0.15 division of the center of the display regardless of the indicated center frequency.

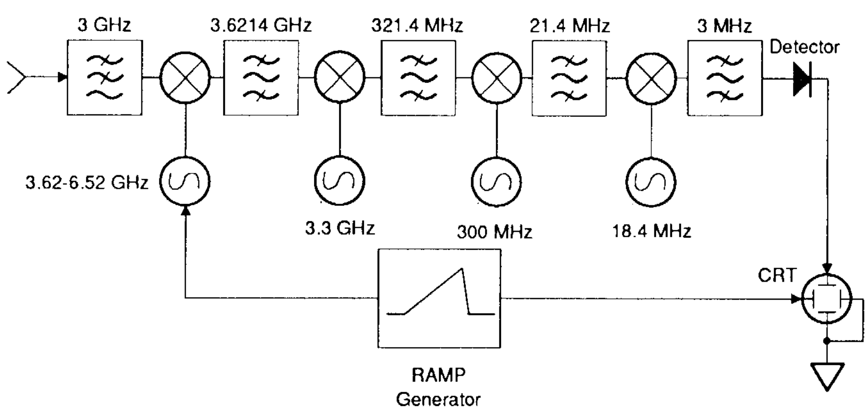


Figure 10. Most spectrum analyzers use two to four mixing steps to reach the final IF

1 In the text we shall round off some of the frequency values for simplicity although the exact values are shown in the figures.

Resolution

Analog filters

Frequency resolution is the ability of a spectrum analyzer to separate two input sinusoids into distinct responses. But why should resolution even be a problem when Fourier tells us that a signal (read sine wave in this case) has energy at only one frequency? It seems that two signals, no matter how close in frequency, should appear as two lines on the display. But a closer look at our superheterodyne receiver shows why signal responses have a definite width on the display. The output of a mixer includes the sum and difference products plus the two original signals (input and LO). The intermediate frequency is determined by a bandpass filter, and this filter selects the desired mixing product and rejects all other signals. Because the input signal is fixed and the local oscillator is swept, the products from the mixer are also swept. If a mixing product happens to sweep past the IF, the characteristics of the bandpass filter are traced on the display. See figure 11. The narrowest filter in the chain determines the overall bandwidth, and in the architecture of figure 10, this filter is in the 3-MHz IF.

So unless two signals are far enough apart, the traces they make fall on top of each other and look like only one response. Fortunately, spectrum analyzers have selectable resolution (IF) filters, so it is usually possible to select one narrow enough to resolve closely spaced signals.

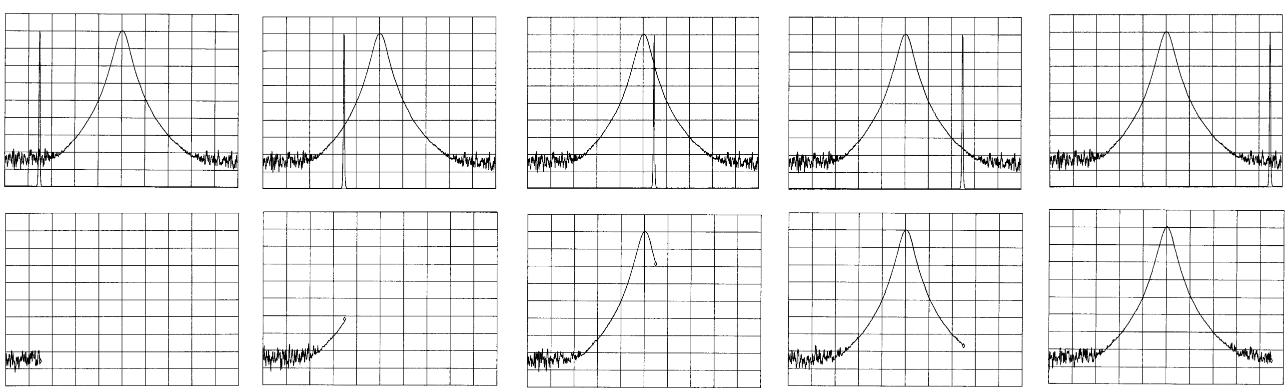


Figure 11. As a mixing product sweeps past the IF filter, the filter shape is traced on the display

Agilent data sheets indicate resolving power by listing the 3-dB bandwidths of the available IF filters. This number tells us how close together equal-amplitude sinusoids can be and still be resolved. In this case there will be about a 3-dB dip between the two peaks traced out by these signals. See figure 12. The signals can be closer together before their traces merge completely, but the 3-dB bandwidth is a good rule of thumb for resolution of equal-amplitude signals.[1]

More often than not we are dealing with sinusoids that are not equal in amplitude. What then? The smaller sinusoid can be lost under the skirt of the response traced out by the larger. This effect is illustrated in figure 13. Thus another specification is listed for the resolution filters: bandwidth selectivity (or selectivity or shape factor). For Agilent analyzers, bandwidth selectivity is specified as the ratio of the 60-dB bandwidth to the 3-dB bandwidth, as shown in figure 14. Some analyzer manufacturers specify the 60:6 dB ratio. The analog filters in Agilent analyzers are synchronously-tuned, have four or five poles, and are nearly Gaussian in shape. Bandwidth selectivity varies from 25:1 for the wider filters of older Agilent analyzers to 11:1 for the narrower filters of newer stabilized and high-performance analyzers.

So what resolution bandwidth must we choose to resolve signals that differ by 4 kHz and 30 dB, assuming 11:1 bandwidth selectivity? We shall start by assuming that the analyzer is in its most commonly used mode: a logarithmic amplitude and a linear frequency scale. In this mode, it is fairly safe to assume that the filter skirt is straight between the 3- and 60-dB points, and since we are concerned with rejection of the larger signal when the analyzer is tuned to the smaller signal, we need to consider not the full band-width but the frequency difference from the filter center frequency to the skirt. To determine how far down the filter skirt is at a given offset, we have:

$$-3\text{ dB} - [(\text{Offset} - BW_{3dB}/2)/(BW_{60dB}/2 - BW_{3dB}/2)]^{*}\text{Diff}_{60,3dB},$$

where

Offset = frequency separation of two signals,
BW_{3dB} = 3-dB bandwidth,
BW_{60dB} = 60-dB bandwidth, and
$\text{Diff}_{60,3dB}$ = difference between 60 and 3 dB (57 dB).

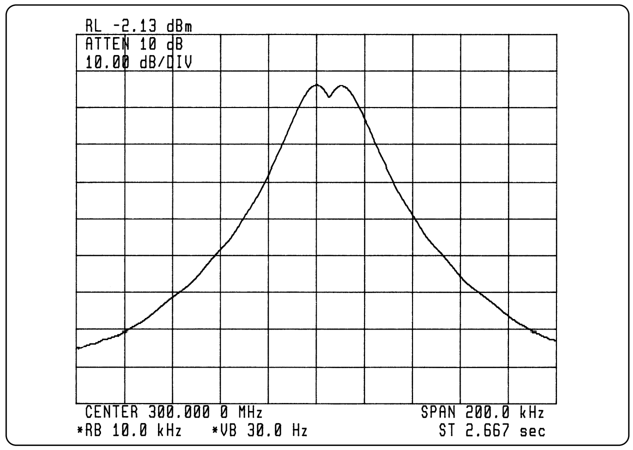


Figure 12. Two equal- amplitude sinusoids separated by the 3 dB BW of the selected IF filter can be resolved

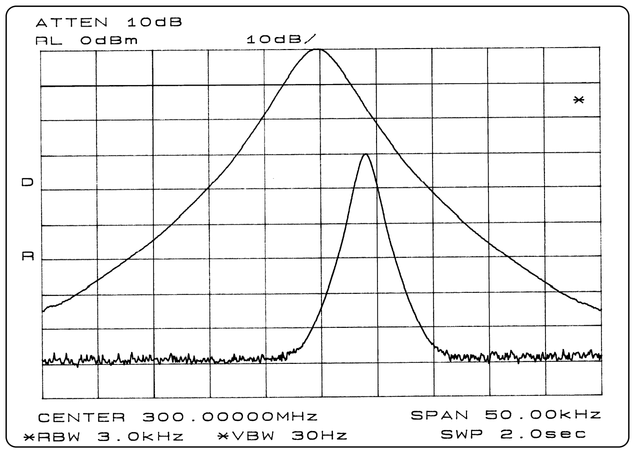


Figure 13. A low-level signal can be lost under skirt of the response to a larger signal

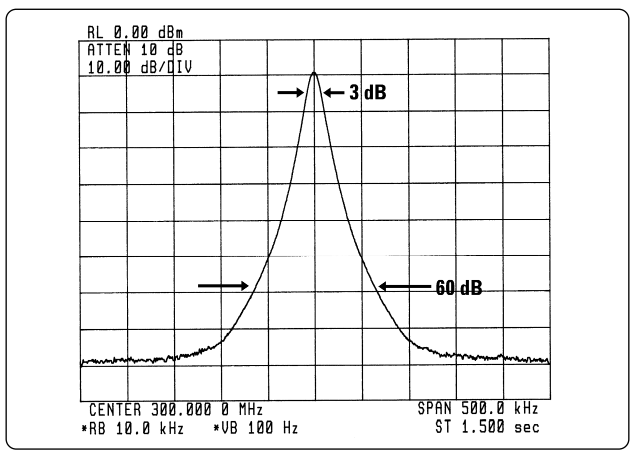


Figure 14. Bandwidth selectivity, ratio of 60 dB to 3 dB bandwidths, helps determine resolving power for unequal sinusoids

1 If you experiment with resolution on a spectrum analyzer that has an analog display or one that has a digital display and Rosenfell display mode, use enough video filtering to create a smooth trace; otherwise, there will be a smearing as the two signals interact. While the smeared trace certainly indicates the presence of more than one signal, it is difficult to determine the amplitudes of the individual signals from that trace. Spectrum analyzers with digital displays and positive peak as their normal display mode may not show the smearing effect. You can observe the smearing by se-lecting the alternate sample display mode.

Let's try the 3-kHz filter. At 60 dB down it is about 33 kHz wide but only about 16.5 kHz from center frequency to the skirt. At an offset of 4 kHz, the filter skirt is down:

$-3 - [(4 - 3/2)/(33/2 - 3/2)]*(60 - 3) = -12.5$ dB,

not far enough to allow us to see the smaller signal. If we use numbers for the 1-kHz filter, we find that the filter skirt is down:

$-3 - [4 - 1/2)/(11/2 - 1/2)1*(60 - 3) = -42.9$ dB,

and the filter resolves the smaller signal. See figure 15. Figure 16 shows a plot of typical resolution versus signal separation for several resolution bandwidths in the 8566B.

Digital filters

Some spectrum analyzers, such as the Agilent 8560 and ESA-E Series, use digital techniques to realize their narrower resolution-bandwidth filters (100 Hz and below for the 8560 family, 300 Hz and below for the ESA-E Series family). As shown in figure 17, the linear analog signal is mixed down to 4.8 kHz and passed through a bandpass filter only 600 Hz wide. This IF signal is then amplified, sampled at a 6.4-kHz rate, and digitized.

Once in digital form, the signal is put through a Fast Fourier Transform algorithm. To transform the appropriate signal, the analyzer must be fixed-tuned (not sweeping); that is, the transform must be done on a time-domain signal. Thus the 8560 analyzers step in 600-Hz increments, instead of sweeping continuously, when we select one of the digital resolution bandwidths. This stepped tuning can be seen on the display, which is updated in 600-Hz increments as the digital processing is completed. The ESA-E Series uses a similar scheme and updates its displays in approximately 900-Hz increments.

An advantage of digital processing as done in these analyzers is a bandwidth selectivity of 5:1[1]. And, this selectivity is available on the narrowest filters, the ones we would choose to separate the most closely spaced signals.

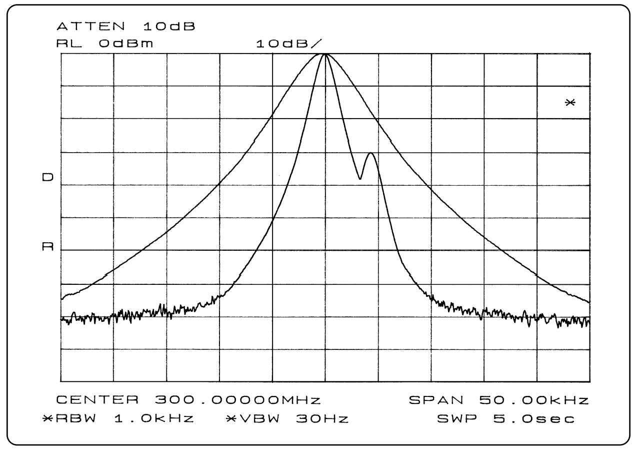


Figure 15. The 3 kHz filter does not resolve smaller signal - the 1 kHz filter does

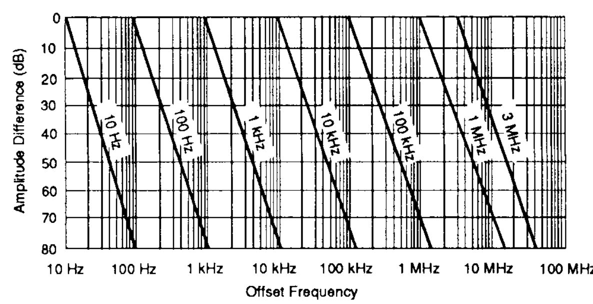


Figure 16. Resolution versus offset for the 8566B

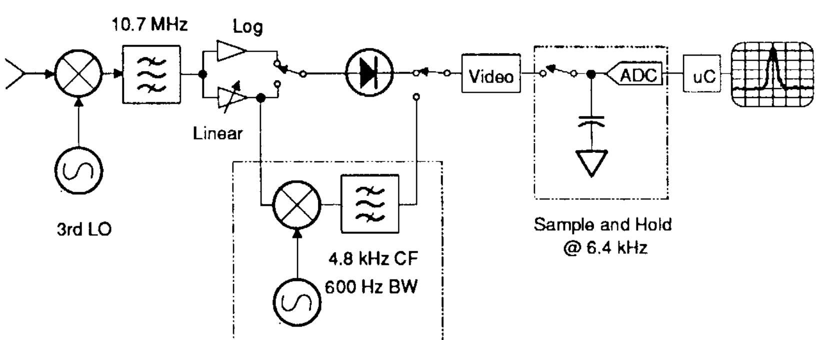


Figure 17. Digital implementation of 10, 30, and 100 Hz resolution filters in 8560A, 8561B, and 8563A

1 Also see sweep time.

Residual FM

Is there any other factor that affects the resolution of a spectrum analyzer? Yes, the stability of the LOs in the analyzer, particularly the first LO. The first LO is typically a YIG-tuned oscillator (tuning somewhere in the 2 to 7 GHz range), and this type of oscillator has a residual FM of 1 kHz or more. This instability is transferred to any mixing products resulting from the LO and incoming signals, and it is not possible to determine whether the signal or LO is the source of the instability.

The effects of LO residual FM are not visible on wide resolution bandwidths. Only when the bandwidth approximates the peak-to-peak excursion of the FM does the FM become apparent. Then we see that a ragged-looking skirt as the response of the resolution filter is traced on the display. If the filter is narrowed further, multiple peaks can be produced even from a single spectral component. Figure 18 illustrates the point. The widest response is created with a 3-kHz bandwidth; the middle, with a 1-kHz bandwidth; the innermost, with a 100-Hz bandwidth. Residual FM in each case is about 1 kHz.

So the minimum resolution bandwidth typically found in a spectrum analyzer is determined at least in part by the LO stability. Low-cost analyzers, in which no steps are taken to improve upon the inherent residual FM of the YIG oscillators, typically have a minimum bandwidth of 1 kHz. In mid-performance analyzers, the first LO is stabilized and filters as narrow as 10 Hz are included. Higher-performance analyzers have more elaborate synthesis schemes to stabilize all their LOs and so have bandwidths down to 1 Hz. With the possible exception of economy analyzers, any instability that we see on a spectrum analyzer is due to the incoming signal.

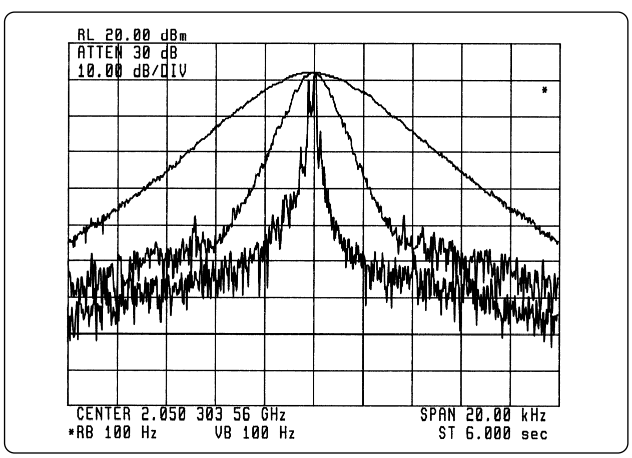


Figure 18. LO residual FM is seen only when the resolution bandwidth is less than the peak-to-peak FM

Phase noise

Even though we may not be able to see the actual frequency jitter of a spectrum analyzer LO system, there is still a manifestation of the LO frequency or phase instability that can be observed: phase noise (also called sideband noise). No oscillator is perfectly stable. All are frequency- or phase-modulated by random noise to some extent. As noted above, any instability in the LO is transferred to any mixing products resulting from the LO and input signals, so the LO phase-noise modulation sidebands appear around any spectral component on the display that is far enough above the broadband noise floor of the system (figure 19). The amplitude difference between a displayed spectral component and the phase noise is a function of the stability of the LO. The more stable the LO, the farther down the phase noise. The amplitude difference is also a function of the resolution bandwidth. If we reduce the resolution bandwidth by a factor of ten, the level of the phase noise decreases by 10 dB[1].

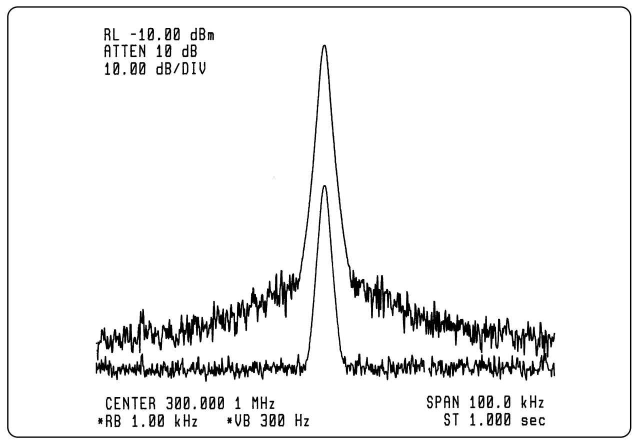


Figure 19. Phase noise is displayed only when a signal is displayed far enough above the system noise floor

1 The effect is the same for the broadband noise floor (or any broadband noise signal). See sensitivity.

The shape of the phase-noise spectrum is a function of analyzer design. In some analyzers the phase noise is a relatively flat pedestal out to the bandwidth of the stabilizing loop. In others, the phase noise may fall away as a function of frequency offset from the signal. Phase noise is specified in terms of dBc or dB relative to a carrier. It is sometimes specified at a specific frequency offset; at other times, a curve is given to show the phase-noise characteristics over a range of offsets.

Generally, we can see the inherent phase noise of a spectrum analyzer only in the two or three narrowest resolution filters, when it obscures the lower skirts of these filters. The use of the digital filters described above does not change this effect. For wider filters, the phase noise is hidden under the filter skirt just as in the case of two unequal sinusoids discussed earlier.

In any case, phase noise becomes the ultimate limitation in an analyzer's ability to resolve signals of unequal amplitude. As shown in figure 20, we may have determined that we can resolve two signals based on the 3-dB bandwidth and selectivity, only to find that the phase noise covers up the smaller signal.

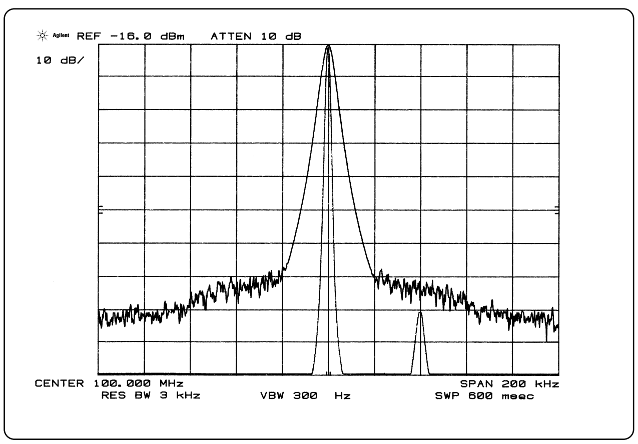


Figure 20. Phase noise can prevent resolution of unequal signals

Sweep time

Analog resolution filters

If resolution was the only criterion on which we judged a spectrum analyzer, we might design our analyzer with the narrowest possible resolution (IF) filter and let it go at that. But resolution affects sweep time, and we care very much about sweep time. Sweep time directly affects bow long it takes to complete a measurement.

Resolution comes into play because the IF filters are band-limited circuits that require finite times to charge and discharge. If the mixing products are swept through them too quickly, there will be a loss of displayed amplitude as shown in figure 21. (See Envelope Detector below for another approach to IF response time.) If we think about bow long a mixing product stays in the passband of the IF filter, that time is directly proportional to bandwidth and inversely proportional to the sweep in Hz per unit time, or:

Time in passband =
(RBW)/[(Span)/(ST)] = [(RBW)(ST)]/(Span),

where RBW = resolution bandwidth and
ST = sweep time.

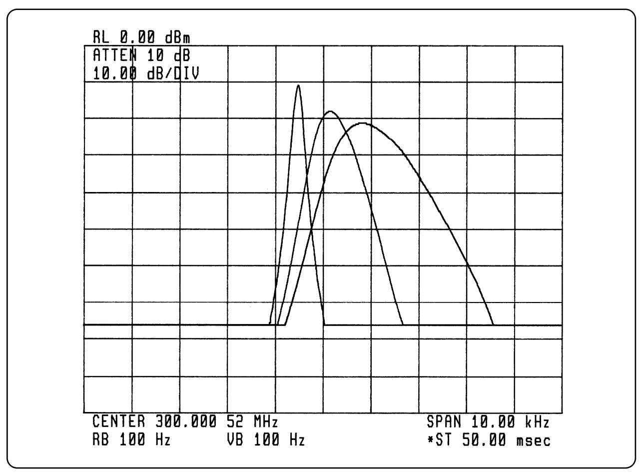


Figure 21. Sweeping an analyzer too fast causes a drop in displayed amplitude and a shift in indicated frequency

On the other hand, the rise time of a filter is inversely proportional to its bandwidth, and if we include a constant of proportionality, k, then:

Rise time = k/(RBW).

If we make the times equal and solve for sweep time, we have:

k/(RBW) = [(RBW)(ST)]/(Span), or:

$$ST = k(\text{Span})/(\text{RBW})^2.$$

The value of k is in the 2 to 3 range for the synchronously-tuned, near-Gaussian filters used in Agilent analyzers. For more nearly square, stagger-tuned filters, k is 10 to 15.

The important message here is that a change in resolution has a dramatic effect on sweep time. Some spectrum analyzers have resolution filters selectable only in decade steps, so selecting the next filter down for better resolution dictates a sweep time that goes up by a factor of 100!

