Theory of Intermodulation and Reciprocal Mixing: Practice, Definitions and Measurements in Devices and Systems, Part 1

A man of renown in the communications field begins an explanation of receiver performance.

By Ulrich L. Rohde, KA2WEU/DJ2LR/HB9AWE

This two-part paper deals with both analog and digital receivers and gives a treatment of intermodulation distortion. Starting from a mathematical—but simple—derivation, it looks at gain compression, intermodulation distortion and dynamic range. Furthermore, a new dynamic measure is introduced that allows one to differentiate between different receive systems much better. Other important characteristics are triple-beat distortion and cross modulation. To describe the performance of broadband systems, noise-power ratio is also considered.

Large-signal effects which go beyond the linear range, such as AM-to-PM conversion, spectral re-growth and adjacent-channel power ratio are considered, as well as phase-response differential group delay and reciprocal mixing.

Synergy Microwave Corporation
201 McLean Blvd
Paterson, NJ 07504
ulrohde@aol.com

This tutorial is a good prerequisite in preparation for looking into high-performance components such as high-IP_3 mixers and modern receiver architectures.

Amplitude Linearity Issues and Figures of Merit[1]

A network's amplitude nonlinearity can be characterized by the expansion:

$$y = k_1 f(x) + k_2[f(x)]^2 + k_3[f(x)]^3 + \text{higher - order terms} \quad \text{(Eq 1)}$$

where y represents the output, the coefficients k_n represent complex quantities whose values can be determined by an analysis of the output waveforms, and $f(x)$ represents the input. Although all practical networks exhibit amplitude nonlinearity, we can (and often do) refer to many networks as "linear." We say this of networks that are *sufficiently amplitude-linear for our purposes*—for example: weakly nonlinear networks in which small-signal opera-

[1]Notes appear on page 15.

tion is assumed even though the signal levels involved are sufficient to cause slight distortion. For many practical purposes, the first three terms of Eq 1 adequately describe such a network's nonlinearity:

$$y = k_1 f(x) + k_2[f(x)]^2 + k_3[f(x)]^3 \qquad \text{(Eq 2)}$$

In adopting this simplification, we assume also that the nonlinearity is frequency-independent; that is, that the network has sufficient bandwidth to allow all of the products predicted by Eq 2 to appear at its output terminals unperturbed.

When multiple signals are present in a network, even weak nonlinearity can result in profound consequences. To illustrate this, we'll let $f(x)$ consist of the sum of two sinusoidal signals:

$$f(x) = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t \qquad \text{(Eq 3)}$$

We'll assume that ω_1 and ω_2 are close enough so that the coefficients k_i can be considered equal for both signals. We'll also assume, for simplicity, that all of the k_i are real. If Eq 2 describes the network's response to an input $f(x)$, the response will be as shown in Eq 4 below.

The k_1 term of Eq 4 represents the results of amplitude-linear behavior. No new frequency components have appeared; the two sine waves have merely been "rescaled" by k_1.

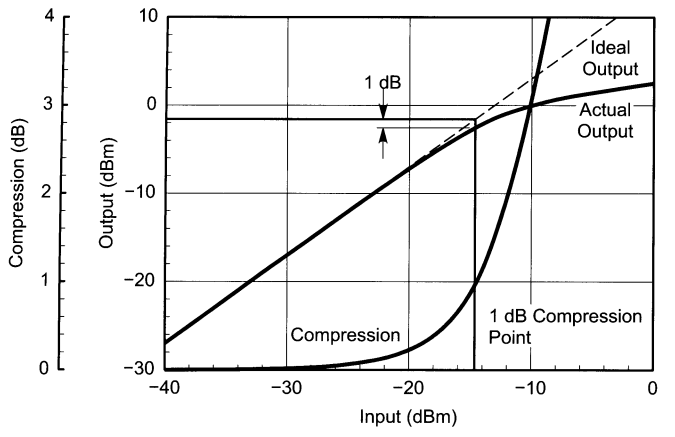


Fig 1—The power level at which a network's power output is down 1 dB relative to that of its ideally linear equivalent is a figure of merit known as the *1-dB compression point* (P_{-1dB}). The 1-dB compression point can be expressed relative to input power ($P_{-1dB,in}$) or output power ($P_{-1dB,out}$). For the amplifier simulated here, $P_{-1dB,in} \approx -14.5$ dBm and $P_{-1dB,out} \approx -1.3$ dBm.