How many filters, then, would be desirable in a spectrum analyzer? The example above seems to indicate that we would want enough filters to provide something less than decade steps. Most Agilent analyzers provide values in a 1,3,10 sequence or in ratios roughly equaling the square root of 10. So sweep time is affected by a factor of about 10 with each step in resolution. Some series of Agilent spectrum analyzers offer bandwidth steps of just 10% for an even better compromise among span, resolution, and sweep time.

Most spectrum analyzers available today automatically couple sweep time to the span and resolution-bandwidth settings. Sweep time is adjusted to maintain a calibrated display. If a sweep time longer than the maximum available is called for, the analyzer indicates that the display is uncalibrated. We are allowed to override the automatic setting and set sweep time manually if the need arises.

Digital resolution filters

The digital resolution filters used in the Agilent 8560 and ESA-E Series have an effect on sweep time that is different from the effects we've just discussed for analog filters. For the Agilent 8560 family, this difference occurs because the signal being analyzed is processed in 600-Hz blocks. So when we select the 10-Hz resolution bandwidth, the analyzer is in effect simultaneously processing the data in each 600-Hz block through 60 contiguous 10-Hz filters. If the digital processing were instantaneous, we would expect a factor-of-60 reduction in sweep time. The reduction factor is somewhat less, but is significant nonetheless.

Envelope detector

Spectrum analyzers typically convert the IF signal to video[1] with an envelope detector. In its simplest form, an envelope detector is a diode followed by a parallel RC combination. See figure 22. The output of the IF chain, usually a sine wave, is applied to the detector. The time constants of the detector are such that the voltage across the capacitor equals the peak value of the IF signal at all times; that is, the detector can follow the fastest possible changes in the envelope of the IF signal but not the instantaneous value of the IF sine wave itself (nominally 3, 10.7, or 21.4 MHz in Agilent spectrum analyzers).

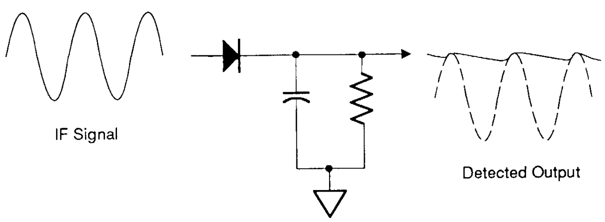


Figure 22. Envelope detector

For most measurements, we choose a resolution bandwidth narrow enough to resolve the individual spectral components of the input signal. If we fix the frequency of the LO so that our analyzer is tuned to one of the spectral components of the signal, the output of the IF is a steady sine wave with a constant peak value. The output of the envelope detector will then be a constant (dc) voltage, and there is no variation for the detector to follow.

However, there are times when we deliberately choose a resolution bandwidth wide enough to include two or more spectral components. At other times, we have no choice. The spectral components are closer in frequency than our narrowest bandwidth. Assuming only two spectral components within the passband, we have two sine waves interacting to create a beat note, and the envelope of the IF signal varies as shown in figure 23 as the phase between the two sine waves varies.

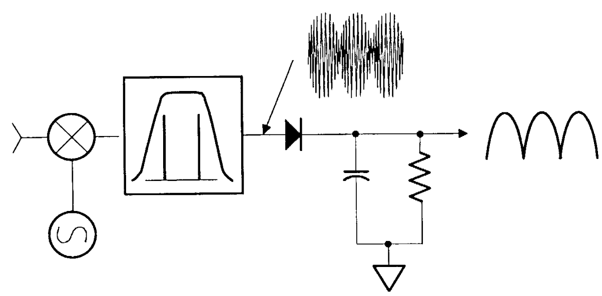


Figure 23. Output of the envelope detector follows the peaks of the IF signal.

1 A signal whose frequency range extends from zero (dc) to some upper frequency determined by the circuit elements.

What determines the maximum rate at which the envelope of the IF signal can change? The width of the resolution (IF) filter. This bandwidth determines how far apart two input sinusoids can be and, after the mixing process, be within the filter at the same time.[1] If we assume a 21.4-MHz final IF and a 100-kHz bandwidth, two input signals separated by 100 kHz would produce, with the appropriate LO frequency and two or three mixing processes, mixing products of 21.35 and 21.45 MHz and so meet the criterion. See figure 23. The detector must be able to follow the changes in the envelope created by these two signals but not the nominal 21.4 MHz IF signal itself.

The envelope detector is what makes the spectrum analyzer a voltmeter. If we duplicate the situation above and have two equal-amplitude signals in the passband of the IF at the same time, what would we expect to see on the display? A power meter would indicate a power level 3 dB above either signal; that is, the total power of the two. Assuming that the two signals are close enough so that, with the analyzer tuned half way between them, there is negligible attenuation due to the roll-off of the filter, the analyzer display will vary between a value twice the voltage of either (6 dB greater) and zero (minus infinity on the log scale). We must remember that the two signals are sine waves (vectors) at different frequencies, and so they continually change in phase with respect to each other. At some time they add exactly in phase; at another, exactly out of phase.

So the envelope detector follows the changing amplitude values of the peaks of the signal from the IF chain but not the instantaneous values. And gives the analyzer its voltmeter characteristics.

Although the digitally-implemented resolution bandwidths do not have an analog envelope detector, one is simulated for consistency with the other bandwidths.

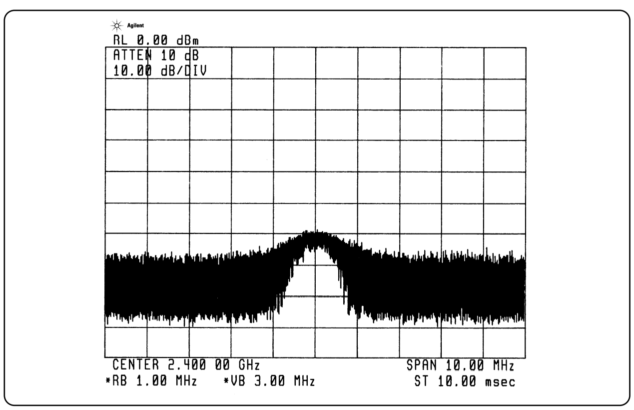


Figure 24. Spectrum analyzers display signal plus noise

Display smoothing

Video filtering

Spectrum analyzers display signals plus their own internal noise,[2] as shown in figure 24. To reduce the effect of noise on the displayed signal amplitude, we often smooth or average the display, as shown in figure 25. All Agilent superheterodyne analyzers include a variable video filter for this purpose. The video filter is a low-pass filter that follows the detector and determines the bandwidth of the video circuits that drive the vertical deflection system of the display. As the cutoff frequency of the video filter is reduced to the point at which it becomes equal to or less than the bandwidth of the selected resolution (IF) filter, the video system can no longer follow the more rapid variations of the envelope of the signal(s) passing through the IF chain. The result is an averaging or smoothing of the displayed signal.

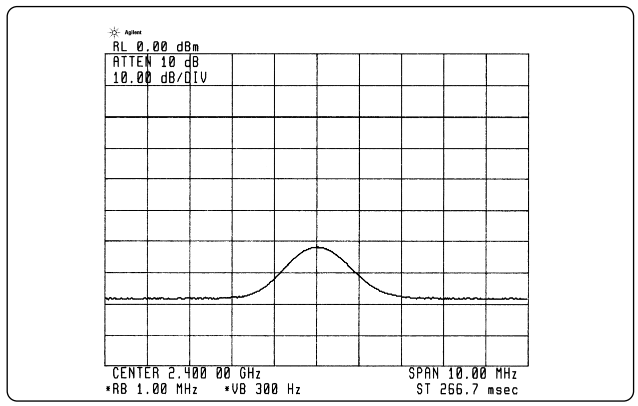


Figure 25. Display of figure 24 after full smoothing

1 For this discussion, we assume that the filter is perfectly rectangular.
2 See dynamic range versus measurement uncertainty.

The effect is most noticeable in measuring noise, particularly when a wide resolution bandwidth is used. As we reduce the video bandwidth, the peak-to-peak variations of the noise are reduced. As figure 26 shows, the degree of reduction (degree of averaging or smoothing) is a function of the ratio of the video to resolution bandwidths. At ratios of 0.01 or less, the smoothing is very good; at higher ratios, not so good. That part of the trace that is already smooth - for example, a sinusoid displayed well out of the noise - is not affected by the video filter.

(If we are using an analyzer with a digital display and "pos peak" display mode[1], we notice two things: changing the resolution bandwidth does not make much difference in the peak-to-peak fluctuations of the noise, and changing the video bandwidth seems to affect the noise level. The fluctuations do not change much because the analyzer is displaying only the peak values of the noise. However, the noise level appears to change with video bandwidth because the averaging [smoothing] changes, thereby changing the peak values of the noise. See figure 27. We can select sample detection to get the full effect.)

Because the video filter has its own response time, the sweep time equation becomes: $ST = k(Span)/[(RBW)(VBW)]$ when the video bandwidth is equal to or less than the resolution bandwidth. However, sweep time is affected only when the value of the signal varies over the span selected. For example, if we were experimenting with the analyzer's own noise in the example above, there would be no need to slow the sweep because the average value of the noise is constant across a very wide frequency range. On the other hand, if there is a discrete signal in addition to the noise, we must slow the sweep to allow the video filter to respond to the voltage changes created as the mixing product of the discrete signal sweeps past the IF. Those analyzers that set sweep time automatically account for video bandwidth as well as span and resolution bandwidth.

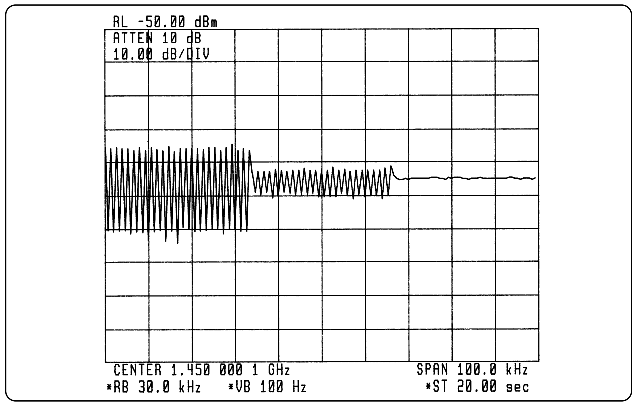


Figure 26. Smoothing effect of video-resolution bandwidth ratios of 3, 1/10, and 1/100 (on a single sweep)

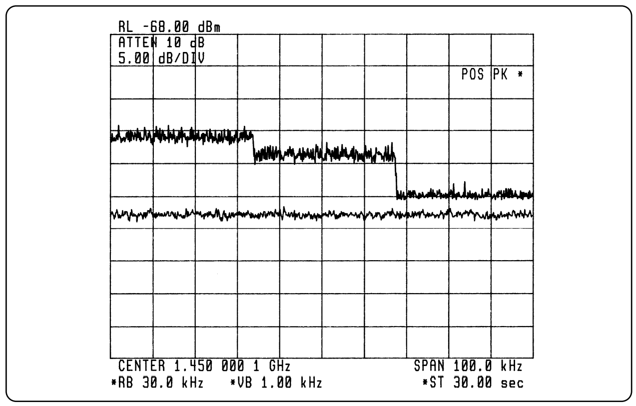


Figure 27. On analyzers using pos peak display mode, reducing the video bandwidth lowers the peak noise but not the average noise. Lower trace shows average noise.

1 See digital displays.

Video averaging

Analyzers with digital displays often offer another choice for smoothing the display: video averaging. In this case, averaging is accomplished over two or more sweeps on a point-by-point basis. At each display point, the new value is averaged in with the previously averaged data:

$$A_{avg} = [(n - 1)/n]A_{prior\ avg} + (1/n)A_n,$$

where A_{avg} = new average value,
$A_{prior\ avg}$ = average from prior sweep,
A_n= measured value on current sweep, and
n = number of current sweep.

Thus the display gradually converges to an average over a number of sweeps. As with video filtering, we can select the degree of averaging or smoothing by setting the number of sweeps over which the averaging occurs. Figure 28 shows video averaging for different numbers of sweeps. While video averaging has no effect on sweep time, the time to reach a given degree of averaging is about the same as with video filtering because of the number of sweeps required[1].

Which form of display smoothing should we pick? In many cases, it does not matter. If the signal is noise or a low-level sinusoid very close to the noise, we get the same results with either video filtering or video averaging.

However, there is a distinct difference between the two. Video filtering performs averaging in real time; that is, we see the full effect of the averaging or smoothing at each point on the display as the sweep progresses. Each point is averaged only once, for a time of about 1/VBW on each sweep. Video averaging, on the other hand, requires multiple sweeps to achieve the full degree of averaging, and the averaging at each point takes place over the full time period needed to complete the multiple sweeps.

As a result, we can get significantly different results from the two averaging methods on certain signals. For example, a signal with a spectrum that changes with time can yield a different average on each sweep when we use video filtering. However, if we choose video averaging over many sweeps, we shall get a value much closer to the true average. See figure 29.

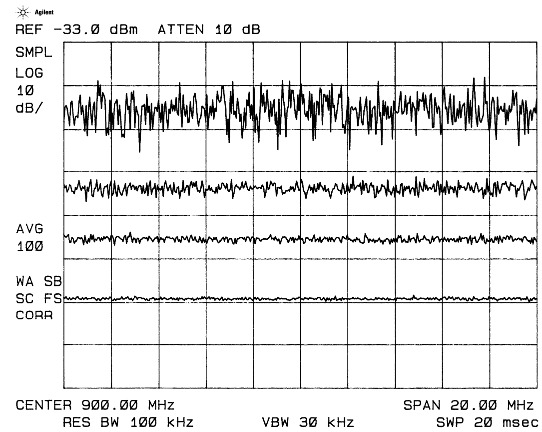


Figure 28. Effect of video (digital) averaging for 1, 5, 20, and 100 sweeps (top to bottom)

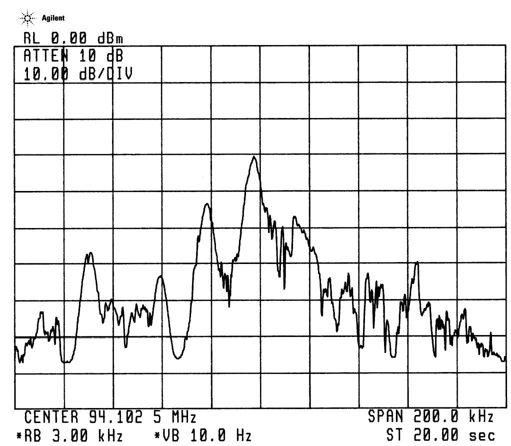


Figure 29A. Analog filtering

Figure 29B. Digital averaging

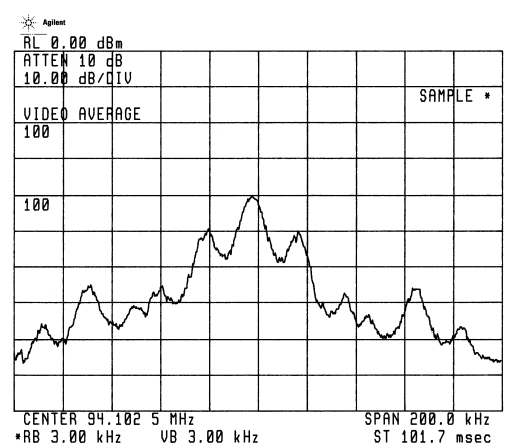


Figure 29. Video (analog) filtering and video (digital) averaging yield different results on an FM broadcast signal

1 Most analyzers automatically switch to a sample display mode when video averaging is selected. See digital displays for potential loss of signal information in the sample mode.

Amplitude measurements

CRT displays

Up until the mid-1970s, spectrum analyzer displays were purely analog. That is, the output of the envelope detector was simply amplified and applied directly to the vertical plates of the CRT. This mode of operation meant that the CRT trace presented a continuous indication of the signal envelope, and no information was lost. However, analog displays had drawbacks. The major problem was in handling the long sweep times when narrow resolution bandwidths were used. In the extreme case, the display became a spot that moved slowly across the CRT with no real trace on the display. Even with long-persistence phosphors such as P7, a meaningful display was not possible with the longer sweep times. One solution in those days was time-lapse photography.

Another solution was the storage CRT. These tubes included mechanisms to store a trace so that it could be displayed on the screen for a reasonable length of time before it faded or became washed out. Initially, storage was binary in nature. We could choose permanent storage, or we could erase the display and start over. Hewlett-Packard (now Agilent) pioneered a variable-persistence mode in which we could adjust the fade rate of the display. When properly adjusted, the old trace would just fade out at the point where the new trace was updating the display. The idea was to provide a display that was continuous, had no flicker, and avoided confusing overwrites. The system worked quite well with the correct trade-off between trace intensity and fade rate. The difficulty was that the intensity and the fade rate had to be readjusted for each new measurement situation.

When digital circuitry became affordable in the mid-1970s, it was quickly put to use in spectrum analyzers. Once a trace had been digitized and put into memory, it was permanently available for display. It became an easy matter to update the display at a flicker-free rate without blooming or fading. The data in memory was updated at the sweep rate, and since the contents of memory were written to the display at a flicker-free rate, we could follow the updating as the analyzer swept through its selected frequency span just as we could with analog systems.

Digital displays

But digital systems were had problems of their own. What value should be displayed? As figure 30 shows, no matter how many data points we use across the CRT, each point must represent what has occurred over some frequency range and, although we usually do not think in terms of time when dealing with a spectrum analyzer, over some time interval. Let us imagine the situation illustrated in figure 30: we have a display that contains a single CW signal and otherwise only noise. Also, we have an analog system whose output we wish to display as faithfully as possible using digital techniques.

As a first method, let us simply digitize the instantaneous value of the signal at the end of each interval (also called a cell or bucket). This is the sample mode. To give the trace a continuous look, we design a system that draws vectors between the points. From the conditions of figure 30, it appears that we get a fairly reasonable display, as shown in figure 31. Of course, the more points in the trace, the better the replication of the analog signal. The number of points is limited, with 400, 600, 800, and 1,000 being typical.[1] As shown in figure 32, more points do indeed get us closer to the analog signal.

While the sample mode does a good job of indicating the randomness of noise, it is not a good mode for a spectrum analyzer's usual function: analyzing sinusoids. If we were to look at a 100-MHz comb on the Agilent 71210, we might set it to span from 0 to 22 GHz. Even with 1,000 display points, each point represents a span of 22 MHz, far wider than the maximum 3-MHz resolution bandwidth.

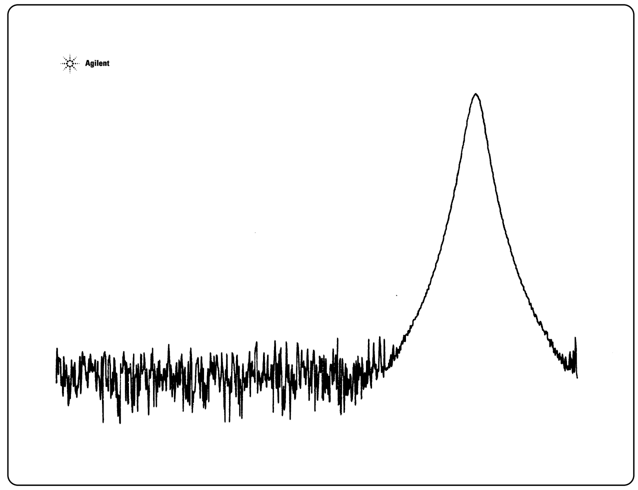


Figure 30. When digitizing an analog signal, what value should be displayed at each point?

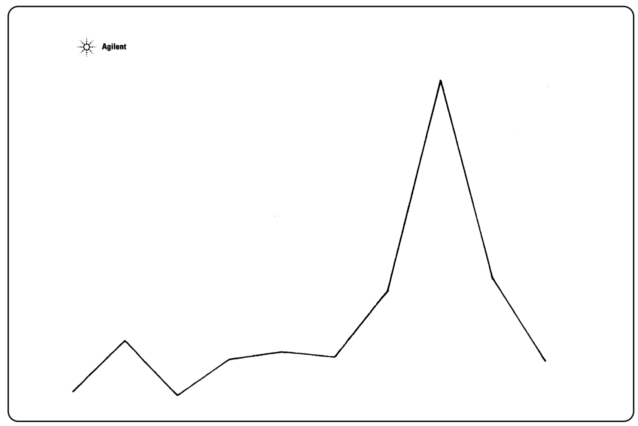


Figure 31. The sample display mode using ten points to display the signal of figure 30

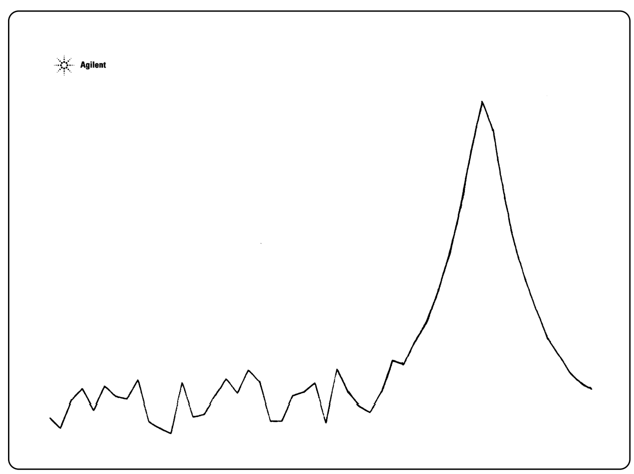


Figure 32. More points produce a display closer to an analog display

1 The Agilent ESA-E and 71000 families allow selection of the number of trace points, 2-8192, for the ESA-E, 3-1024, for the 71000.

As a result, the true amplitude of a comb tooth is shown only if its mixing product happens to fall at the center of the IF when the sample is taken. Figure 33 shows a 5-GHz span with a 1-MHz bandwidth; the comb teeth should be relatively equal in amplitude. Figure 34 shows a 500-MHz span comparing the true comb with the results from the sample mode; only a few points are used to exaggerate the effect. (The sample trace appears shifted to the left because the value is plotted at the beginning of each interval.)

One way to insure that all sinusoids are reported is to display the maximum value encountered in each cell. This is the positive-peak display mode, or pos peak. This display mode is illustrated in figure 35. Figure 36 compares pos peak and sample display modes. Pos peak is the normal or default display mode offered on many spectrum analyzers because it ensures that no sinusoid is missed, regardless of the ratio between resolution bandwidth and cell width. However, unlike sample mode, pos peak does not give a good representation of random noise because it captures the crests of the noise. So spectrum analyzers using the pos peak mode as their primary display mode generally also offer the sample mode as an alternative.

To provide a better visual display of random noise than pos peak and yet avoid the missed-signal problem of the sample mode, the Rosenfell display mode is offered on many spectrum analyzers. Rosenfell is not a person's name but rather a description of the algorithm that tests to see if the signal rose and fell within the cell represented by a given data point. Should the signal both rise and fall, as determined by pos-peak and neg-peak detectors, then the algorithm classifies the signal as noise. In that case, an odd-numbered data point indicates the maximum value encountered during its cell. On the other band, an even-numbered data point indicates the minimum value encountered during its cell. Rosenfell and sample modes are compared in figure 37.

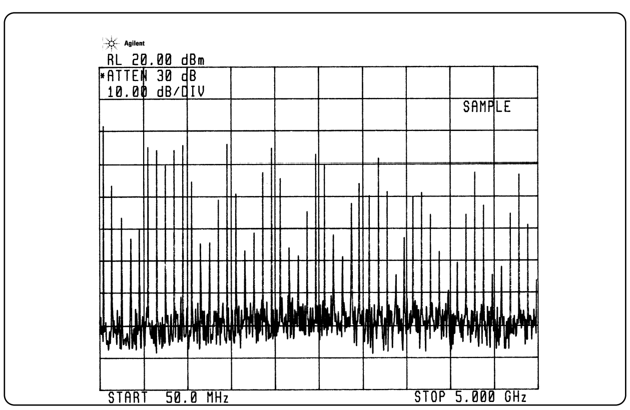


Figure 33. A 5-GHz span of a 100-MHz comb in the sample display mode. The actual comb values are relatively constant over this range.

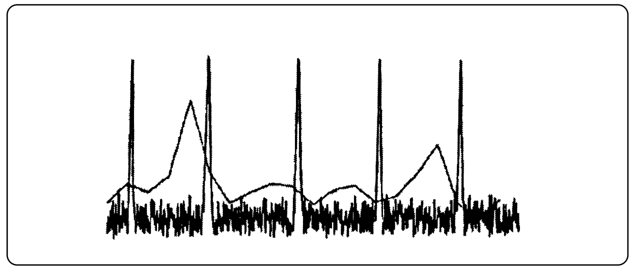


Figure 34. The actual comb and results of the sample display mode over a 500-MHz span. When resolution bandwidth is narrower than the sample interval, the sample mode can give erroneous results. (The sample trace has only 20 points to exaggerate the effect.)

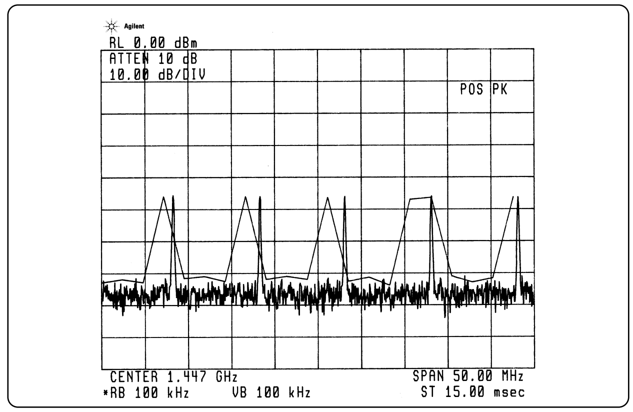


Figure 35. Pos peak display mode versus actual comb

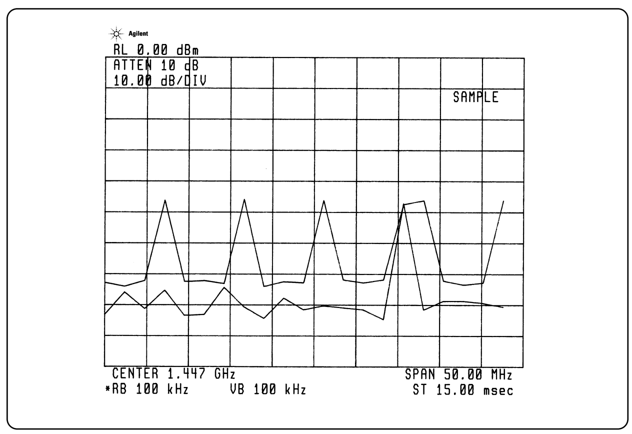


Figure 36. Comparison of sample and pos peak display modes

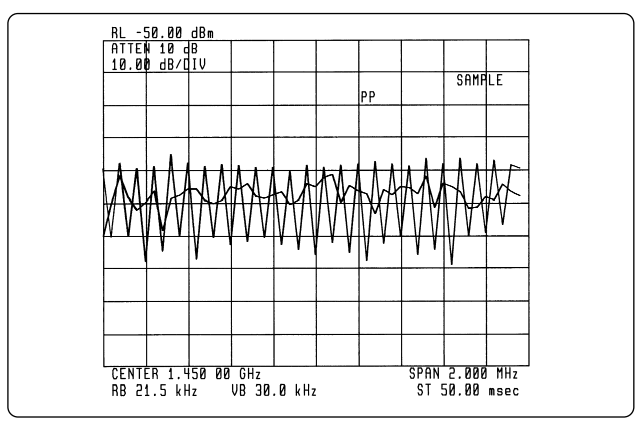


Figure 37. Comparison of Rosenfell and sample display modes

What happens when a sinusoidal signal is encountered? We know that as a mixing product is swept past the IF filter, an analyzer traces out the shape of the filter on the display. If the filter shape is spread over many display points, then we encounter a situation in which the displayed signal only rises as the mixing product approaches the center frequency of the filter and only falls as the mixing product moves away from the filter center frequency. In either of these cases, the pos-peak and neg-peak detectors sense an amplitude change in only one direction, and, according to the Rosenfell algorithm, the maximum value in each cell is displayed. See figure 38.

What happens when the resolution bandwidth is narrow relative to a cell? If the peak of the response occurs anywhere but at the very end of the cell, the signal will both rise and fall during the cell. If the cell happens to be an odd-numbered one, all is well. The maximum value encountered in the cell is simply plotted as the next data point. However, if the cell is even-numbered, then the minimum value in the cell is plotted. Depending on the ratio of resolution bandwidth to cell width, the minimum value can differ from the true peak value (the one we want displayed) by a little or a lot. In the extreme, when the cell is much wider than the resolution bandwidth, the difference between the maximum and minimum values encountered in the cell is the full difference between the peak signal value and the noise. Since the Rosenfell algorithm calls for the minimum value to be indicated during an even-numbered cell, the algorithm must include some provision for preserving the maximum value encountered in this cell.

To ensure no loss of signals, the pos-peak detector is reset only after the peak value has been used on the display. Otherwise, the peak value is carried over to the next cell. Thus when a signal both rises and falls in an even-numbered cell, and the minimum value is displayed, the pos-peak detector is not reset. The pos-peak value is carried over to the next cell, an odd-numbered cell. During this cell, the pos-peak value is updated only if the signal value exceeds the value carried over. The displayed value, then, is the larger of the held-over value and the maximum value encountered in the new, odd-numbered cell. Only then is the pos-peak detector reset.

This process may cause a maximum value to be displayed one data point too far to the right, but the offset is usually only a small percentage of the span. Figure 39 shows what might happen in such a case. A small number of data points exaggerate the effect.

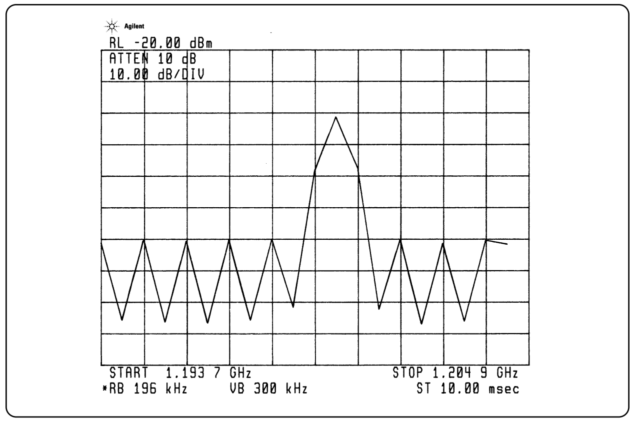


Figure 38. When detected signal only rises or falls, as when mixing product sweeps past resolution filter, Rosenfell displays maximum values

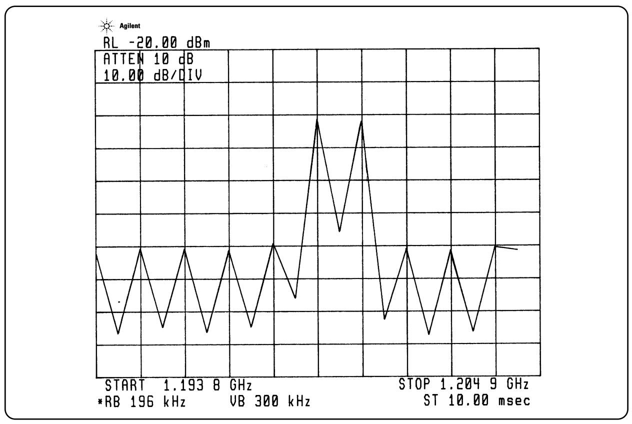


Figure 39. Rosenfell when signal peak falls between data points (fewer trace points exaggerate the effect)

The Rosenfell display mode does a better job of combining noise and discrete spectral components on the display than does pos peak. We get a much better feeling for the noise with Rosenfell. However, because it allows only maxima and minima to be displayed, Rosenfell does not give us the true randomness of noise as the sample mode does. For noise signals, then, the sample display mode is the best.

Agilent analyzers that use Rosenfell as their default, or normal, display mode also allow selection of the other display modes - pos peak, neg peak, and sample.

As we have seen, digital displays distort signals in the process of getting them to the screen. However, the pluses of digital displays greatly outweigh the minuses. Not only can the digital information be stored indefinitely and refreshed on the screen without flicker, blooming, or fade, but once data is in memory, we can add capabilities such as markers and display arithmetic or output data to a computer for analysis or further digital signal processing.

Amplitude accuracy

Now that we have our signal displayed on the CRT, let's look at amplitude accuracy. Or, perhaps better, amplitude uncertainty. Most spectrum analyzers these days are specified in terms of both absolute and relative accuracy. However, relative performance affects both, so let us look at those factors affecting relative measurement uncertainty first.

Relative uncertainty

When we make relative measurements on an incoming signal, we use some part of the signal as a reference. For example, when we make second-harmonic distortion measurements, we use the fundamental of the signal as our reference. Absolute values do not come into play[1]; we are interested only in how the second harmonic differs in amplitude from the fundamental.

So what factors come into play? Table 1 gives us a reasonable shopping list. The range of values given covers a wide variety of spectrum analyzers. For example, frequency response, or flatness, is frequency-range dependent. A low-frequency RF analyzer might have a frequency response of ±0.5 dB[2]. On the other hand, a microwave spectrum analyzer tuning in the 20-GHz range could well have a frequency response in excess of ±4 dB. Display fidelity covers a variety of factors. Among them are the log amplifier (how true the logarithmic characteristic is), the detector (how linear), and the digitizing circuits (how linear). The CRT itself is not a factor for those analyzers using digital techniques and offering digital markers because the marker information is taken from trace memory, not the CRT. The display fidelity is better over small amplitude differences, so a typical specification for display fidelity might read 0.1 dB/dB, but no more than the value shown in table 1 for large amplitude differences.

Table 1. Amplitude uncertainties

Relative		**±dB**
	Frequency response	0.5-4
	Display fidelity	0.5-2
	ΔRF attenuator	0.5-2
	ΔIF attenuator/gain	0.1-1
	ΔResolution bandwidth	0.1-1
	ΔCRT scaling	0.1-1
Absolute		
	Calibrator	0.2-1

The remaining items in the table involve control changes made during the course of a measurement. See figure 40. Because an RF input attenuator must operate over the entire frequency range of the analyzer, its step accuracy, like frequency response, is a function of frequency. At low RF frequencies, we expect the attenuator to be quite good; at 20 GHz, not as good. On the other hand, the IF attenuator (or gain control) should be more accurate because it operates at only one frequency. Another parameter that we might change during the course of a measurement is resolution bandwidth. Different filters have different insertion losses. Generally we see the greatest difference when switching between inductor-capacitor (LC) filters, typically used for the wider resolution bandwidths, and crystal filters. Finally, we may wish to change display scaling from, say, 10 dB/div to 1 dB/div or linear.

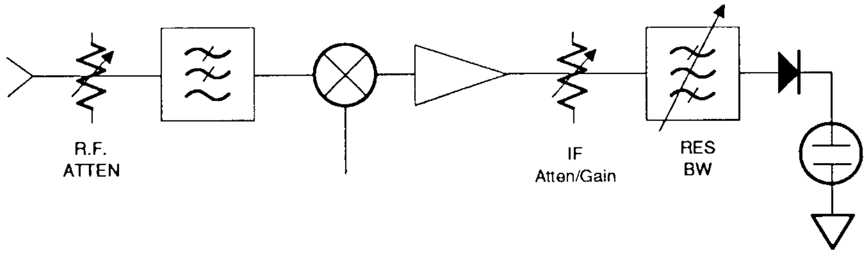


Figure 40. Controls that affect amplitude accuracy

1 Except to the extent that dynamic range is affected. See dynamic range.
2 Generally, frequency response is defined as half the peak-to-peak response.

A factor in measurement uncertainty not covered in the table is impedance mismatch. Analyzers do not have perfect input impedances, nor do most signal sources have ideal output impedances. However, in most cases uncertainty due to mismatch is relatively small. Improving the match of either the source or analyzer reduces uncertainty. Since an analyzer's match is worst with its input attenuator set to 0 dB, we should avoid the 0-dB setting if we can. If need be, we can attach a well-matched pad (attenuator) to the analyzer input and so effectively remove mismatch as a factor.

Absolute accuracy

The last item in table 1 is the calibrator, which gives the spectrum analyzer its absolute calibration. For convenience, calibrators are typically built into today's spectrum analyzers and provide a signal with a specified amplitude at a given frequency. We then rely on the relative accuracy of the analyzer to translate the absolute calibration to other frequencies and amplitudes.

Improving overall uncertainty

If we are looking at measurement uncertainty for the first time, we may well be concerned as we mentally add up the uncertainty figures. And even though we tell ourselves that these are worst-case values and that almost never are all factors at their worst and in the same direction at the same time, still we must add the figures directly if we are to certify the accuracy of a specific measurement.

There are some things that we can do to improve the situation. First of all, we should know the specifications for our particular spectrum analyzer. These specifications may be good enough over the range in which we are making our measurement. If not, table 1 suggests some opportunities to improve accuracy.

Before taking any data, we can step through a measurement to see if any controls can be left unchanged. We might find that a given RF attenuator setting, a given resolution bandwidth, and a given display scaling suffice for the measurement. If so, all uncertainties associated with changing these controls drop out. We may be able to trade off IF attenuation against display fidelity, using whichever is more accurate and eliminating the other as an uncertainty factor. We can even get around frequency response if we are willing to go to the trouble of characterizing our particular analyzer[1]. The same applies to the calibrator. If we have a more accurate calibrator, or one closer to the frequency of interest, we may wish to use that in lieu of the built-in calibrator.

Finally, many analyzers available today have self-calibration routines. These routines generate error coefficients (for example, amplitude changes versus resolution bandwidth) that the analyzer uses later to correct measured data. The smaller values shown in table 1, 0.5 dB for display fidelity and 0.1 dB for changes in IF attenuation, resolution bandwidth, and display scaling, are based on corrected data. As a result, these self-calibration routines allow us to make good amplitude measurements with a spectrum analyzer and give us more freedom to change controls during the course of a measurement.

1 Should we do so, then mismatch may become a more significant factor.

Sensitivity

One of the primary uses of a spectrum analyzer is to search out and measure low-level signals. The ultimate limitation in these measurements is the random noise generated by the spectrum analyzer itself. This noise, generated by the random electron motion throughout the various circuit elements, is amplified by the various gain stages in the analyzer and ultimately appears on the display as a noise signal below which we cannot make measurements. A likely starting point for noise seen on the display is the first stage of gain in the analyzer. This amplifier boosts the noise generated by its input termination plus adds some of its own. As the noise signal passes on through the system, it is typically high enough in amplitude that the noise generated in subsequent gain stages adds only a small amount to the noise power. It is true that the input attenuator and one or more mixers may be between the input connector of a spectrum analyzer and the first stage of gain, and all of these components generate noise. However, the noise that they generate is at or near the absolute minimum of –174 dBm/Hz (kTB), the same as at the input termination of the first gain stage, so they do not significantly affect the noise level input to, and amplified by, the first gain stage.

While the input attenuator, mixer, and other circuit elements between the input connector and first gain stage have little effect on the actual system noise, they do have a marked effect on the ability of an analyzer to display low-level signals because they attenuate the input signal. That is, they reduce the signal-to-noise ratio and so degrade sensitivity.

We can determine sensitivity simply by noting the noise level indicated on the display with no input signal applied. This level is the analyzer's own noise floor. Signals below this level are masked by the noise and cannot be seen or measured. However, the displayed noise floor is not the actual noise level at the input but rather the effective noise level. An analyzer display is calibrated to reflect the level of a signal at the analyzer input, so the displayed noise floor represents a fictitious (we have called it an effective) noise floor at the input below which we cannot make measurements. The actual noise level at the input is a function of the input signal. Indeed, noise is sometimes the signal of interest. Like any discrete signal, a noise signal must be above the effective (displayed) noise floor to be measured. The effective input noise floor includes the losses (attenuation) of the input attenuator, mixer(s), etc., prior to the first gain stage.

We cannot do anything about the conversion loss of the mixers, but we do have control over the RF input attenuator. By changing the value of input attenuation, we change the attenuation of the input signal and so change the displayed signal-to-noise-floor ratio, the level of the effective noise floor at the input of the analyzer, and the sensitivity. We get the best sensitivity by selecting minimum (zero) RF attenuation.

Different analyzers handle the change of input attenuation in different ways. Because the input attenuator has no effect on the actual noise generated in the system, some analyzers simply leave the displayed noise at the same position on the display regardless of the input-attenuator setting. That is, the IF gain remains constant. This being the case, the input attenuator will affect the location of a true input signal on the display. As we increase input attenuation, further attenuating the input signal, the location of the signal on the display goes down while the noise remains stationary. To maintain absolute calibration so that the actual input signal always has the same reading, the analyzer changes the indicated reference level (the value of the top line of the graticule). This design is used in older Agilent analyzers.

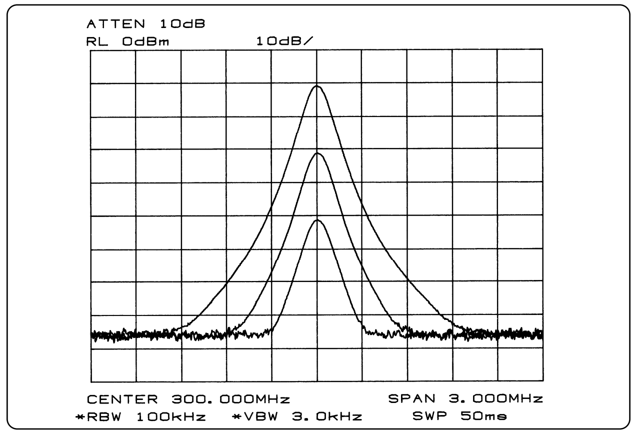


Figure 41. Some spectrum analyzers change reference level when RF attenuator is changed, so an input signal moves on the display, but the analyzer's noise does not

In newer Agilent analyzers, starting with the 8568A, an internal microprocessor changes the IF gain to offset changes in the input attenuator. Thus, true input signals remain stationary on the display as we change the input attenuator, while the displayed noise moves up and down. In this case, the reference level remains unchanged. See figures 41 and 42. In either case, we get the best signal-to-noise ratio (sensitivity) by selecting minimum input attenuation.

Resolution bandwidth also affects signal-to-noise ratio, or sensitivity. The noise generated in the analyzer is random and has a constant amplitude over a wide frequency range. Since the resolution, or IF, bandwidth filters come after the first gain stage, the total noise power that passes through the filters is determined by the width of the filters. This noise signal is detected and ultimately reaches the display. The random nature of the noise signal causes the displayed level to vary as:

$$10*\log(bw_2/bw_1),$$

where bw_1 = starting resolution bandwidth and
bw_2 = ending resolution bandwidth.

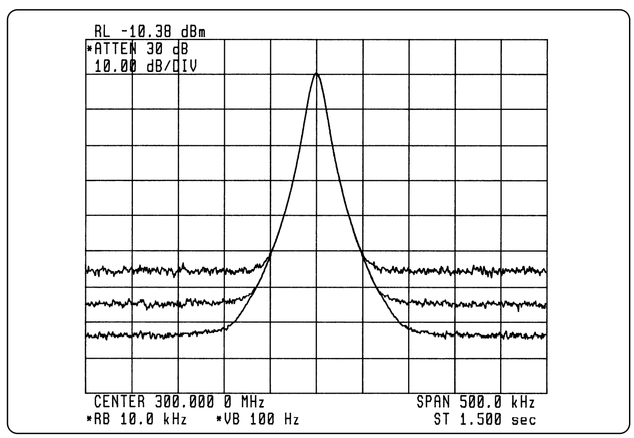


Figure 42. Other analyzers keep reference level constant by changing IF gain, so as RF attenuator is changed, the analyzer's noise moves, but an input signal does not

So if we change the resolution bandwidth by a factor of 10, the displayed noise level changes by 10 dB[1], as shown in figure 43. We get best signal-to-noise ratio, or best sensitivity, using the minimum resolution bandwidth available in our spectrum analyzer.

A spectrum analyzer displays signal plus noise, and a low signal-to-noise ratio makes the signal difficult to distinguish. We noted above that the video filter can be used to reduce the amplitude fluctuations of noisy signals while at the same time having no effect on constant signals. Figure 44 shows how the video filter can improve our ability to discern low-level signals. It should be noted that the video filter does not affect the average noise level and so does not, strictly speaking, affect the sensitivity of an analyzer.

In summary, we get best sensitivity by selecting the minimum resolution bandwidth and minimum input attenuation. These settings give us best signal-to-noise ratio. We can also select minimum video bandwidth to help us see a signal at or close to the noise level[2]. Of course, selecting narrow resolution and video bandwidths does lengthen the sweep time.

Noise figure

Many receiver manufacturers specify the performance of their receivers in terms of noise figure rather than sensitivity. As we shall see, the two can be equated. A spectrum analyzer is a receiver, and we shall examine noise figure on the basis of a sinusoidal input.

Noise figure can be defined as the degradation of signal-to-noise ratio as a signal passes through a device, a spectrum analyzer in our case. We can express noise figure as:

$$F = (S_i/N_i)/(S_o/N_o),$$

where F = noise figure as power ratio,
S_i = input signal power,
N_i = true input noise power,
S_o = output signal power, and
N_o = output noise power.

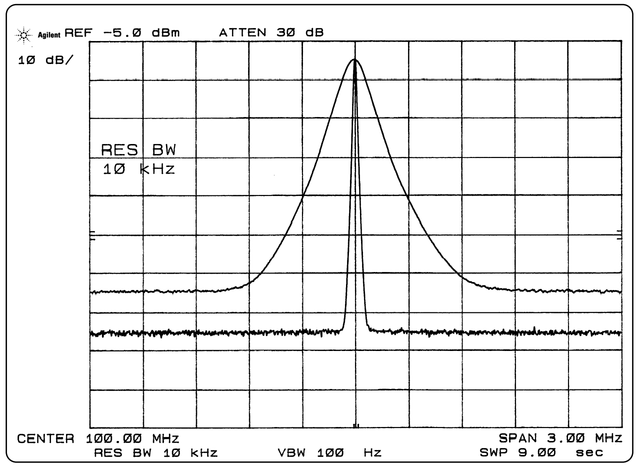


Figure 43. Displayed noise level changes as 10*log(BW_2/BW_1)

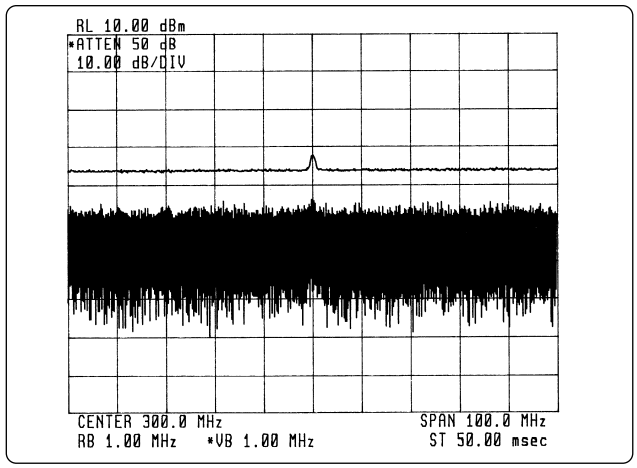


Figure 44. Video filtering makes low-level signals more discernable. (The average trace was offset for visibility.)

If we examine this expression, we can simplify it for our spectrum analyzer. First of all, the output signal is the input signal times the gain of the analyzer. Second, the gain of our analyzer is unity because the signal level at the output (indicated on the display) is the same as the level at the input (input connector). So our expression, after substitution, cancellation, and rearrangement, becomes:

$$F = N_o/N_i$$

This expression tells us that all we need to do to determine the noise figure is compare the noise level as read on the display to the true (not the effective) noise level at the input connector. Noise figure is usually expressed in terms of dB, or:

$$NF = 10*\log(F) = 10*\log(N_o) - 10*\log(N_i).$$

1 Not always true for the analyzer's own noise because of the way IF step gain and filter poles are distributed throughout the IF chain. However, the relationship does hold true when the noise is the external signal being measured.
2 For the effect of noise on accuracy, see measurement uncertainty under dynamic range.

We use the true noise level at the input rather than the effective noise level because our input signal-to-noise ratio was based on the true noise. Now we can obtain the true noise at the input simply by terminating the input in 50 ohms. The input noise level then becomes:

$N_i = kTB,$

where k = Boltzmann's constant,
T = absolute temperature in degrees Kelvin, and
B = bandwidth.

At room temperature and for a 1-Hz bandwidth,

kTB = –174 dBm.

We know that the displayed level of noise on the analyzer changes with bandwidth. So all we need to do to determine the noise figure of our spectrum analyzer is to measure the noise power in some bandwidth, calculate the noise power that we would have measured in a 1-Hz bandwidth using $10*\log(bw_2/bw_1)$, and compare that to –174 dBm. For example, if we measured –110 dBm in a 10-kHz resolution bandwidth, we would get:

$$NF = (\text{measured noise})_{dBm/RBW} - 10*\log(RBW/1) - kTB_{B=1}$$

$$= -110 \text{ dBm} - 10*\log(10{,}000/1) - (-174 \text{ dBm})$$
$$= -110 - 40 + 174$$
$$= 24 \text{ dB}.$$

Noise figure is independent of bandwidth[1]. Had we selected a different resolution bandwidth, our results would have been exactly the same. For example, had we chosen a 1-kHz resolution bandwidth, the measured noise would have been –120 dBm and 10*log(RBW/1) would have been 30. Combining all terms would have given –120 – 30 + 174 = 24 dB, the same noise figure as above.

The 24-dB noise figure in our example tells us that a sinusoidal signal must be 24 dB above kTB to be equal to the average displayed noise on this particular analyzer. Thus we can use noise figure to determine sensitivity for a given bandwidth or to compare sensitivities of different analyzers on the same bandwidth[2].

Preamplifiers

One reason for introducing noise figure is that it helps us determine how much benefit we can derive from the use of a preamplifier. A 24-dB noise figure, while good for a spectrum analyzer, is not so good for a dedicated receiver. However, by placing an appropriate preamplifier in front of the spectrum analyzer, we can obtain a system (preamplifier/spectrum analyzer) noise figure that is lower than that of the spectrum analyzer alone. To the extent that we lower the noise figure, we also improve the system sensitivity.

When we introduced noise figure above, we did so on the basis of a sinusoidal input signal. We shall examine the benefits of a preamplifier on the same basis. However, a preamplifier also amplifies noise, and this output noise can be higher than the effective input noise of the analyzer. As we shall see in the Noise as a Signal section below, a spectrum analyzer displays a random noise signal 2.5 dB below its actual value. As we explore preamplifiers, we shall account for this 2.5-dB factor where appropriate.

1 This may not be precisely true for a given analyzer because of the way resolution filter sections and gain are distributed in the IF chain.
2 The noise figure computed in this manner cannot be compared directly to that of a receiver or amplifier because the "measured noise" term in the equation understates the actual noise by 2.5 dB. See noise as a signal.

Rather than develop a lot of formulas to see what benefit we get from a preamplifier, let us look at two extreme cases and see when each might apply. First, if the noise power out of the preamplifier (in a bandwidth equal to that of the spectrum analyzer) is at least 15 dB higher than the displayed average noise level (noise floor) of the spectrum analyzer, then the noise figure of the system is approximately that of the preamplifier less 2.5 dB. How can we tell if this is the case? Simply connect the preamplifier to the analyzer and note what happens to the noise on the CRT. If it goes up 15 dB or more, we have fulfilled this requirement.

On the other hand, if the noise power out of the preamplifier (again, in the same bandwidth as that of the spectrum analyzer) is 10 dB or more lower than the average displayed noise level on the analyzer, then the noise figure of the system is that of the spectrum analyzer less the gain of the preamplifier. Again we can test by inspection. Connect the preamplifier to the analyzer; if the displayed noise does not change, we have fulfilled the requirement.

But testing by experiment means that we have the equipment at hand. We do not need to worry about numbers. We simply connect the preamplifier to the analyzer, note the average displayed noise level and subtract the gain of the preamplifier. Then we have the sensitivity of the system.

What we really want is to know ahead of time what a preamplifier will do for us. We can state the two cases above as follows:

$$\text{if} \quad NF_{PRE} + G_{PRE} \geq NF_{SA} + 15\text{ dB},$$
$$\text{then} \quad NF_{SYS} = NF_{PRE} - 2.5\text{ dB},$$

and

$$\text{if} \quad NF_{PRE} + G_{PRE} \leq NF_{SA} - 10\text{ dB},$$
$$\text{then} \quad NF_{SYS} = NF_{SA} - G_{PRE}.$$

Using these expressions, let's see how a preamplifier affects our sensitivity. Assume that our spectrum analyzer has a noise figure of 24 dB and the preamplifier has a gain of 36 dB and a noise figure of 8 dB. All we need to do is to compare the gain plus noise figure of the preamplifier to the noise figure of the spectrum analyzer. The gain plus noise figure of the preamplifier is 44 dB, more than 15 dB higher than the noise figure of the spectrum analyzer, so the noise figure of the preamplifier/spectrum-analyzer combination is that of the preamplifier less 2.5 dB, or 5.5 dB. In a 10-kHz resolution bandwidth our preamplifier/analyzer system has a sensitivity of:

$$kTB_{B=1} + 10{*}\log(RBW/1) + NF_{SYS}$$

$$= -174\text{ dBm} + 40\text{ dB} + 5.5\text{ dB}$$
$$= -128.5\text{ dBm}.$$

This is an improvement of 18.5 db over the −110 dBm noise floor without the preamplifier.

Is there any drawback to using this preamplifier? That depends upon our ultimate measurement objective. If we want the best sensitivity but no loss of measurement range, then this preamplifier is not the right choice. Figure 45 illustrates this point. A spectrum analyzer with a 24-dB noise figure will have an average displayed noise level of −110 dBm in a 10-kHz resolution bandwidth. If the 1-dB compression point[1] for that analyzer is −10 dBm, the measurement range is 100 dB. When we connect the preamplifier, we must reduce the maximum input to the system by the gain of the pre-amplifier to −46 dBm. However, when we connect the preamplifier, the noise as displayed on the CRT will rise by about 17.5 dB because the noise power out of the preamplifier is that much higher than the analyzer's own noise floor, even after accounting for the 2.5-dB factor. It is from this higher noise level that we now subtract the gain of the preamplifier. With the preamplifier in place, our measurement range is 82.5 dB, 17.5 dB less than without the preamplifier. The loss in measurement range equals the change in the displayed noise when the preamplifier is connected.

1 See mixer compression (page 41).

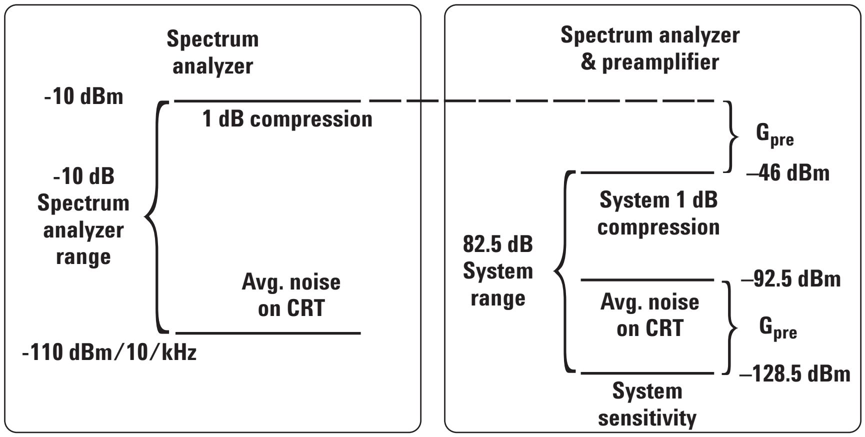


Figure 45. If the displayed noise goes up when a preamplifier is connected, measurement range is diminished by the amount the noise changes

Is there a preamplifier that will give us better sensitivity without costing us measurement range? Yes. But it must meet the second of the above criteria; that is, the sum of its gain and noise figure must be at least 10 dB less than the noise figure of the spectrum analyzer. In this case the displayed noise floor will not change noticeably when we connect the preamplifier, so although we shift the whole measurement range down by the gain of the preamplifier, we end up with the same overall range that we started with.

To choose the correct preamplifier, we must look at our measurement needs. If we want absolutely the best sensitivity and are not concerned about measurement range, we would choose a high-gain, low-noise-figure preamplifier so that our system would take on the noise figure of the preamplifier less 2.5 dB. If we want better sensitivity but cannot afford to give up any measurement range, we must choose a lower-gain preamplifier.

Interestingly enough, we can use the input attenuator of the spectrum analyzer to effectively degrade its the noise figure (or reduce the gain of the preamplifier, if you prefer). For example, if we need slightly better sensitivity but cannot afford to give up any measurement range, we can use the above preamplifier with 30 dB of RF input attenuation on the spectrum analyzer. This attenuation increases the noise figure of the analyzer from 24 to 54 dB. Now the gain plus noise figure of the preamplifier (36 + 8) is 10 dB less than the noise figure of the analyzer, and we have met the conditions of the second criterion above. The noise figure of the system is now: $NF_{sys} = NF_{SA} – G_{PRE} = 54\text{ dB} – 36\text{ dB} = 18\text{ dB}$, a 6-dB improvement over the noise figure of the analyzer alone with 0 dB of input attenuation. So we have improved sensitivity by 6 dB and given up virtually no measurement range.

Of course, there are preamplifiers that fall in between the extremes. Figure 46 enables us to determine system noise figure from a knowledge of the noise figures of the spectrum analyzer and preamplifier and the gain of the amplifier. We enter the graph of figure 46 by determining $NF_{PRE} + G_{PRE} – NF_{SA}$. If the value is less than zero, we find the corresponding point on the dashed curve and read system noise figure as the left ordinate in terms of dB above $NF_{SA} – G_{PRE}$. If $NF_{PRE} + G_{PRE} – NF_{SA}$ is a positive value, we find the corresponding point on the solid curve and read system noise figure as the right ordinate in terms of dB above NF_{PRE}.

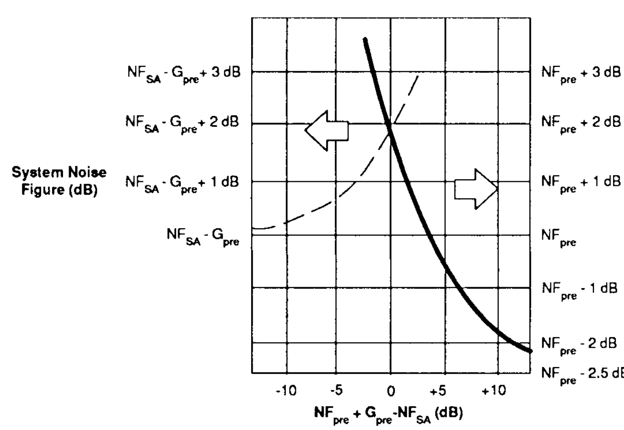


Figure 46. System noise figure for sinusoidal signals

Let's first test the two extreme cases from above. As $NF_{PRE} + G_{PRE} – NF_{SA}$ becomes less than –10 dB, we find that system noise figure asymptotically approaches $NF_{SA} – G_{PRE}$. As the value becomes greater than +15 dB, system noise figure asymptotically approaches NF_{PRE} less 2.5 dB. Next, let's try two numerical examples. Above, we determined that the noise figure of our analyzer is 24 dB. What would the system noise figure be if we add an Agilent 8447D, a preamplifier with a noise figure of about 8 dB and a gain of 26 dB? First, $NF_{PRE} + G_{PRE} – NF_{SA}$ is +10 dB. From the graph of figure 46 we find a system noise figure of about NF_{PRE} – 1.8 dB, or about 8 – 1.8 = 6.2 dB. The graph accounts for the 2.5-dB factor. On the other hand, if the gain of the preamplifier is just 10 dB, then $NF_{PRE} + G_{PRE} – NF_{SA}$ is –6 dB. This time the graph indicates a system noise figure of $NF_{SA} – G_{PRE}$ + 0.6 dB, or 24 – 10 + 0.6 = 14.6 dB.[1] (We did not introduce the 2.5-dB factor above when we determined the noise figure of the analyzer alone because we read the measured noise directly from the display. The displayed noise included the 2.5-dB factor.)

Noise as a signal

So far, we have limited our concern with noise to the noise generated within the measurement system - analyzer or analyzer/preamplifier. We described sensitivity in terms of the smallest sinusoidal signal that we could measure: one that is equal to the displayed average noise level.

However, random noise is sometimes the signal that we want to measure. Because of the nature of noise, the superheterodyne spectrum analyzer indicates a value that is lower than the actual value of the noise. Let's see why this is so and how we can correct for it.

By random noise, we mean a signal whose instantaneous amplitude has a Gaussian distribution versus time, as shown in figure 47. For example, thermal or Johnson noise has this characteristic. Such a signal has no discrete spectral components, so we cannot select some particular component and measure it to get an indication of signal strength. In fact, we must define what we mean by signal strength. If we sample the signal at an arbitrary instant, we could theoretically get any amplitude value. We need some measure that expresses the noise level averaged over time. Power and rms voltage both satisfy that requirement.

We have already seen that both video filtering and video averaging reduce the peak-to-peak fluctuations of a signal and can give us a steady value. We must equate this value to either power or rms voltage. The rms value of a Gaussian distribution equals its standard deviation, σ.

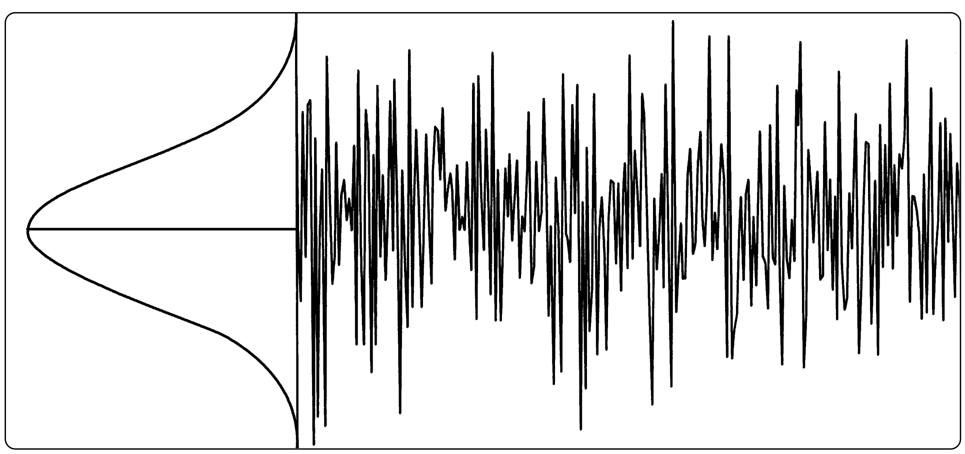


Figure 47. Random noise has a Gaussian amplitude distribution

1 For more details on noise figure, see Application Note 57-1, "Fundamentals of RF and Microwave Noise Figure Measurements," (literature number 5952-8255E).

Let's start with our analyzer in the linear display mode. The Gaussian noise at the input is band limited as it passes through the IF chain, and its envelope takes on a Rayleigh distribution (figure 48). The noise that we see on our analyzer display, the output of the envelope detector, is the Rayleigh-distributed envelope of the input noise signal. To get a steady value, the mean value, we use video filtering or averaging. The mean value of a Rayleigh distribution is 1.253σ.

But our analyzer is a peak-responding voltmeter calibrated to indicate the rms value of a sine wave. To convert from peak to rms, our analyzer scales its readout by 0.707 (–3 dB). The mean value of the Rayleigh-distributed noise is scaled by the same factor, giving us a reading that is 0.886σ (1.05 dB below σ). To equate the mean value displayed by the analyzer to the rms voltage of the input noise signal, then, we must account for the error in the displayed value. Note, however, that the error is not an ambiguity; it is a constant error that we can correct for by adding 1.05 dB to the displayed value.

Normally, we use our analyzer in the log display mode, and this mode adds to the error in our noise measurement. The gain of a log amplifier is a function of signal amplitude, so the higher noise values are not amplified as much as the lower values. As a result, the output of the envelope detector is a skewed Rayleigh distribution, and the mean value that we get from video filtering or averaging is another 1.45 dB lower. In the log mode, then, the mean or average noise is displayed 2.5 dB too low. Again, this error is not an ambiguity, and we can correct for it.

This is the 2.5-dB factor that we accounted for in the preamplifier discussion above whenever the noise power out of the preamplifier was approximately equal to or greater than the analyzer's own noise.

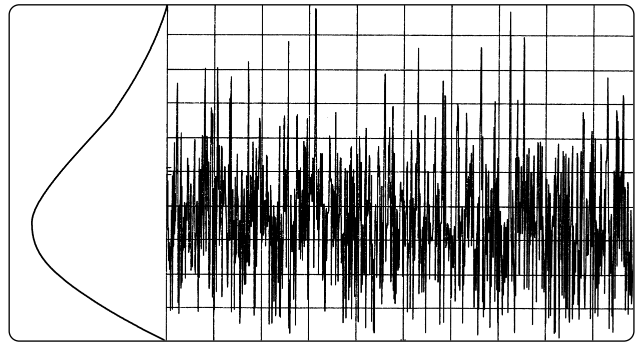


Figure 48. The envelope of band-limited Gaussian noise has a Rayleigh distribution

Another factor that affects noise measurements is the bandwidth in which the measurement is made. We have seen how changing resolution bandwidth affects the displayed level of the analyzer's internally generated noise. Bandwidth affects external noise signals in the same way. To compare measurements made on different analyzers, then, we must know the bandwidths used in each case.

Not only does the 3-dB (or 6-dB) bandwidth of the analyzer affect the measured noise level, the shape of the resolution filter also plays a role. To make comparisons possible, we define a standard noise-power bandwidth: the width of a rectangular filter that passes the same noise power as our analyzer's filter. For the near-Gaussian filters in Agilent analyzers, the equivalent noise-power bandwidth is about 1.05 to 1.13 times the 3-dB bandwidth, depending on bandwidth selectivity. For example, a 10-kHz resolution bandwidth filter has a noise-power bandwidth in the range of 10.5 to 11.3 kHz.

If we use $10*\log(bw_2/bw_1)$ to adjust the displayed noise level to what we would have measured in a noise power bandwidth of the same numeric value as our 3-dB bandwidth, we find that the adjustment varies from:

$$10*\log(10{,}000/10{,}500) = -0.21 \text{ dB to}$$
$$10*\log(10{,}000/11{,}300) = -0.53 \text{ dB.}$$

In other words, if we subtract something between 0.21 and 0.53 dB from the indicated noise level, we shall have the noise level in a noise-power bandwidth that is convenient for computations.

Let's consider all three factors and calculate a total correction:

Rayleigh distribution (linear mode):	1.05 dB
Log amplifier (log mode):	1.45 dB
3-dB/noise power bandwidths:	–0.5 dB

Total correction:	2.0 dB

Here we use –0.5 dB as a reasonable compromise for the bandwidth correction. The total correction is thus a convenient value.

Many of today's microprocessor-controlled analyzers allow us to activate a noise marker. When we do so, the microprocessor switches the analyzer into the sample display mode, computes the mean value of the 32 display points about the marker, adds the above 2-dB amplitude correction, normalizes the value to a 1-Hz noise-power bandwidth, and displays the normalized value.[1]

The analyzer does the hard part. It is reasonably easy to convert the noise-marker value to other bandwidths. For example, if we want to know the total noise in a 4-MHz communication channel, we add 66 dB to the noise-marker value (60 dB for the 1,000,000/1 and another 6 dB for the additional factor of four).

Preamplifier for noise measurements

Since noise signals are typically low-level signals, we often need a preamplifier to have sufficient sensitivity to measure them. However, we must recalculate sensitivity of our analyzer first. Above, we defined sensitivity as the level of a sinusoidal signal that is equal to the displayed average noise floor. Since the analyzer is calibrated to show the proper amplitude of a sinusoid, no correction for the signal was needed. But noise is displayed 2.5 dB too low, so an input noise signal must be 2.5 dB above the analyzer's displayed noise floor to be at the same level by the time it reaches the display. The input and internal noise signals add to raise the displayed noise by 3 dB, a factor of two in power. So we can define the noise figure of our analyzer for a noise signal as:

$$NF_{SA(N)} = (\text{noise floor})_{dBm/RBW} - 10*\log(RBW/1) - kTB_{B=1} + 2.5 \text{ dB.}$$

If we use the same noise floor as above, –110 dBm in a 10-kHz resolution bandwidth, we get:

$$NF_{SA(N)} = -110 \text{ dBm} - 10*\log(10{,}000/1) - (174 \text{ dBm}) + 2.5 \text{ dB} = 26.5 \text{ dB.}$$

As was the case for a sinusoidal signal, $NFS_{SA(N)}$ is independent of resolution bandwidth and tells us how far above kTB a noise signal must be to be equal to the noise floor of our analyzer.

When we add a preamplifier to our analyzer, the system noise figure and sensitivity improve. However, we have accounted for the 2.5-dB factor in our definition of NF SAMI so the graph of system noise figure becomes that of figure 49. We determine system noise figure for noise the same way that we did for a sinusoidal signal above.

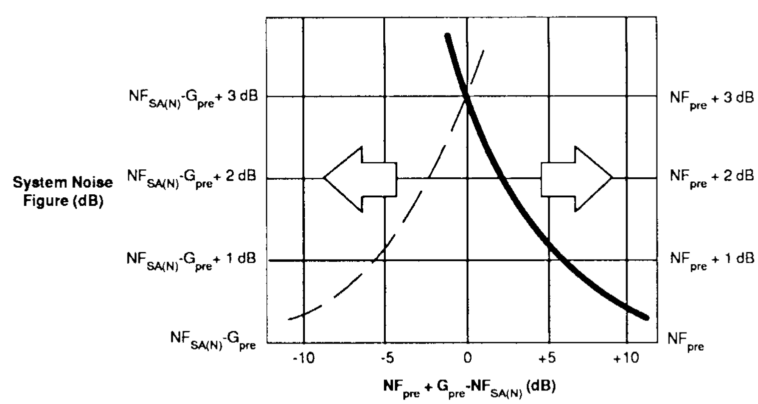


Figure 49. System noise figure for noise signals

1 The ESA-E family computes the mean over half a division, regardless of the number of display points.

Dynamic range

Definition

Dynamic range is generally thought of as the ability of an analyzer to measure harmonically related signals and the interaction of two or more signals; for example, to measure second- or third-harmonic distortion or third-order intermodulation. In dealing with such measurements, remember that the input mixer of a spectrum analyzer is a non-linear device and so always generates distortion of its own. The mixer is non-linear for a reason. It must be nonlinear to translate an input signal to the desired IF. But the unwanted distortion products generated in the mixer fall at the same frequencies as do the distortion products we wish to measure on the input signal.

So we might define dynamic range in this way: it is the ratio, expressed in dB, of the largest to the smallest signals simultaneously present at the input of the spectrum analyzer that allows measurement of the smaller signal to a given degree of uncertainty.

Notice that accuracy of the measurement is part of the definition. We shall see how both internally generated noise and distortion affect accuracy below.

Dynamic range versus internal distortion

To determine dynamic range versus distortion, we must first determine just how our input mixer behaves. Most analyzers, particularly those utilizing harmonic mixing to extend their tuning range[1], use diode mixers. (Other types of mixers would behave similarly.) The current through an ideal diode can be expressed as:

$$i = I_s(e^{qv/kT} - 1),$$

where q = electronic charge,
v = instantaneous voltage,
k = Boltzmann's constant, and
T = temperature in degrees Kelvin.

We can expand this expression into a power series:

$$i = I_s(k_1 v + k_2 v^2 + k_3 v^3 + \ldots),$$

where $k_1 = q/kT$
$k_2 = k_1^2/2!$,
$k_3 = k_1^3!$, etc.

Let's now apply two signals to the mixer. One will be the input signal that we wish to analyze; the other, the local oscillator signal necessary to create the IF:

$$v = V_{LO}\sin(w_{LO}t) + V_1\sin(w_1 t).$$

If we go through the mathematics, we arrive at the desired mixing product that, with the correct LO frequency, equals the IF:

$$k_2 V_{LO} V_1 \cos[(w_{LO} - w_1)t].$$

A $k_2 V_{LO} V_1 \cos[(w_{LO} + w_1)t]$ term is also generated, but in our discussion of the tuning equation, we found that we want the LO to be above the IF, so $(w_{LO} + w_1)$ is also always above the IF.

With a constant LO level, the mixer output is linearly related to the input signal level. For all practical purposes, this is true as long as the input signal is more than 15 to 20 dB below the level of the LO. There are also terms involving harmonics of the input signal:

$$(3k_3/4)V_{LO}V_1^2\sin(w_{LO} - 2w_1)t,$$
$$(k_4/8)V_{LO}V_1^3\sin(w_{LO} - 3w_1)t, \text{ etc.}$$

1 See Chapter 3.

These terms tell us that dynamic range due to internal distortion is a function of the input signal level at the input mixer. Let's see how this works, using as our definition of dynamic range the difference in dB between the fundamental tone and the internally generated distortion.

The argument of the sine in the first term includes $2w_1$, so it represents the second harmonic of the input signal. The level of this second harmonic is a function of the square of the voltage of the fundamental, V_1^2. This fact tells us that for every dB that we drop the level of the fundamental at the input mixer, the internally generated second harmonic drops by 2 dB. See figure 50. The second term includes $3w_1$, the third harmonic, and the cube of the input-signal voltage, V_1^3. So a 1-dB change in the fundamental at the input mixer changes the internally generated third harmonic by 3 dB.

Distortion is often described by its order. The order can be determined by noting the coefficient associated with the signal frequency or the exponent associated with the signal amplitude. Thus second-harmonic distortion is second order and third harmonic distortion is third order. The order also indicates the change in internally generated distortion relative to the change in the fundamental tone that created it.

Now let us add a second input signal:

$$v = V_{LO}\sin(w_{LO}t) + V_1\sin(w_1t) + V_2\sin(w_2t).$$

This time when we go through the math to find internally generated distortion, in addition to harmonic distortion, we get:

$$(k_4/8)V_{LO}V_1^2V_2\cos[w_{LO} - (2w_1 - w_2)]t,$$

$$(k_4/8)V_{LO}V_1V_2^2\cos[w_{LO} - (2w_2 - w_1)]t, \text{ etc.}$$

These represent intermodulation distortion, the interaction of the two input signals with each other. The lower distortion product, $2w_1 - w_2$, falls below w_1 by a frequency equal to the difference between the two fundamental tones, $w_2 - w_1$. The higher distortion product, $2w_2 - w_1$, falls above w_2 by the same frequency. See figure 50.

Once again, dynamic range is a function of the level at the input mixer. The internally generated distortion changes as the product of V_1^2 and V_2 in the first case, of V_1 and V_2^2 in the second. If V_1 and V_2 have the same amplitude, the usual case when testing for distortion, we can treat their products as cubed terms (V_1^3 or V_2^3). Thus, for every dB that we simultaneously change the level of the two input signals, there is a 3-dB change in the distortion components as shown in figure 50.

This is the same degree of change that we saw for third harmonic distortion above. And in fact, this, too, is third-order distortion. In this case, we can determine the degree of distortion by summing the coefficients of w_1 and w_2 (e.g., $2w_1 - 1w_2$ yields $2 + 1 = 3$) or the exponents of V_1 and V_2.

All this says that dynamic range depends upon the signal level at the mixer. How do we know what level we need at the mixer for a particular measurement? Many analyzer data sheets now include graphs to tell us how dynamic range varies. However, if no graph is provided, we can draw our own.

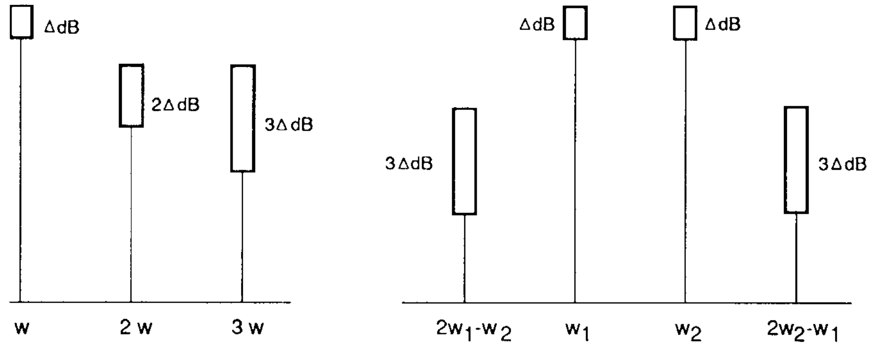


Figure 50. Changing the level of fundamental tone (w) or tones (w_1, w_2) at the mixer affects internally generated distortion

We do need a starting point, and this we must get from the data sheet. We shall look at second-order distortion first. Let's assume the data sheet says that second-harmonic distortion is 70 dB down for a signal –40 dBm at the mixer. Because distortion is a relative measurement, and, at least for the moment, we are calling our dynamic range the difference in dB between fundamental tone or tones and the internally generated distortion, we have our starting point. Internally generated second-order distortion is 70 dB down, so we can measure distortion down 70 dB. We plot that point on a graph whose axes are labeled distortion (dBc) versus level at the mixer (level at the input connector minus the input-attenuator setting). See figure 51. What happens if the level at the mixer drops to –50 dBm? As noted in figure 50, for every dB change in the level of the fundamental at the mixer there is a 2-dB change in the internally generated second harmonic. But for measurement purposes, we are only interested in the relative change, that is, in what happened to our measurement range. In this case, for every dB that the fundamental changes at the mixer, our measurement range also changes by 1 dB. In our second-harmonic example, then, when the level at the mixer changes from –40 to –50 dBm, the internal distortion, and thus our measurement range, changes from –70 to –80 dBc. In fact, these points fall on a line with a slope of 1 that describes the dynamic range for any input level at the mixer.

We can construct a similar line for third-order distortion. For example, a data sheet might say third-order distortion is –70 dBc for a level of –30 dBm at this mixer. Again, this is our starting point, and we would plot the point shown in figure 51. If we now drop the level at the mixer to –40 dBm, what happens? Referring again to figure 50, we see that both third-harmonic distortion and third-order intermodulation distortion fall by 3 dB for every dB that the fundamental tone or tones fall. Again it is the difference that is important. If the level at the mixer changes from –30 to –40 dBm, the difference between fundamental tone or tones and internally generated distortion changes by 20 dB. So the internal distortion is –90 dBc. These two points fall on a line having a slope of 2, giving us the third-order performance for any level at the mixer.

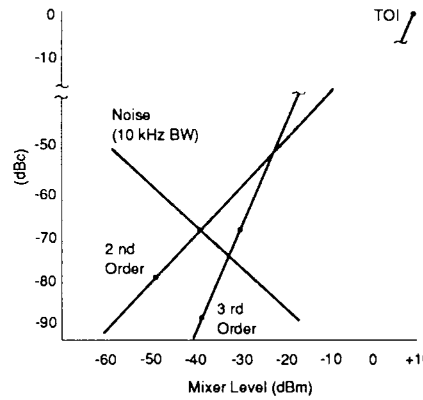


Figure 51. Dynamic range versus distortion and noise

Sometimes third-order performance is given as TOI (Third Order Intercept). This is the mixer level at which the internally generated third-order distortion would be equal to the fundamental(s), or 0 dBc. This situation cannot be realized in practice because the mixer would be well into saturation. However, from a mathematical standpoint TOI is a perfectly good data point because we know the slope of the line. So even with TOI as a starting point, we can still determine the degree of internally generated distortion at a given mixer level.

We can calculate TOI from data-sheet information. Because third-order dynamic range changes 2 dB for every dB change in the level of the fundamental tone(s) at the mixer, we get TOI by subtracting half of the specified dynamic range in dBc from the level of the fundamental(s):

$$\text{TOI} = 1_{\text{fund}} - d/2,$$

where 1_{fund} = level of the fundamental in dBm and
d = difference in dBc between fundamental and distortion.

Using the values from the discussion above:

$$\text{TOI} = -30\text{ dBm} - (-70\text{ dBc})/2 = +5\text{ dBm}.$$

Attenuator test

Understanding the distortion graph is important, but we can use a simple test to determine whether displayed distortion components are true input signals or internally generated signals. Change the input attenuator. If the displayed value of the distortion components remains the same, the components are part of the input signal. If the displayed value changes, the distortion components are generated internally or are the sum of external and internally generated signals. We continue changing the attenuator until the displayed distortion does not change and then complete the measurement.

Noise

There is another constraint on dynamic range, and that is the noise floor of our spectrum analyzer. Going back to our definition of dynamic range as the ratio of the largest to the smallest signal that we can measure, the average noise of our spectrum analyzer puts the limit on the smaller signal. So dynamic range versus noise becomes signal-to-noise ratio in which the signal is the fundamental whose distortion we wish to measure.

So how do we plot noise on our dynamic range chart? Using the same numbers for sensitivity that we used earlier (24-dB noise figure), we would calculate an average noise level of –110 dBm in a 10-kHz resolution bandwidth. If our signal fundamental has a level of -40 dBm at the mixer, it is 70 dB above the average noise, so we have 70 dB signal-to-noise ratio. For every dB that we reduce the signal level at the mixer, we lose 1 dB of signal-to-noise ratio. Our noise curve is a straight line having a slope of –1, as shown in figure 51.

Under what conditions, then, do we get the best dynamic range? Without regard to measurement accuracy, it would be at the intersection of the appropriate distortion curve and the noise curve. Figure 51 tells us that our maximum dynamic range for second-order distortion is 70 dB; for third-order distortion, 77 dB.

Figure 51 shows the dynamic range for one resolution bandwidth. We certainly can improve dynamic range by narrowing the resolution bandwidth, but there is not a one-to-one correspondence between the lowered noise floor and the improvement in dynamic range. For second-order distortion the improvement is one half the change in the noise floor; for third-order distortion, two-thirds the change in the noise floor. See figure 52.

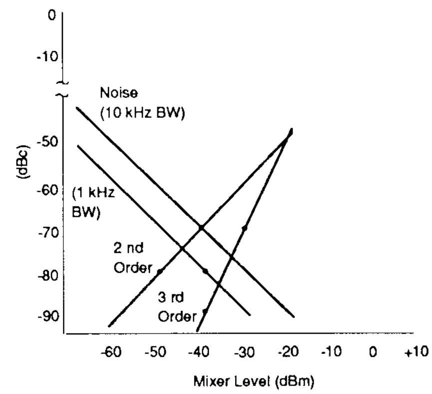


Figure 52. Reducing resolution bandwidth improves dynamic range

The final factor in dynamic range is the phase noise on our spectrum analyzer LO, and this affects only third-order distortion measurements. For example, suppose we are making a two-tone, third-order distortion measurement on an amplifier, and our test tones are separated by 10 kHz. The third-order distortion components will be separated from the test tones by 10 kHz also. For this measurement we might find ourselves using a 1-kHz resolution bandwidth. Referring to figure 52 and allowing for a 10-dB decrease in the noise curve, we would find a maximum dynamic range of about 84 dB. However, what happens if our phase noise at a 10-kHz offset is only –75 dBc? Then 75 dB becomes the ultimate limit of dynamic range for this measurement, as shown in figure 53.

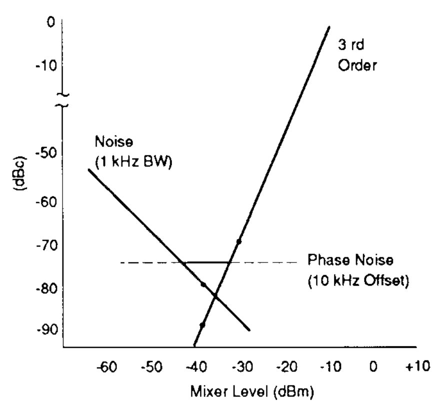


Figure 53. Phase noise can limit third-order intermodulation tests

In summary, the dynamic range of a spectrum analyzer is limited by three factors: the distortion performance of the input mixer, the broadband noise floor (sensitivity) of the system, and the phase noise of the local oscillator.

Dynamic range versus measurement uncertainty

In our previous discussion of amplitude accuracy, we included only those items listed in table 1 plus mismatch. We did not cover the possibility of an internally generated distortion product (a sinusoid) being at the same frequency as an external signal that we wished to measure. However, internally generated distortion components fall at exactly the same frequencies as the distortion components we wish to measure on external signals. The problem is that we have no way of knowing the phase relationship between the external and internal signals. So we can only determine a potential range of uncertainty:

Uncertainty (in dB) = $20*\log(1 \pm 10^{d/20})$,

where d = difference in dB between larger and smaller sinusoid (a negative number).

See figure 54. For example, if we set up conditions such that the internally generated distortion is equal in amplitude to the distortion on the incoming signal, the error in the measurement could range from +6 dB (the two signals exactly in phase) to -infinity (the two signals exactly out of phase and so canceling). Such uncertainty is unacceptable in most cases. If we put a limit of 1 dB on the measurement uncertainty, figure 54 shows us that the internally generated distortion product must be about 18 dB below the distortion product that we wish to measure. To draw dynamic-range curves for second- and third-order measurements with no more than 1 dB of measurement error, we must then offset the curves of figure 51 by 18 dB as shown in figure 55.

Next let's look at uncertainty due to low signal-to-noise ratio. The distortion components we wish to measure are, we hope, low-level signals, and often they are at or very close to the noise level of our spectrum analyzer. In such cases we often use the video filter to make these low-level signals more discernable. Figure 56 shows the error in displayed signal level as a function of displayed signal-to-noise for a typical spectrum analyzer. Note that the error is only in one direction, so we could correct for it. However, we usually do not. So for our dynamic-range measurement, let's accept a 0.5-dB error due to noise and offset the noise curve in our dynamic-range chart by 5 dB as shown in figure 55. Where the distortion and noise curves intersect, the maximum error possible would be less than 1.5 dB.

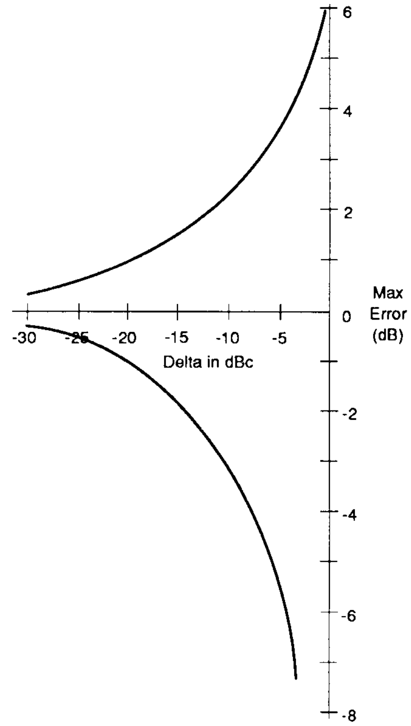


Figure 54. Uncertainty versus difference in amplitude between two sinusoids at the same frequency

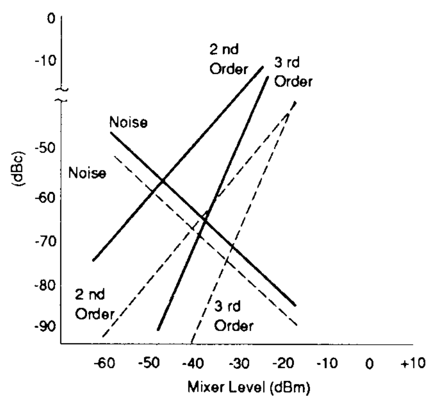


Figure 55. Dynamic range for 1.5-dB maximum error

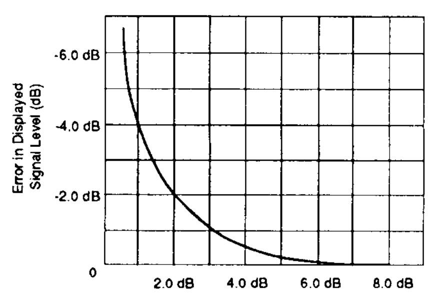


Figure 56. Error in displayed signal amplitude due to noise

Let's see what happened to our dynamic range as a result of our concern with measurement error. As figure 55 shows, second-order-distortion dynamic range changes from 70 to 58.5 dB, a change of 11.5 dB. This is one half the total offsets for the two curves (18 dB for distortion; 5 dB for noise). Third-order distortion changes from 77 dB to about 68 dB for a change of about 9 dB. In this case the change is one third of the 18-dB offset for the distortion curve plus two thirds of the 5-dB offset for the noise curve.

Mixer compression

In our discussion of dynamic range, we did not concern ourselves with how accurately the larger tone is displayed, even on a relative basis. As we raise the level of a sinusoidal input signal, eventually the level at the input mixer becomes so high that the desired output mixing product no longer changes linearly with respect to the input signal. The mixer is in saturation, and the displayed signal amplitude is too low.

Saturation is gradual rather than sudden. To help us stay away from the saturation condition, the 0.5-dB or 1-dB compression point is usually specified. A mixer level of –10 to –5 dBm is typical. Thus we can determine what input attenuator setting to use for accurate measurement of high-level signals[1].

Actually, there are three different methods of evaluating compression. The traditional method, called CW compression, measures the change in gain of a device (amplifier or mixer or system) as the input signal power is swept upward. This method is the one just described. Note that the CW compression point is considerably higher than the levels for the fundamentals indicated above for even moderate dynamic range. So we were correct in not concerning ourselves with the possibility of compression of the larger signal(s).

A second method, called two-tone compression, measures the change in system gain for a small signal while the power of a larger signal is swept upward. Two-tone compression applies to the measurement of multiple CW signals, such as sidebands and independent signals. The threshold of compression of this method is usually a few dB lower than that of the CW method.

A final method, called pulse compression, measures the change in system gain to a narrow (broadband) RF pulse while the power of the pulse is swept upward. When measuring pulses, we often use a resolution bandwidth much narrower than the bandwidth of the pulse, so our analyzer displays the signal level well below the peak pulse power. As a result, we could be unaware of the fact that the total signal power is above the mixer compression threshold. A high threshold improves signal-to-noise ratio for high-power, ultra-narrow or widely chirped pulses. The threshold is about 12 dB higher than for two-tone compression in the Agilent 8560A, 8561A/B, and 8562A/B/C analyzers. Nevertheless, because different compression mechanisms affect CW, two-tone, and pulse compression differently, any of the compression thresholds can be lower than any other.

1 Many analyzers internally control the combined settings of the input attenuator and IF gain so that a cw as high as the compression level at the input mixer creates a deflection above the top line of the graticule. Thus we cannot make incorrect measure-ments inadvertently.

Display range and measurement range

There are two additional ranges that are often confused with dynamic range: display range and measurement range. Display range, often called display dynamic range, refers to the calibrated amplitude range of the CRT display. For example, a display with eight divisions would seem to have an 80-dB display range when we select 10 dB per division. However, in most cases, as with Agilent, the bottom division is not calibrated. The bottom line of the graticule represents a signal amplitude of zero, so the bottom division of the display covers the range from –70 dB to infinity relative to the reference level (top line)[1]. The bottom division of ten-division displays are also typically uncalibrated. Another factor is the range of the log amplifier. Typical ranges are 70 and 90 dB for analyzers with eight and ten divisions, respectively. Some analyzers do have log amplifiers, or use autoranging, to utilize the full 10 divisions of their displays. The Agilent ESA-E and 8560 families use a combination of digital signal processing and autoranging for a full 100-dB display when we select one of the digitally-implemented resolution bandwidths.

The question is, can the full display range be used? From the discussion of dynamic range above, we know that the answer is generally yes. In fact, dynamic range often exceeds display range or log amplifier range. What then? To bring the smaller signals into the calibrated area of the display, we must increase IF gain. But in so doing, we move the larger signals off the top of the display, above the reference level. In Agilent analyzers, we can move signals at least 20 dB above the reference level without affecting the accuracy with which the smaller signals are displayed. So we can indeed take advantage of the full dynamic range of an analyzer even when the dynamic range exceeds the display range.

Measurement range is the ratio of the largest to the smallest signal that can be measured under any circumstances. The upper limit is determined by the maximum safe input level, +30 dBm (1 watt) for most analyzers. These analyzers have input attenuators settable to 60 or 70 dB, so we can reduce +30-dBm signals to levels well below the compression point of the input mixer and measure them accurately. Sensitivity sets the other end of the range. Depending on the minimum resolution bandwidth of the particular analyzer, sensitivity typically ranges from –115 to –135 dBm. Measurement range, then, can vary from 145 to 160 dB. Of course, we cannot view a –135 dBm signal while a +30 dBm signal is also present at the input.

Frequency measurements

So far, we have focused almost exclusively on amplitude measurements. What about frequency measurements? Up until the late 1970s, absolute frequency uncertainty was measured in megahertz because the first LO was a high-frequency oscillator operating above the RF range of the analyzer, and there was no attempt to tie the LO to a more accurate reference oscillator. (An exception was the Agilent 8580, an automatic spectrum analyzer based on the Agilent 8555A, in which an external synthesizer was substituted for the internal LO. However, the cost of the system prevented its use as a general -purpose analyzer.) Many analyzers of this type are still available and in general use. Examples are the Agilent 8590A and 8592A.

Absolute frequency uncertainty of even many megahertz is not a hindrance in many cases. For example, many times we are measuring an isolated signal. Or we need just enough accuracy to be able to identify the signal of interest among other signals. Absolute frequency is often listed under the Frequency Readout Accuracy specification and refers to center frequency and, for analyzers with microprocessors and digital displays, start, stop, and marker frequencies.

More important, usually, is relative frequency uncertainty. How far apart are spectral components? What is the modulation frequency? Here the span accuracy comes into play. For Agilent analyzers, span accuracy generally means the uncertainty in the indicated separation of any two spectral components on the display. For example, suppose span accuracy is 3 percent and we have two signals separated by two divisions on a 1-MHz span (100 kHz per division). The uncertainty of the signal separation would be 6 kHz. The uncertainty would be the same if we used delta markers and the delta reading was 200 kHz.

Span accuracy can be used to improve low-frequency accuracy. How would we tune to a 100-kHz signal on an analyzer having 5 MHz frequency uncertainty? We can use the LO feed through (the response created when the first LO sweeps past the first IF) as a zero-frequency marker and the span accuracy to locate the signal. The LO feed through indicates 0 Hz with no error, and we can place it at the left side of the display graticule with a span of 200 kHz. Again assuming 3 percent span accuracy, our signal should appear at the center of the display ±0.15 divisions.

1 Because of the internally generated noise, analyzers always display some signal above the bottom line of the graticule on 10 dB/div and higher scale factors.

With the introduction of the Agilent 8568A in 1978, counter-like frequency accuracy became available in a general-purpose spectrum analyzer. A low-drift, ovenized crystal oscillator was added as a reference for all of the LOs in the instrument. Over the years, crystal reference oscillators, some ovenized, some not, have been added to analyzers in all cost ranges.

A comment on stabilized oscillators: if we use the broadest definition of indirect synthesis, that the frequency of the oscillator in question is in some way determined by a reference oscillator, then the actual technique used is irrelevant. Phase lock, frequency discrimination, counter lock all fall within this definition of indirect synthesis.

What we really care about is the effect on frequency accuracy (and drift). A typical readout accuracy might be stated as follows:

±[(freq readout x freq ref error) + A% of span + B% of RBW + C Hz].

Note that we cannot determine an exact frequency error unless we know something about the frequency reference. In some cases we are given an annual aging rate (for example, ±2 x 10^{-6}/year); in others, aging over a shorter period (for example, ±5 x 10^{-10}/day). In addition, we need to know when the oscillator was last adjusted and how close it was set to its nominal frequency (usually 10 MHz). Other factors that we often overlook when we think about frequency accuracy are whether or not the instrument was unplugged from the power line before we use it (some reference oscillators require 72 hours to reach their specified drift rate) and the temperature coefficient (it can be worse than the drift rate). In short, there are a number of factors to consider before we can determine frequency uncertainty.

In a factory setting there is often an in-house frequency standard available that is traceable to a national standard. Most analyzers with internal reference oscillators allow substitution of an external reference. The frequency reference error in the above expression then becomes that of the in-house standard.

When making measurements in the field, we typically want to turn our analyzer on, complete our task, and move on as quickly as possible. It is helpful to know how the reference in our analyzer behaves under short warm up conditions. For the Agilent 8560 series of portable spectrum analyzers, specifications for the standard reference give performance after a five-minute warm up; specifications for the precision frequency reference give performance for both five- and fifteen-minute warm up.

Most analyzers with digital displays include markers. When a single marker is activated, it gives us absolute frequency (as well as amplitude). However, the indicated frequency of the marker is a function of the frequency calibration of the display and the location of the marker on the display. To get best frequency accuracy, then, we must be careful to place the marker exactly at the peak of the response to a spectral component. If we place the marker at some other point on the response, we shall get a different frequency reading. For the best accuracy, we may narrow the span and resolution bandwidth to minimize their effects and to make it easier to place the marker at the peak of the response.

Many analyzers that have markers include internal counter schemes that eliminate the effects of span and resolution bandwidth on frequency accuracy. The counter does not count the input signal directly but instead counts the IF signal and perhaps one or more of the LOs, and the microprocessor computes the frequency of the input signal. A minimum signal-to-noise ratio is required to eliminate noise as a factor in the count. But counting the signal in the IF also eliminates the need to place the marker at the exact peak of the signal response on the display. Anywhere sufficiently out of the noise will do. Marker count accuracy might be stated as:

±[(marker freq x freq ref error) + counter resolution + A Hz].

We must still deal with the frequency reference error as above. Counter resolution refers to the least significant digit in the counter readout, a factor here just as with any digital counter. Some analyzers allow the counter mode to be used with delta markers. In that case, the effects of counter resolution and the fixed frequency would be doubled.

Summary

In this chapter we have described the RF superheterodyne spectrum analyzer. We went through the block diagram and noted bow the various sections affect our ability to make measurements. We looked at amplitude accuracy, sensitivity, and dynamic range, and ended with a discussion of frequency measurements. In the next chapter we shall see bow we might extend the frequency range to enable us to analyze microwave signals.

Chapter 3
Extending the frequency range

Harmonic mixing

In chapter 2, we described a single-range spectrum analyzer that tunes to 2.9 GHz. Now we wish to tune higher in frequency, perhaps to 22 GHz. The most economical way to achieve such an extended range is to use harmonic mixing.

But let us take one step at a time. In developing our tuning equation in chapter 2, we found that we needed the low-pass filter of figure 7 to prevent higher-frequency signals from reaching the mixer. The result was a uniquely-responding, single-band analyzer that tuned to 2.9 GHz. Now we wish to observe and measure higher-frequency signals, so we must remove the low-pass filter.

Another factor that we explored in developing the tuning equation was the choice of LO and intermediate frequencies. We decided that the IF should not be within the band of interest because it created a hole in our tuning range in which we could not make measurements. So we chose 3.6 GHz, moving the IF above the highest tuning range of interest (2.9 GHz). Since our new tuning range will be above 2.9 GHz, it seems logical to move the new IF to a frequency below 2.9 GHz. A typical first IF for these higher-frequency ranges in Agilent spectrum analyzers is 321.4 MHz. We shall use this frequency in our examples. In summary, for the low band, up to 2.9 GHz, our first IF is 3.6 GHz. For the upper frequency bands, we must switch to a first IF of 321.4 MHz. Note that in figure 10 the second IF is 321.4 MHz, so all we need to do when we wish to tune to the higher ranges is bypass the first IF, as shown in figure 57.

In chapter 2 we used a mathematical approach to conclude that we needed a low-pass filter. As we shall see, things become more complex in the situation here, so we shall use a graphical approach as an easier method to see what is happening. The low band is the simpler case, so we shall start with that. In all of our graphs, we shall plot the LO frequency along the horizontal axis and signal frequency along the vertical axis, as shown in figure 58. Since we know that we get a mixing product equal to the IF (and therefore a response on the display) whenever the input signal differs from the LO by the IF, we can determine the frequency to which the analyzer is tuned simply by adding the IF to or subtracting it from the LO frequency. To determine our tuning range, then, we start by plotting the LO frequency against the signal-frequency axis as shown by the dashed line in figure 58. Subtracting the IF from the dashed line gives us a tuning range of 0 to 2.9 GHz, the range that we developed in chapter 2. Note that this line in figure 58 is labeled "1$^-$" to indicate fundamental mixing and the use of the minus sign in the tuning equation. We can use the graph to determine what LO frequency is required to receive a particular signal (to display a 1-GHz signal, the LO must be tuned to 4.6 GHz) or to what signal the analyzer is tuned for a given LO frequency (for an LO frequency of 6 GHz, the spectrum analyzer is tuned to receive a signal frequency of 2.4 GHz). In our text, we shall round off the first IF to one decimal place; the true IF is shown on the block diagram.

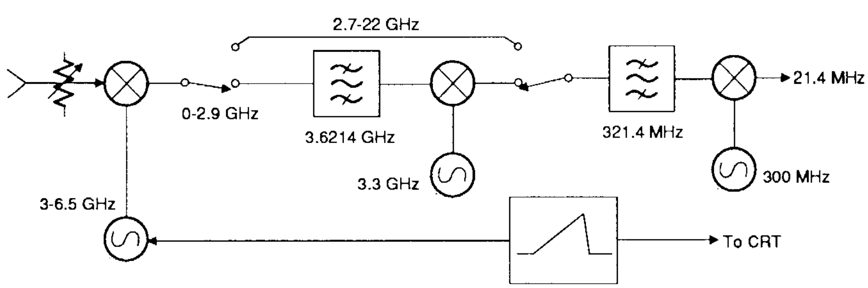


Figure 57. Switching arrangement to provide a high IF for the low band and a low IF for the high bands

Now let's add the other fundamental-mixing band by adding the IF to the LO line in figure 58. This gives us the solid upper line, labeled 1^+, that indicates a tuning range from 7.2 to 10.1 GHz. Note that for a given LO frequency, the two frequencies to which the analyzer is tuned are separated by twice the IF. It would seem valid to argue that while measuring signals in the low band, we shall most likely not be bothered by signals in the 1^+ frequency range.

Next let us see to what extent harmonic mixing complicates the situation. Harmonic mixing comes about because the LO provides a high-level drive signal to the mixer for efficient mixing, and since the mixer is a non-linear device, it generates harmonics of the LO signal. Incoming signals can mix against LO harmonics just as well as the fundamental, and any mixing product that equals the IF produces a response on the display. In other words, our tuning (mixing) equation now becomes:

$$F_s = nf_{LO} \pm f_{IF},$$

where n = LO harmonic, and the other parameters remain the same as before.

Let's add second-harmonic mixing to our graph in figure 58 and see to what extent this complicates our measurement procedure. As before, we shall first plot the LO frequency against the signal- frequency axis. Multiplying the LO frequency by two yields the upper dashed line of figure 59. As we did for fundamental mixing, we simply subtract the IF (3.6 GHz) from and add it to the LO second-harmonic curve to produce the 2^- and 2^+ tuning ranges. Since neither of these overlap the desired 1- tuning range, we can again argue that they do not really complicate the measurement process. In other words, signals in the 1-tuning range produce unique, unambiguous responses on our analyzer display. Should we be concerned about signals in the 2^-, 1^+, or higher bands producing ambiguous responses, we can add a simple external low-pass filter in front of our analyzer to eliminate them.

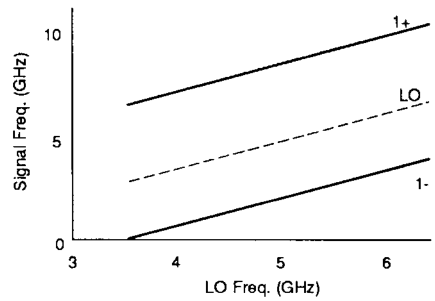


Figure 58. Tuning curves for fundamental mixing in the low-band, high-IF case

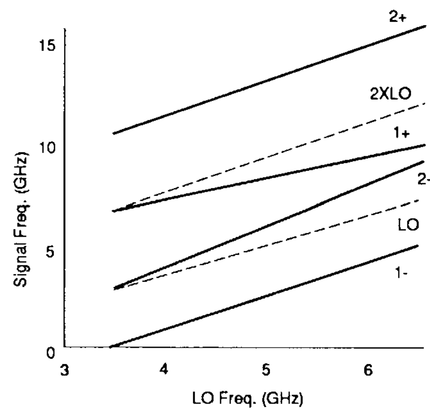


Figure 59. Signals in the 1-frequency range produce single, unambiguous responses in the low-band, high-IF case

The situation is considerably different for the high-band, low-IF case. As before, we shall start by plotting the LO fundamental against the signal-frequency axis and then add and subtract the IF, producing the results shown in figure 60. Note that the 1^- and 1^+ tuning ranges are much closer together, and in fact overlap, because the IF is a much lower frequency, 321.4 MHz in this case. Does the close spacing of the tuning ranges complicate the measurement process? Yes and no. First of all, our system can be calibrated for only one tuning range at a time. In this case, we would choose the 1^- tuning to give us a low-end frequency of 2.7 GHz so that we have some overlap with the 2.9 GHz upper end of our lowband tuning range. So what are we likely to see on the display? If we enter the graph at an LO frequency of 5 GHz, we find that there are two possible signal frequencies that would give us responses at the same point on the display: 4.7 and 5.3 GHz (rounding the numbers again). On the other hand, if we enter the signal frequency axis at 5.3 GHz, we find that in addition to the 1^+ response at an LO frequency of 5 GHz, we could also get a 1^- response if we allowed the LO to sweep as high as 5.6 GHz, twice the IF above 5 GHz. Also, if we entered the signal frequency graph at 4.7 GHz, we would find a 1^+ response at an LO frequency of about 4.4 GHz (twice the IF below 5 GHz) in addition to the 1^- response at an LO frequency of 5 GHz.

Here we see cases of images and multiple responses. Images are signals at different frequencies that produce responses at the same point on the display, that is, at the same LO frequency. As we can see from figure 60, images are separated by twice the IF. The multiple-response case results when, a single input signal (sinusoid) causes more than one response on the display, that is, a response at two or more LO frequencies (two in this case). Again, note that the LO frequencies producing the multiple responses are spaced by twice the IF.

Clearly we need some mechanism to differentiate between responses generated on the 1^- tuning curve for which our analyzer is calibrated and those produced on the 1^+ tuning curve. However, before we look at signal identification solutions, let's add harmonic-mixing curves to 22 GHz and see if there are any additional factors that we must consider in the signal-identification process. Figure 61 shows tuning curves up to the fourth LO harmonic.

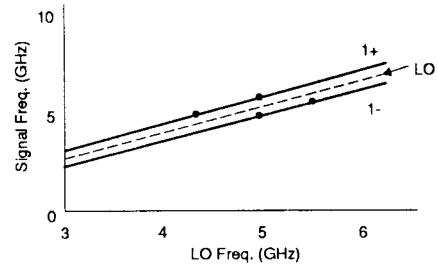


Figure 60. Tuning curves for fundamental mixing in the high-band, low-IF case

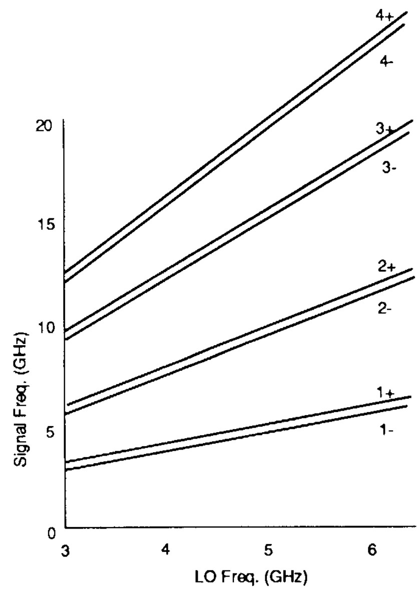


Figure 61. Tuning curves for n = 1 through 4 in the high-band, low-IF case

In examining figure 61, we find nothing really new, but rather an extension of the multiples and images that we discussed in figure 60. For example, we have image pairs for each of the LO harmonics. For an LO frequency of 5 GHz, we have a pair for fundamental mixing that we discussed in figure 60 at 4.7 and 5.3 GHz. For the second, third, and fourth harmonics of the LO, we have image pairs of 9.7 and 10.3, 14.7 and 15.3, and 19.7 and 20.3 GHz, respectively. The number of multiple responses that we get is a function of signal frequency and how far we sweep the LO. For example, if we sweep the LO over its full 3 to 6.5 GHz range, we get two responses for a 5-GHz input signal and four responses for an input signal at 10 GHz. Figure 62 shows two cases on an Agilent 71200, a spectrum analyzer with a wide-open front end (no filtering at the input prior to the first mixer).

Can we conclude from figures 61 and 62 that such a spectrum analyzer is not practical? Certainly not. Many of us work in controlled environments in which we deal with only one or two signals at a time. In such environments, analyzers like the Agilent 71200 work just fine. From figures 60 and 61 we conclude that image signals, if they exist, can be filtered away with simple bandpass filters and that multiple responses will not bother us if we limit our frequency span to something less than 600 MHz (twice the IF). And, knowing the signal frequencies, we can tune to the signal directly knowing that the analyzer will select the appropriate mode (1^-, 2^-, 3^+, or 4^+) for which it is calibrated.

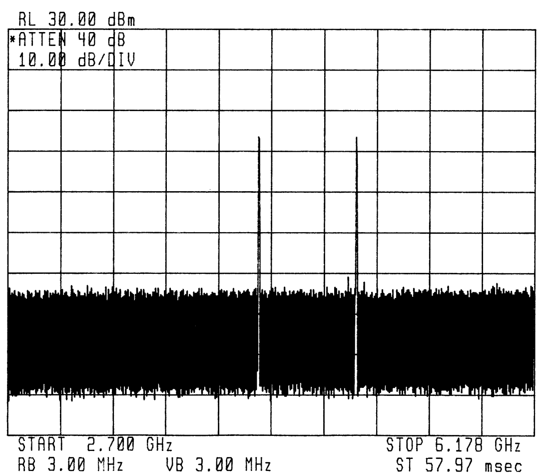


Figure 62A.

Figure 62B.

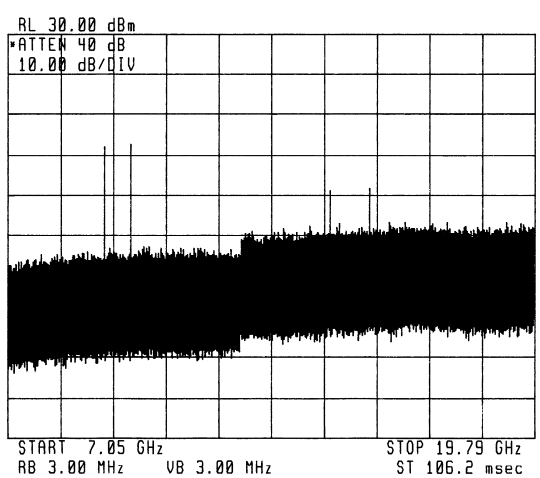


Figure 62. The number of responses is a function of signal frequency and analyzer span

Amplitude calibration

So far, we have seen that a harmonic-mixing spectrum analyzer does not always indicate the correct frequency of a given response. What about amplitude?

The conversion loss of a mixer is a function of harmonic number, and the loss goes up as the harmonic number goes up. (Here we are considering only those cases in which we observe a particular response on the correct mixing mode or tuning range.) This means that signals of equal amplitude would appear at different levels on the display if they involved different mixing modes. To preserve amplitude calibration, then, something must be done. For example, the reference level or the IF gain could be changed to compensate for the changing conversion loss. In Agilent spectrum analyzers, the IF gain is changed[1].

The increased conversion loss at higher LO harmonics causes a loss of sensitivity just as if we had increased the input attenuator. And since the IF gain change occurs after the conversion loss, the gain change is reflected by a corresponding change in the displayed noise level. See figure 63. So we can determine analyzer sensitivity on the harmonic-mixing ranges by noting the average displayed noise level just as we did on fundamental mixing.

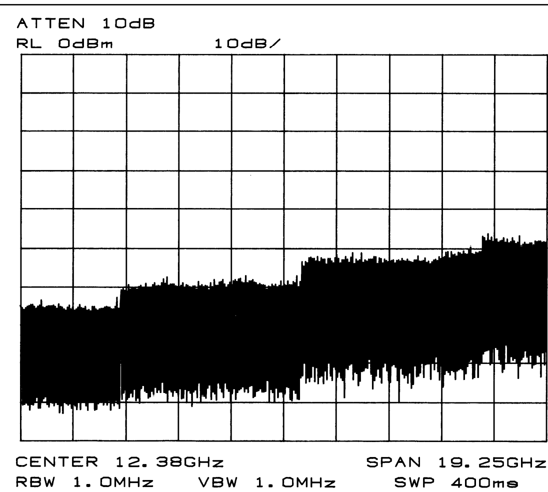


Figure 63. Steps in the noise floor indicate changes in sensitivity with changes in LO harmonic used in the mixing process

1 In the Agilent Modular Series, the display is shifted digitally in between the 10-dB IF gain steps.

Phase noise

In chapter 2 we noted that instability of an analyzer LO appears as phase noise around signals that are displayed far enough above the noise floor. We also noted that this phase noise can impose a limit on our ability to measure closely-spaced signals that differ in amplitude. Refer to figures 20 and 53. The level of the phase noise indicates the angular, or frequency, deviation of the LO.

What happens to phase noise when a harmonic of the LO is used in the mixing process? Relative to fundamental mixing, phase noise increases by:

$$20*\log(n),$$

where n = harmonic of the LO.

For example, suppose that the LO fundamental has a peak-to-peak deviation of 100 Hz. The second harmonic then has a 200-Hz peak-to-peak deviation; the third harmonic, 300 Hz; and so on. Since the phase noise indicates the signal (noise in this case) producing the modulation, the level of the phase noise must be higher to produce greater deviation. When the degree of modulation is very small, as in the situation here, the amplitude of the modulation side bands is directly proportional to the deviation of the carrier (LO). If the deviation doubles, then, the level of the side bands must also double in voltage; that is, increase by 6 dB or 20*log(2). As a result, the ability of our analyzer to measure closely spaced signals that are unequal in amplitude decreases as higher harmonies of the LO are used for mixing. Phase-noise levels for fundamental and fourth-harmonic mixing are shown in figure 64.

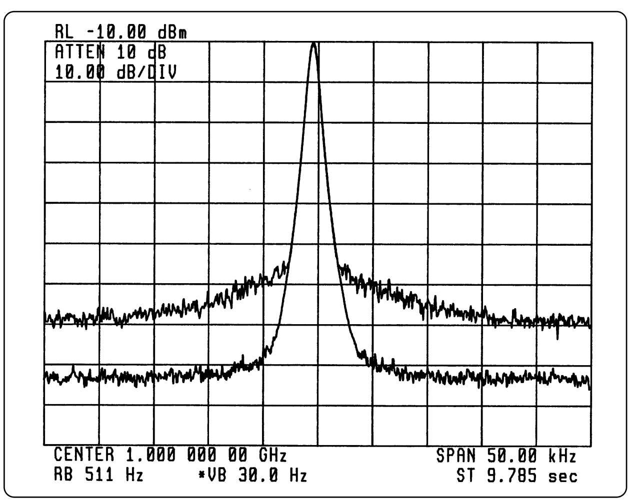


Figure 64. Difference in phase noise between fundamental and fourth-harmonic mixing

Signal identification

Even in a controlled situation, there are times when we must contend with unknown signals. In such cases, it is quite possible that the particular response we have tuned onto the display has been generated on an LO harmonic or mixing mode other than the one for which the display is calibrated. So our analyzer must have some way to tell us whether or not the display is calibrated for the signal response in question.

The Agilent 71200 offers two different identification methods: image and shift. We shall consider the image method first. Going back to figure 60, let's assume that we have tuned the analyzer to a frequency of 4.7 GHz (an LO frequency of 5 GHz), and we see a response in the center of the display. Let's further assume that the signal is either 4.7 or 5.3 GHz, but that we do not know which. If we use the image-identification process, the analyzer changes the first LO by twice the IF, first in one direction and then the other. If our signal is indeed at 4.7 GHz, when the analyzer changes its LO down in frequency, there is still a response (due to the 1^+ mixing mode) in the center of the display. On the other hand, when the LO is moved up, there is no response on the display. Thus we can conclude that the signal is indeed at 4.7 GHz and that the analyzer is properly tuned.

If, on the other hand, we had tuned our analyzer to 4.7 GHz (5 GHz-LO) and the input signal is actually 5.3 GHz, we would still have a response in the middle of the display. In this case, however, when we activate the image identification routine, there is no response when the LO is moved down by twice the IF and there is a response when the LO is moved up. This result tells us that when we are tuned to 4.7 GHz, we are actually observing the image of 4.7 GHz. So we must tune our analyzer higher in frequency by twice the IF, to 5.3 GHz (5.6-GHz LO), to observe the response on the 1^- mixing mode for which the analyzer is calibrated.

What happens if the response on the display is created by a harmonic of the LO different from the one for which the analyzer is calibrated? Referring to figure 65, suppose that we have tuned our analyzer to 4.7 GHz (5-GHz LO), but our input signal is actually 10.3 GHz. We shall see a response in the middle of the display from the 2^+ mixing mode. When we activate the image- identification process, the analyzer again moves the LO, up and down, by twice the IF. But neither change produces a response on the display. The test fails for both cases. We know that multiple responses for a given LO harmonic are separated by an LO difference of twice the IF. But here the response is generated by the second harmonic of the LO, so it is the second harmonic of the LO that we must change by twice the IF to tune from one response to the other. The image routine, at least as a first step, changed the fundamental of the LO by twice the IF and so changed the second harmonic by four times the IF. Hence the failure. Having failed, the system then divides the change by two and tries again. In this case the analyzer changes the LO fundamental by just 1*IF and so moves the second harmonic of the LO by the required 2*IF. Now when the LO moves up, the second of the response pair comes to mid-screen, and the test is successful.

The tests described so far are automatic, and a message appears on the display that tells us the signal frequency and gives us the opportunity to either tune the analyzer to that frequency or ignore the signal. The identification process can also be done manually. The manual routine is offered because noisy or modulated signals can sometimes fool the automatic process.

The image-identification method does not work on the low band (0 to 2.9 GHz) because, due to the high IF, we get only a single response in this band rather than a response pair. The second identification routine, the shift routine, works on this band as well as on the higher bands. This method involves changing the frequencies of two LOs in the analyzer rather than just one. Referring back to figure 57, consider what happens if we reduce the frequency of the 300 MHz LO to 298 MHz. To have a signal in the middle of the 21.4 MHz IF, the signal coming from the second IF must be 319.4 MHz; that is, the sum of 21.4 and 298 MHz. And if we are in the low band, as shown in figure 57, the new center frequency of the first IF is 3.6194 GHz (319.4 MHz plus 3.3 GHz). In any case, whether we are in the low-frequency, high-IF or the high-frequency, low-IF band, we have reduced the effective first IF by 2 MHz.

Although this method is called the shift method, we are actually looking for the absence of a shift to indicate that we are on the correct response on the correct band. To negate the downward change in the first IF, the first LO is also changed. If the band that we have selected on the analyzer uses a minus mixing mode - for example, 1^- or 2^- - the first LO is moved up in frequency. For the 3^+ and 4^+ mixing modes, the first LO is moved down. Since the appropriate harmonic for the band selected must shift 2 MHz, the actual change to the LO fundamental is 2/n MHz, where n is the appropriate harmonic number. As noted above, there is no

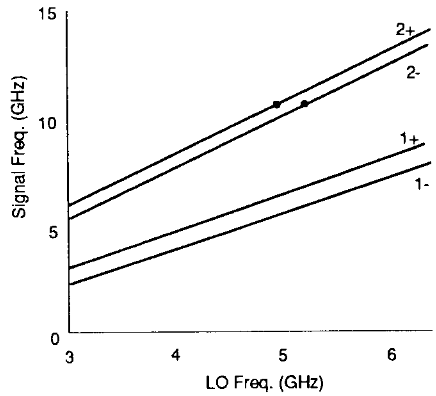


Figure 65. A response at a given LO frequency does not uniquely determine signal frequency

frequency shift of the displayed response when we are tuned to the correct response on the correct band. In all other cases there is a shift.

As with the image method, the shift method can be run automatically or manually. When run automatically, the Agilent 71200 indicates on its display whether the identified signal is in or out of band and, if the signal does not match the current tuning of the analyzer, gives us a choice of either tuning to the signal or ignoring it.

There is yet a third, totally manual identification routine. This method takes advantage of the fact that in the high-frequency, low-IF band the response pairs are easily located, as in figure 62. This method works particularly well when external mixers are used for measurements above 22 GHz. It also works well for modulated and noisy signals. This method has us tune halfway between the two responses of a given pair (for example, those in figure 62) and set the frequency span wide enough to see both responses. Then we simply note the indicated separation of the two responses. If the separation is twice the IF (642 MHz), then we have chosen the band with the correct harmonic number. If the responses are closer together than twice the IF, then they are produced on an LO harmonic higher than the harmonic utilized for the band we are on. If the indicated separation is greater than twice the IF, then the responses are produced on a lower LO harmonic. Once we have chosen the correct LO harmonic (by selecting a center frequency that yields a 642-MHz separation of the response pair), we can choose the correct response. For a minus mixing mode (1^- or 2^- on the Agilent 71200, for example), we would select the response displayed to the right; for a plus mixing mode (3^+ or 4^+ on the Agilent 71200), to the left.

Preselection

We made the case for the spectrum analyzer with a wide-open front end on the basis of a controlled measurement environment involving few, if any, unknown signals. However, there are many cases in which we have no idea how many signals are involved or what their frequencies might be. For example, we could be searching for unknown spurious signals, conducting site surveillance tests as part of a frequency-monitoring program, or performing EMI tests to measure unwanted device emissions. In all these cases, we could be looking for totally unknown signals in a potentially crowded spectral environment. Having to perform some form of identification routine on each and every response would make measurement time intolerably long. Hence the need for some form of pre-filtering or preselection.

What form must our preselection take? Referring back to figure 60, assume that we have the image pair 4.7 and 5.3 GHz present at the input of our analyzer. If we were particularly interested in one, we could use a band pass filter to allow that signal into the analyzer and reject the other. However, the fixed filter does not eliminate multiple responses, so if the spectrum is crowded there is still potential for confusion. More important, perhaps, is the restriction that a fixed filter puts on the flexibility of the analyzer. If we are doing broadband testing, we certainly do not want to be continually forced to change band pass filters.

The solution is a tunable filter configured in such a way that it automatically tracks the frequency of the appropriate mixing mode. Figure 66 shows the effect of such a preselector. Here we take advantage of the fact that our superheterodyne spectrum analyzer is not a real-time analyzer; that is, it tunes to only one frequency at a time. The dashed lines in figure 66 represent the bandwidth of the tracking preselector. Signals beyond the dashed lines are rejected. Suppose we have signals at 4.7 and 5.3 GHz present at the analyzer input. If we set a center frequency of 5 GHz and a span of 2 GHz, let's see what happens as the analyzer tunes across this range. As the LO sweeps past 4.4 GHz (the frequency at which it could mix with the 4.7 GHz input signal on its 1^+ mixing mode), the preselector is tuned to 4.1 GHz and therefore rejects the 4.7 GHz signal. Since the input signal does not reach the mixer, no mixing occurs, and no response appears on the display. As the LO sweeps past 5 GHz, the preselector allows the 4.7 GHz

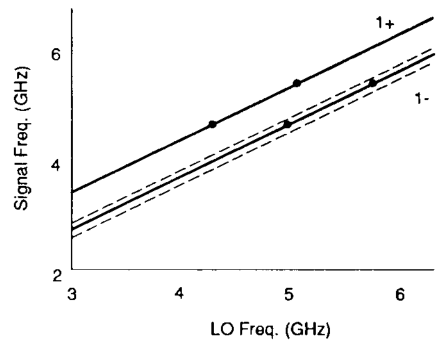


Figure 66. A preselector allows a signal to reach the mixer only when the analyzer is tuned to receive the signal

signal to reach the mixer, and we see the appropriate response on the display. The 5.3 GHz image signal is rejected, so it creates no mixing product to interact with the mixing product from the 4.7 GHz signal and cause a false display. Finally, as the LO sweeps past 5.6 GHz, the preselector allows the 5.3 GHz signal to reach the mixer, and we see it properly displayed. Note in figure 61 that nowhere do the various mixing modes intersect. So as long as the preselector bandwidth is narrow enough (it typically varies from 20 MHz at low frequencies to 80 MHz at high frequencies) it will eliminate all image and multiple responses.

The word eliminate may be a little strong. Preselectors do not have infinite rejection. Something in the 70- to 80-dB range is more likely. So if we are looking for very low-level signals in the presence of very high-level signals, we might see low-level images or multiples of the high-level signals.

What about the low band? Most tracking preselectors use YIG technology, and YIG filters do not operate well at low frequencies. Fortunately, there is a simple solution. Figure 59 shows that no other mixing mode overlaps the 1-mixing mode in the low frequency, high-IF case. So a simple low-pass filter attenuates both image and multiple responses. Figure 67 shows the input architecture of a typical microwave spectrum analyzer.

Improved dynamic range

A preselector improves dynamic range if the signals in question have sufficient frequency separation. The discussion of dynamic range in chapter 2 assumed implicitly that both the large and small signals were always present at the mixer and that their amplitudes did not change during the course of the measurement. But as we have seen, if signals are far enough apart, a preselector allows one to reach the mixer while rejecting the others. For example, if we were to test a microwave oscillator for harmonies, a preselector would reject the fundamental when we tuned the analyzer to one of the harmonics.

Let's look at the dynamic range of a second-harmonic test of a 3GHz oscillator. In figure 51 of chapter 2, we suggested a second-order specification for the input mixer such that a -40 dBm signal at the mixer would produce a second harmonic product –70 dBc. We also know from our discussion that for every dB the level of the fundamental changes at the mixer, measurement range also changes by 1 dB. The second-harmonic distortion curve has been regraphed in figure 68 with an extended range. For this example, we shall assume plenty of power from the oscillator and set the input attenuator so that when we measure the oscillator fundamental, the level at the mixer is −10 dBm, below the 1-dB compression point.

From the graph, we see that a −10 dBm signal at the mixer produces a second-harmonic distortion component 40 dB down. Now we tune the analyzer to the 6-GHz second harmonic. If the preselector has 70-dB rejection, the fundamental at the mixer has dropped to –80 dBm. Figure 68 indicates that for a signal of –80 dBm at the mixer, the internally generated distortion is −110 dBc, meaning 110 dB below the new fundamental level of –80 dBm. This puts the absolute level of the harmonic at −190 dBm. So the difference between the fundamental we tuned to and the internally generated second harmonic we tuned to is 180 dB!

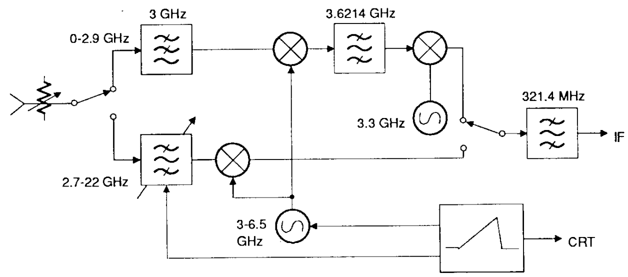


Figure 67. Front-end architecture of a typical preselected spectrum analyzer.

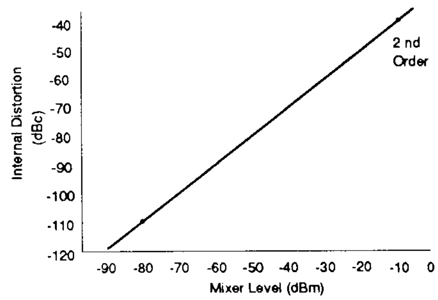


Figure 68. Extended second-order distortion graph

Clearly, for harmonic distortion, dynamic range is limited on the low-level (harmonic) end only by the noise floor (sensitivity) of the analyzer.

What about the upper, high-level end? When measuring the oscillator fundamental, we must limit the power at the mixer to get an accurate reading of the level. We can use either internal or external attenuation to limit the level of the fundamental at the mixer to something less than the 1-dB compression point. However, since the preselector highly attenuates the fundamental when we are tuned to the second harmonic, we can remove some attenuation if we need better sensitivity to measure the harmonic. A fundamental level of +20 dBm at the preselector should not affect our ability to measure the harmonic.[1]

1 Some sources can be damaged if high-level external signals are applied their output circuits. The preselector achieves its out-of-band rejection by reflecting the signal. If we select 0 dB of input attenuation for best sensitivity when measuring harmonics, we must remember that the fundamental is almost totally reflected.

Any improvement in dynamic range for third-order intermodulation measurements depends upon separation of the test tones versus preselector bandwidth. As we noted, typical preselector bandwidth is about 20 MHz at the low end and 80 MHz at the high end. As a conservative figure, we might use 18 dB per octave roll off of a typical three-sphere YIG filter beyond the 3-dB point. So to determine the improvement in dynamic range, we must determine to what extent each of the fundamental tones is attenuated and how that affects internally generated distortion. From the expressions in chapter 2 for third-order intermodulation, we have:

$$(k_4/8)V_{LO}V_1^2V_2 \cos[w_{LO} - (2w_1 - w_2)]t$$

and

$$(k_4/8)V_{LO}V_1V_2^2\cos[w_{LO} - (2w_2 - w_1)]t.$$

Looking at these expressions, we see that the amplitude of the lower distortion component $(2w_1 - w_2)$ varies as the square of V_1 and linearly with V_2. On the other side, the amplitude of the upper distortion component $(2w_2 - w_1)$ varies linearly with V1 and as the square of V_2. However, unlike the case in figure 50 of chapter 2, the preselector will not attenuate the two fundamental tones equally. Figure 69 illustrates the situation in which we are tuned to the lower distortion component and the two fundamental tones are separated by half the preselector bandwidth. In this case the lower-frequency test tone is attenuated 3 dB; the upper test tone, 21 dB (3 dB plus an additional 18 dB per octave away from center frequency). Since we are tuned to the lower distortion component, internally generated distortion at this frequency drops by a factor of two relative to the attenuation of V_1 and equally as fast as the attenuation of V_2. The improvement in dynamic range is a total of 27 dB. Improvements for other signal separations appear in the table included in figure 69. As in the case of second harmonic distortion, the noise floor of the analyzer must be considered, too. For very closely spaced test tones, the preselector provides no improvement, and we determine dynamic range as if the preselector was not there.

The discussion of dynamic range in chapter 2 also applies to the low-pass-filtered low band. The only exceptions occur when a particular harmonic of a low-band signal falls within the preselected range. For example, if we measure the second harmonic of a 1.5-GHz fundamental, we get the benefit of the preselector when we tune to the 3-GHz harmonic.

Dynamic Range Improvement

Sig. Sep (Presel BW's)	Improvement (dB)
0.25	3
0.5	27
1	81

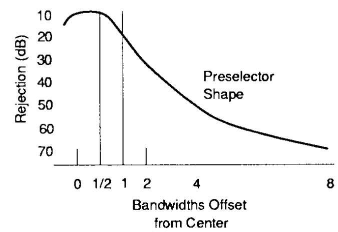


Fig. 69. Preselector attenuation and improvement in third-order intermodulation dynamic range.

Multiband tuning

Not only does a preselector effectively eliminate image and multiple responses, it makes tuning across wide frequency ranges practical. All Agilent spectrum analyzers with built-in preselectors allow tuning across the entire preselected range in a single sweep, as shown in figure 70A. Analyzers with microprocessors also allow spans less than the full preselected range that nevertheless involve more than one mixing mode.

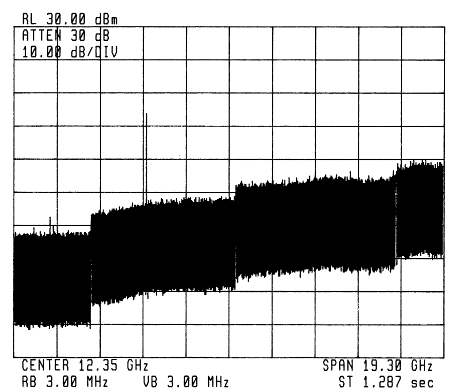


Figure 70A.

Figure 70B.

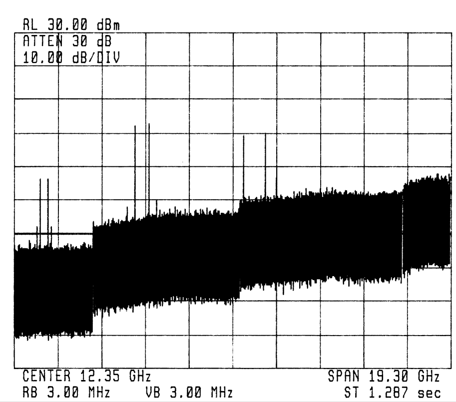


Figure 70. Preselection makes wide spans practical

The wide frequency spans are accomplished by continuously tuning the preselector while repeatedly retuning the LO as appropriate for the harmonic used in the particular mixing mode. The abrupt steps in the displayed noise floor occur because the IF gain is changed to compensate for the changing conversion loss in the mixer as the LO harmonic changes. For all practical purposes, then, the preselected range becomes a single tuning band. However, continual sweeping across the switch point between the low-pass-filtered low band and the preselected high band is not allowed because a mechanical switch is used to select the band, and continual operation of the switch would cause excessive wear.

The Agilent 71200 allows tuning over its entire tuning range because the same mixer is used on both low and high bands, and therefore no band switch is involved. However, because it is not preselected, this wide tuning is not as useful as on a preselected analyzer. See figure 70B.

Pluses and minuses of preselection

We have seen the pluses of preselection: simpler analyzer operation, uncluttered displays, improved dynamic range, and wide spans. But there are some minuses relative to the unpreselected analyzer as well.

First of all, the preselector has insertion loss, typically 6 to 8 dB. This loss comes prior to the first stage of gain, so system sensitivity is degraded by the full loss. In addition, when a preselector is connected directly to a mixer as shown in figure 67, the interaction of the mismatch of the preselector (typically 2.5 VSWR) with that of the input mixer (typically 3 VSWR) can cause a degradation of frequency response approaching ±2 dB. To minimize this interaction, a matching pad (fixed attenuator) or isolator is often inserted between the preselector and mixer. Sensitivity is degraded by the full value of the pad (6 to 10 dB) or isolator (1 to 2 dB). The lower loss of the isolator yields better sensitivity, but the better match of the pad yields better flatness.

Some architectures eliminate the need for the matching pad or isolator. As the electrical length between the preselector and mixer increases, the rate of change of phase of the reflected and rereflected signals becomes more rapid for a given change in input frequency. The result is a more exaggerated ripple effect on flatness. Architectures such as those used in the 8566A and B and the Agilent 71210 include the mixer diode as an integral part of the preselector/mixer assembly. In such an assembly, there is minimal electrical length between the preselector and mixer. This architecture thus removes the ripple effect on frequency response and improves sensitivity by eliminating the matching pad or isolator.

Even aside from the its interaction with the mixer, a preselector causes some degradation of frequency response. In most configurations, the tuning ramp for the preselector and local oscillator come from the same source, but there is no feedback mechanism to ensure that the preselector exactly tracks the tuning of the analyzer. As a result, analyzers such as the 8566B have both manual and automatic preselector-peak routines, and best flatness is obtained by peaking the preselector at each signal. The 8562A, on the other band, has preselector-peak values programmed into the firmware for each GHz along the frequency range, so specified frequency response is obtained without taking extra steps to peak the preselector.

Wideband fundamental mixing

Even though figure 67 is a simplified block diagram, if we look at it closely, we can find three areas for improved operation: ability to sweep across the low-band/high-band switch point, fundamental mixing across the entire frequency range for better sensitivity, and automatic preselector peaking for better amplitude accuracy and faster measurements.

All three areas are addressed in the Agilent 71210. First of all, this analyzer uses a solid-state switch that is part of the preselector circuit to switch between the low and high bands. As a result, the Agilent 71210 can sweep across the switch point continuously and simplify the analysis of spectra that straddle the switch point. The solid-state switch also permits continuous sweeps across the entire 0 to 22 GHz frequency range.

Second, fundamental mixing avoids the loss of sensitivity that results from harmonic mixing. Fundamental mixing could be achieved by using a 3 to 22 GHz fundamental oscillator (if one existed). The actual scheme used in the Agilent 71210 multiplies the 3 to 6.5 GHz LO as appropriate before it is applied to the mixer. Such an arrangement is illustrated in figure 71. In this case the sensitivity (noise floor) remains essentially constant across the entire frequency range, as shown in figure 72. The slight rise of the noise at the high end of the low band results from an increased loss in the solid-state switch.

The improved sensitivity gives the Agilent 71210 an advantage over harmonic-mixing analyzers when it comes to the measurement of low-level signals. Perhaps more important is the potential for reduced test times. For example, at 20 GHz the Agilent 71210 enjoys a sensitivity advantage of about 20 dB over the 8566B. For a test requiring a given sensitivity, then, the resolution bandwidth selected on the 71210 can be one hundred times wider than the bandwidth on the 8566B. We know from chapter 2 that sweep time (for analog filters) is inversely proportional to the square of the resolution bandwidth. So the 71210 has a potential measurement time advantage over the 8566B of 10,000:1!

Finally, proper preselector peaking plays a role in both amplitude accuracy and measurement time. In an open-loop configuration, the tuning of the preselector may not exactly match that of the analyzer. As a result, the preselector - a bandpass filter - will be mistuned to varying degrees as a function of frequency and so add to the non-flatness of the system. Stopping to optimize preselector tuning at each and every measurement point would add considerably to measurement time.

The 71210 achieves dynamic preselector peaking by including a fourth YIG sphere in the same assembly that includes the three spheres used to form the preselector filter. The fourth sphere is the resonant element in a discriminator circuit. The resonant frequency of a YIG sphere is determined by the strength of the magnetic field in which it is placed. All four spheres of the preselector/ discriminator are placed in the magnetic field of an electromagnet. The tuning ramp of the analyzer determines the current in the coil of the electromagnet and thus tunes the preselector/

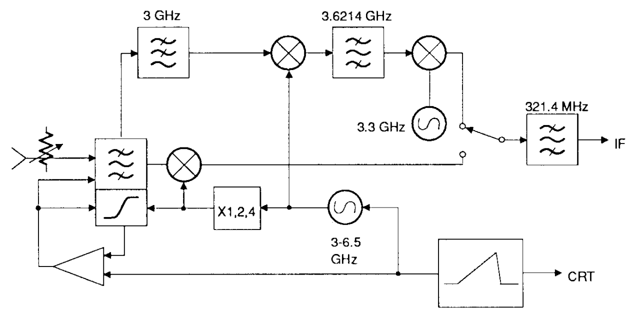


Figure 71. Front-end architecture of the Agilent 71210 with solid-state band switch, fundamental mixing to 22 GHz, and dynamic preselector peaking

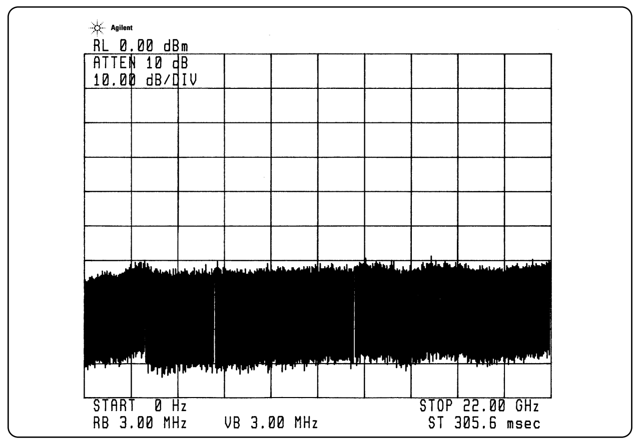


Figure 72. Fundamental mixing across the entire tuning range gives the 71210 the same sensitivity at 22 GHz as at 1 GHz

discriminator. There is a second, small coil within the preselector/discriminator assembly to adjust the magnetic field of only the discriminator sphere. The current in this small coil is such that the resonant frequency of the discriminator sphere is higher than the resonant frequency of the preselector spheres by 321.4 MHz, the first IF in the high-frequency, low-IF range. From the tuning equation, we know that 321.4 MHz is the frequency difference between the LO and an input signal for proper tuning of the analyzer to receive that signal.

Since the discriminator sphere resonates at a frequency 321.4 MHz higher than do the preselector spheres, if we can devise a scheme to adjust the current in the electromagnet to keep the discriminator sphere resonating at the LO frequency, the preselector will be properly tuned by definition. As shown in figure 71, there is indeed a feedback mechanism between the discriminator and the main tuning coil. When the discriminator sphere resonates at the LO frequency, there is no output, and no correction is added to the tuning ramp. Should the resonant frequency of the discriminator sphere differ from the LO frequency, the current through the electromagnet is not correct for the tuned frequency of the analyzer, and not only is the discriminator sphere mistuned, but the preselector is mistuned as well. But if the discriminator is mistuned, there is an output voltage that adds to or subtracts from the tuning ramp as appropriate to adjust the current in the electromagnet to bring the resonant frequency of the discriminator sphere back to the LO frequency. Again, because the discriminator sphere is properly tuned, the preselector is also properly tuned. Because this is a truly dynamic, real-time system, the preselector is always properly tuned and no other tuning or peaking mechanism is needed.

So an architecture based on figure 71 addresses all three areas of improvement suggested in reference to figure 67.

Summary

In this chapter we looked at harmonic mixing as a means of extending the frequency range of a spectrum analyzer. We found that, without some form of filtering ahead of the first mixer, the display can be complicated by image and multiple responses, and signal identification might be necessary. We next introduced the preselector, a tracking bandpass filter that essentially eliminates the unwanted responses. Finally, we looked at an improved input architecture that provides fundamental mixing over the entire frequency range, full-range sweeps, and a dynamically peaked preselector.

Glossary of terms

Absolute Amplitude Accuracy: The uncertainty of an amplitude measurement in absolute terms, either volts or power. Includes relative uncertainties (see Relative Amplitude Accuracy) plus calibrator uncertainty. For improved accuracy, some spectrum analyzers have frequency response specified relative to the calibrator as well as relative to the mid-point between peak-to-peak extremes.

Amplitude Accuracy: The uncertainty of an amplitude measurement, whether relative or absolute.

Analog Display: The case in which the analog signal information (from the envelope detector) is written directly to the display. Analyzers with vector displays, as opposed to raster displays, typically revert to an analog display on fast sweeps in zero span even though the normal display mode is digital.

Average Noise Level: See Displayed Average Noise Level.

Bandwidth Selectivity: A measure of an analyzer's ability to resolve signals unequal in amplitude. Also called shape factor, bandwidth selectivity is the ratio of the 60-dB bandwidth to the 3-dB bandwidth for a given resolution (IF) filter. For some analyzers, the 6-dB bandwidth is used in lieu of the 3-dB bandwidth. In either case, bandwidth selectivity tells us how steep the filter skirts are.

CRT Persistence: An indication of the rate at which the image fades on the display. In analyzers that digitize the trace (video) information before writing it to the screen, the refresh rate is high enough to prevent any flicker in the display, so short-persistence CRTs are used. Purely analog (older) analyzers typically use long persistence or variable -persistence CRTs because the refresh rate is the same as the sweep rate.

Delta Marker: A mode in which a fixed, reference marker has been established and a second, active marker is available that we can place anywhere on the displayed trace. A read out indicates the relative frequency separation and amplitude difference between the reference and active markers.

Digital Display: A mode in which trace (analog video) information is digitized and stored in memory prior to being displayed. The displayed trace is a series of points. The number of points is a function of the particular analyzer. Agilent analyzers draw vectors between the points to present a continuous looking trace. The display is refreshed (rewritten from data in memory) at a flicker-free rate; the data in memory is updated at the sweep rate.

Display Detector Mode: The manner in which the analog video information is processed prior to being digitized and stored in memory. See Neg Peak, Pos Peak, Rosenfell, and Sample.

Display Dynamic Range: The maximum dynamic range for which both the larger and smaller signal may be viewed simultaneously on the CRT. For analyzers with a maximum logarithmic display of 10 dB/div, the actual dynamic range (see Dynamic Range) may be greater than the display dynamic range.

Display Fidelity: The uncertainty in measuring relative differences in amplitude on a spectrum analyzer. On purely analog analyzers, differences must be read directly on the CRT display. Many analyzers with digital displays have markers, so differences are taken from stored data, and the ambiguity of the CRT display is removed from the measurement.

Display Range: The calibrated range of the CRT for the particular display mode and scale factor. See Linear and Log Display and Scale Factor.

Displayed Average Noise Level: The noise level as seen on the analyzer's display after setting the video bandwidth narrow enough to reduce the peak-to-peak noise fluctuations such that the displayed noise is seen as an essentially straight line. Usually refers to the analyzer's own internally generated noise as a measure of sensitivity and is typically specified in dBm under conditions of minimum resolution bandwidth and minimum input attenuation.

Drift: The slow (relative to sweep time) change of signal position on the display as a result of a change in LO frequency versus sweep voltage. While drift may require us to retune the analyzer periodically, it does not impair frequency resolution.

Dynamic Range: The ratio in dB between the largest and smallest signals, present at the input of an analyzer simultaneously, that can be measured to a given degree of accuracy. Dynamic range generally refers to measurement of distortion or intermodulation products.

Envelope Detector: Also called a peak detector, a circuit element whose output follows the envelope, but not the instantaneous variation, of its input signal. In a superheterodyne spectrum analyzer, the input to the envelope detector comes from the final IF, and the output is a video signal. When we put our analyzer in zero span, the envelope detector demodulates the input signal, and we can observe the modulating signal as a function of time on the display.

External Mixer: An independent mixer, usually with a waveguide input port, used to extend the frequency range of those spectrum analyzers designed to utilize them. The analyzer provides the LO signal and, if needed, mixer bias, and mixing products are returned to the analyzer's IF input.

FFT (Fast Fourier Transform): A mathematical operation performed on a time-domain signal to yield the individual spectral components that constitute the signal. See Spectrum.

Flatness: See Frequency Response.

Frequency Accuracy: The uncertainty with which the frequency of a signal or spectral component is indicated, either in an absolute sense or relative to some other signal or spectral component. Absolute and relative frequency accuracies are specified independently.

Frequency Range: The minimum to maximum frequencies over which a spectrum analyzer can tune. While the maximum frequency is generally thought of in terms of an analyzer's coaxial input, the range of many microwave analyzers can be extended through use of external wave guide mixers.

Frequency Resolution: The ability of a spectrum analyzer to separate closely spaced spectral components and display them individually. Resolution of equal amplitude components is determined by resolution bandwidth; of unequal amplitude signals, by resolution bandwidth and bandwidth selectivity.

Frequency Response: Variation in the displayed amplitude of a signal as a function of frequency. Typically specified in terms of ± dB relative to the value midway between the extremes. Also may be specified relative to the calibrator signal.

Frequency Span: The frequency range represented by the horizontal axis of the display. Generally, frequency span is given as the total span across the full display. Older analyzers usually indicate frequency span (scan width) on a per-division basis.

Frequency Stability: A general phrase that covers both short- and long-term LO instability. The sweep ramp that tunes the LO also determines where a signal should appear on the display. Any long-term variation in LO frequency (drift) with respect to the sweep ramp causes a signal to slowly shift its horizontal position on the display. Shorter-term LO instability can appear as random FM or phase noise on an otherwise stable signal.

Full Span: Depends upon the spectrum analyzer. For some, full span means a frequency span that covers the entire tuning range of the analyzer. These analyzers include single-band RF analyzers and microwave analyzers such as the 71210 that use a solid-state switch to switch between the low and preselected ranges. For other analyzers, full span refers to a sub-range. For example, for a microwave spectrum analyzer such as the AGILENT 8566B that uses a mechanical switch to switch between the low and preselected ranges, full span may refer to either the low, non-preselected range or the high, preselected range.

Gain Compression: That signal level at the input mixer of a spectrum analyzer at which the displayed amplitude of the signal is a specified number of dB too low due just to mixer saturation. The signal level is generally specified for 1-dB or 0.5-dB compression and is usually between –3 and –10 dBm.

Harmonic Distortion: Unwanted frequency components added to a signal as the result of the nonlinear behavior of the device (e.g. mixer, amplifier) through which the signal passes. These unwanted components are harmonically related to the original signal.

Harmonic Mixing: The utilization of the LO harmonics generated in a mixer to extend the tuning range of a spectrum analyzer beyond the range achievable using just the LO fundamental.

IF Gain/IF Attenuation: A control to adjust vertical the position of signals on the display without affecting the signal level at the input mixer. When changed, the value of the reference level is changed accordingly.

IF Feedthrough: A raising of the baseline trace on the display due to an input signal at the intermediate frequency passing through the input mixer. Generally a potential problem only on non-preselected spectrum analyzers. The entire trace is raised because the signal is always at the IF, i.e. mixing with the LO is not required.

Image Response: A displayed signal that is actually twice the IF away from the frequency indicated by the spectrum analyzer. For each harmonic of the LO there is an image pair, one below and one above the LO frequency by the IF. Images usually appear only on non-preselected spectrum analyzers.

Incidental FM: Unwanted frequency modulation on the output of a device (signal source, amplifier) caused by (incidental to) some other form of modulation, e.g. amplitude modulation.

Input Attenuator: A step attenuator between the input connector and first mixer of a spectrum analyzer. Also called the RF attenuator. The input attenuator is used to adjust level of the signal incident upon the first mixer. The attenuator is used to prevent gain compression due to high-level and/or broadband signals and to set dynamic range by controlling the degree of internally-generated distortion. In some analyzers, the vertical position of displayed signals is changed when the input attenuator setting is changed, so the reference level is also changed accordingly. In AGILENT microprocessor-controlled analyzers, the IF gain is changed to compensate for input attenuator changes, so signals remain stationary on the CRT display, and the reference level is not changed.

Input Impedance: The terminating impedance that the analyzer presents to the signal source. The nominal impedance for RF and microwave analyzers is usually 50 ohms. For some systems, e.g. cable TV, 75 ohms is standard. The degree of mismatch between the nominal and actual input impedance is given in terms of VSWR (voltage standing wave ratio).

Intermodulation Distortion: Unwanted frequency components resulting from the interaction of two or more spectral components passing through a device with non-linear behavior (e.g. mixer, amplifier). The unwanted components are related to the fundamental components by sums and differences of the fundamentals and various harmonics, e.g. $f_1 \pm f_2$, $2*f_1 \pm f_2$, $2*f_2 \pm f_1$, $3*f_1 \pm 2*f_2$, etc.

LO Emission or Feedout: The emergence of the LO signal from the input of a spectrum analyzer. The level can be greater than 0 dBm on non-preselected spectrum analyzers but is usually less than -70 dBm on preselected analyzers.

LO Feedthrough: The response on the display when a spectrum analyzer is tuned to 0 Hz, i.e. when the LO is tuned to the IF. The LO feedthrough can be used as a 0-Hz marker, and there is no frequency error.

Linear Display: The display mode in which vertical deflection on the display is directly proportional to the voltage of the input signal. The bottom line of the graticule represents 0 V, and the top line, the reference level, some non-zero value that depends upon the particular spectrum analyzer. On the Agilent 140 series of analyzers, we select a specific scale factor in V/div. On most newer analyzers, we select the reference level, and the scale factor becomes the reference level value divided by the number of graticule divisions. Although the display is linear, analyzers with microprocessors allow reference level and marker values to be indicated in dBm, dBmV, dBuV, and, in some cases, watts as well as volts.

Log Display: The display mode in which vertical deflection on the display is a logarithmic function of the voltage of the input signal. We set the display calibration by selecting the value of the top line of the graticule, the reference level, and scale factor in dB/div. On Agilent analyzers, the bottom line of the graticule represents zero volts for scale factors of 10 dB/div or more, so the bottom division is not calibrated in these cases. Analyzers with microprocessors allow reference level and marker values to be indicated in dBm, dBmV, dBuV, volts, and, in some cases, watts. Non-microprocessor based analyzers usually offer only one choice of units, and dBm is the usual choice.

Marker: A visible indicator that we can place anywhere along the CRT trace. A read out indicates the absolute value of both the frequency and amplitude of the trace at the marked point. The amplitude value is given in the currently selected units. Also see Delta Marker and Noise Marker.

Measurement Range: The ratio, expressed in dB, of the maximum signal level that can be measured (usually the maximum safe input level) to the lowest achievable average noise level. This ratio is almost always much greater than can be realized in a single measurement. See Dynamic Range.

Mixing Mode: A description of the particular circumstance that creates a given response on a spectrum analyzer. The mixing mode, e.g. 1^+, indicates the harmonic of the LO used in the mixing process and whether the input signal is above (+) or below (-) that harmonic.

Multiple Responses: Two or more responses on a spectrum analyzer display from a single input signal. Multiple responses occur only when mixing modes overlap and the LO is swept over a wide enough range to allow the input signal to mix on more that one mixing mode. Normally not encountered in analyzers with preselectors.

Neg Peak: For digital displays, the display detection mode in which each displayed point indicates the minimum value of the video signal for that part of the frequency span and/or time interval represented by the point.

Noise figure: The ratio, usually expressed in dB, of the signal-to-noise ratio at the input of a device (mixer, amplifier) to the signal-to-noise ratio at the output of the device.

Noise Marker: A marker whose value indicates the noise level in a 1-Hz noise power bandwidth. When the noise marker is selected, the sample display detection mode is activated, the values of a number of consecutive trace points (the number depends upon the analyzer) about the marker are averaged, and this average value is normalized to an equivalent value in a 1-Hz noise power bandwidth. The normalization process accounts for detection and bandwidth plus the effect of the log amplifier when we select the log display mode.

Noise Sidebands: Modulation sidebands that indicate the short-term instability of the LO (primarily the first LO) system of a spectrum analyzer. The modulating signal is noise, in the LO circuit itself and/or in the LO stabilizing circuit, and the sidebands comprise a noise spectrum. The mixing process transfers any LO instability to the mixing products, so the noise sidebands appear on any spectral component displayed on the analyzer far enough above the broadband noise floor. Because the sidebands are noise, their level relative to a spectral component is a function of resolution bandwidth. Noise sidebands are typically specified in terms of dBc/Hz (amplitude in a 1-Hz bandwidth relative to the -carrier) at a given offset from the carrier, the carrier being a spectral component viewed on the display.

Phase Noise: See Noise Sidebands.

Pos Peak: For digital displays, the display detection mode in which each displayed point indicates the maximum value of the video signal for that part of the frequency span and/or time interval represented by the point.

Preamplifier: An external, low-noise-figure amplifier that improves system (preamplifier/spectrum analyzer) sensitivity over that of the analyzer itself.

Preselector: A tunable, bandpass filter that precedes the input mixer of a spectrum analyzer and tracks the appropriate mixing mode. Preselectors are typically used only above 2 GHz. They essentially eliminate multiple and image responses and, for certain signal conditions, improve dynamic range.

Raster Display: A TV-like display in which the image is formed by scanning the electron beam rapidly across and slowly down the CRT face and gating the beam on as appropriate. The scanning rates are fast enough to produce a flicker-free display. Also see Vector Display and Sweep Time.

Reference Level: The calibrated vertical position on the display used as a reference for amplitude measurements. The reference level position is normally the top line of the graticule, but on 71000 spectrum analyzers the reference level position may be located anywhere.

Relative Amplitude Accuracy: The uncertainty of an amplitude measurement in which the amplitude of one signal is compared to the amplitude of another regardless of the absolute amplitude of either. Distortion measurements are relative measurements. Contributors to uncertainty include frequency response and display fidelity and changes of input attenuation, IF gain, scale factor, and resolution bandwidth.

Residual FM: The inherent short-term frequency instability of an oscillator in the absence of any other modulation. In the case of a spectrum analyzer, we usually expand the definition to include the case in which the LO is swept. Residual FM is usually specified in peak-to-peak values because they are most easily measured on the display if visible at all.

Residual Responses: Discrete responses seen on a spectrum analyzer display with no input signal present.

Resolution: See Frequency Resolution.

Resolution Bandwidth: The width of the resolution bandwidth (IF) filter of a spectrum analyzer at some level below the minimum insertion-loss point (maximum deflection point on the display). For Agilent analyzers, the 3-dB bandwidth is specified; for some others, it is the 6-dB bandwidth.

Rosenfell: For digital displays, the display detection mode in which the value displayed at each point is based upon whether or not the video signal both rose and fell during the frequency and/or time interval represented by the point. If the video signal only rose or only fell, the maximum value is displayed. If the video signal did both rise and fall, then the maximum value during the interval is displayed by odd-numbered points, the minimum value, by even-numbered points. To prevent the loss of a signal that occurs only in an even-numbered interval, the maximum value during this interval is preserved, and in the next (odd-numbered) interval, the displayed value is the greater of the value carried over or the maximum that occurs in the current interval.

Sample: For digital displays, the display detection mode in which the value displayed at each point is the instantaneous value of the video signal at the end of the frequency span and/or time interval represented by the point.

Scale Factor: The per-division calibration of the vertical axis of the display.

Sensitivity: The level of the smallest sinusoid that can be observed on a spectrum analyzer, usually under optimized conditions of minimum resolution bandwidth, 0 dB RF input attenuation, and minimum video bandwidth. Agilent defines sensitivity as the displayed average noise level. A sinusoid at that level will appear to be about 2 dB above the noise.

Shape Factor: See Bandwidth Selectivity.

Signal Identification: A routine, either manual or automatic, that indicates whether or not a particular response on the spectrum analyzer's display is from the mixing mode for which the display is calibrated. If automatic, the routine may change the analyzer's tuning to show the signal on the correct mixing mode, or it may tell us the signal's frequency and give us the option of ignoring the signal or having the analyzer tune itself properly for the signal. Generally not needed on preselected analyzers.

Span Accuracy: The uncertainty of the indicated frequency separation of any two signals on the display.

Spectral Purity: See Noise Sidebands.

Spectral Component: One of the sine waves comprising a spectrum.

Spectrum: An array of sine waves of differing frequencies and amplitudes and properly related with respect to phase that, taken as a whole, constitute a particular time-domain signal.

Spectrum Analyzer: A device that effectively performs a Fourier transform and displays the individual spectral components (sine waves) that constitute a time-domain signal. Phase may, or may not, be preserved, depending upon the analyzer type and design.

Spurious Responses: The improper responses that appear on a spectrum analyzer display as a result of the input signal. Internally generated distortion products are spurious responses, as are image and multiple responses.

Sweep Time: The time to tune the LO across the selected span. Sweep time does not include the dead time between the completion of one sweep and the start of the next. In zero span, the spectrum analyzer's LO is fixed, so the horizontal axis of the display is calibrated in time only. In non-zero spans, the horizontal axis is calibrated in both frequency and time, and sweep time is usually a function of frequency span, resolution bandwidth, and video bandwidth.

Units: Dimensions on the measured quantities. Units usually refer to amplitude quantities because they can be changed. In spectrum analyzers with microprocessors, available units are dBm (dB relative to 1 milliwatt dissipated in the nominal input impedance of the analyzer), dBmV (dB relative to 1 millivolt), dBuV (dB relative to 1 microvolt), volts, and, in some analyzers, watts. In AGILENT analyzers, we can specify any units in both log and linear displays.

Variable Persistence: That property of a CRT that allows adjustment of the fade rate of a trace created by the CRTs electron beam. For purely analog displays. Used in the Agilent 141T Spectrum Analyzer Mainframe for flicker-free displays regardless of sweep time.

Vector Display:
The CRT in which the electron beam is directed so that the image (trace, graticule, annotation) is written directly on the CRT face, not created from a series of dots as in the raster display.

Video: In a spectrum analyzer, a term describing the output of the envelope detector. The frequency range extends from 0 Hz to a frequency typically well beyond the widest resolution bandwidth available in the analyzer. However, the ultimate bandwidth of the video chain is determined by the setting of the video filter.

Video Amplifier: A post-detection, dc-coupled amplifier that drives the vertical deflection plates of the CRT. See Video Bandwidth and Video Filter.

Video Average: A digital averaging of a spectrum analyzer's trace information. Available only on analyzers with digital displays. The averaging is done at each point of the display independently and is completed over the number of sweeps selected by the user. The averaging algorithm applies a weighting factor (1/n, where n is the number of the current sweep) to the amplitude value of a given point on the current sweep, applies another weighting factor [(n - 1)/n] to the previously stored average, and combines the two for a current average. After the designated number if sweeps are completed, the weighting factors remain constant, and the display becomes a running average.

Video Bandwidth: The cutoff frequency (3-dB point) of an adjustable low pass filter in the-video circuit. When the video bandwidth is equal to or less than the resolution bandwidth, the video circuit cannot fully respond to the more rapid fluctuations of the output of the envelope detector. The result is a smoothing of the trace, i.e. a reduction in the peak-to-peak excursion of broadband signals such as noise and pulsed RF when viewed in the broadband mode. The degree of averaging or smoothing is a function of the ratio of the video bandwidth to the resolution bandwidth.

Video Filter: A post-detection, low-pass filter that determines the bandwidth of the video amplifier. Used to average or smooth a trace. See Video Bandwidth.

Zero Span: That case in which a spectrum analyzer's LO remains fixed at a given frequency so the analyzer becomes a fixed-tuned receiver. The bandwidth of the receiver is that of the resolution (IF) bandwidth. Signal amplitude variations are displayed as a function of time. To avoid any loss of signal information, the resolution bandwidth must be as wide as the signal bandwidth. To avoid any smoothing, the video bandwidth must be set wider than the resolution bandwidth.

Index

Agilent Technologies' Test and Measurement Support, Services, and Assistance

Agilent Technologies aims to maximize the value you receive, while minimizing your risk and problems. We strive to ensure that you get the test and measurement capabilities you paid for and obtain the support you need. Our extensive support resources and services can help you choose the right Agilent products for your applications and apply them successfully. Every instrument and system we sell has a global warranty. Support is available for at least five years beyond the production life of the product. Two concepts underlie Agilent's overall support policy: "Our Promise" and "Your Advantage."

Our Promise

Our Promise means your Agilent test and measurement equipment will meet its advertised performance and functionality. When you are choosing new equipment, we will help you with product information, including realistic performance specifications and practical recommendations from experienced test engineers. When you use Agilent equipment, we can verify that it works properly, help with product operation, and provide basic measurement assistance for the use of specified capabilities, at no extra cost upon request. Many self-help tools are available.

Your Advantage

Your Advantage means that Agilent offers a wide range of additional expert test and measurement services, which you can purchase according to your unique technical and business needs. Solve problems efficiently and gain a competitive edge by contacting us for calibration, extra-cost upgrades, out-of-warranty repairs, and onsite education and training, as well as design, system integration, project management, and other professional services.Experienced Agilent engineers and technicians worldwide can help you maximize your productivity, optimize the return on investment of your Agilent instruments and systems, and obtain dependable measurement accuracy for the life of those products.

For more assistance with your test and measurement needs go to

www.agilent.com/find/assist

Or contact the test and measurement experts at Agilent Technologies
(During normal business hours)

United States:
(tel) 1 800 452 4844

Canada:
(tel) 1 877 894 4414
(fax) (905) 206 4120

Europe:
(tel) (31 20) 547 2000

Japan:
(tel) (81) 426 56 7832
(fax) (81) 426 56 7840

Latin America:
(tel) (305) 267 4245
(fax) (305) 267 4286

Australia:
(tel) 1 800 629 485
(fax) (61 3) 9272 0749

New Zealand:
(tel) 0 800 738 378
(fax) 64 4 495 8950

Asia Pacific:
(tel) (852) 3197 7777
(fax) (852) 2506 9284

Product specifications and descriptions in this document subject to change without notice.

Printed in USA November 30, 2000
5952-0292

Agilent Technologies
Innovating the HP Way