The second- and third-order terms of Eq 4 represent the effects of harmonic distortion and intermodulation distortion. Second-order effects include second-harmonic distortion (the production of new signals at $2\omega_1$ and $2\omega_2$) and IMD (the production of new signals at $\omega_1 + \omega 2$ and $\omega_1 - \omega_2$). Third-order effects include gain compression, third-harmonic distortion (the production of new signals at $3\omega_1$ and $3\omega_2$), and IMD (the production of new signals at $2\omega_1 \pm \omega_2$ and $2\omega_2 \pm \omega_1$).

Gain Compression

Gain compression occurs when a network cannot increase its output amplitude in linear proportion to an amplitude increase at its input. Gain *saturation* occurs when a network's output amplitude stops increasing (in practice, it may actually decrease) with increases in input amplitude. We can deduce from Eq 4 that the amplitude of the cos $\omega_1 t$ signal has become:

$$A_1' = k_1 A_1 + k_3\left(\tfrac{3}{4} A_1^3 + \tfrac{3}{2} A_1 A_2^2\right) \qquad \text{(Eq 5)}$$

Because k_3 will normally be negative, a large signal $A_2 \cos \omega_2 t$ can effectively mask a smaller signal $A_1 \cos \omega_1 t$ by reducing the network's gain. This third-order effect, known as *blocking* or *desensitization* when it occurs in a receiver, is a special case of gain compression. The presence of additional signals results in a greater reduction of gain; the gain reduction for each signal is a function of the relative levels of all signals present. A receiver's blocking behavior may be characterized by the level of off-channel signal necessary to reduce the strength of an in-band signal by a specified value, typically 1 dB. Alternatively, the decibel ratio of the off-channel signal's power to the receiver's noise-floor power may be cited as *blocking dynamic range*. Desensitization may be also characterized in terms of the

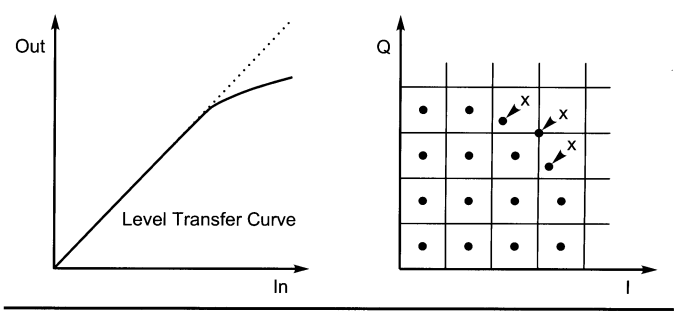


Fig 2—The influence of differential amplitude error (compression) on a QAM constellation.

$$\begin{aligned} y &= k_1(A_1\cos\omega_1 t + A_2\cos\omega_2 t) + k_2(A_1\cos\omega_1 t + A_2\cos\omega_2 t)^2 + k_3(A_1\cos\omega_1 t + A_2\cos\omega_2 t)^3 \\ &= k_1(A_1\cos\omega_1 t + A_2\cos\omega_2 t) + k_2\left[A_1^2\frac{1+\cos 2\omega_1 t}{2} + A_2^2\frac{1+\cos 2\omega_2 t}{2} + A_1A_2\frac{\cos(\omega_1+\omega_2)t + \cos(\omega_1-\omega_2)t}{2}\right] + \\ &\quad k_3\left\{\begin{array}{l}\left[A_1^3\left(\frac{\cos\omega_1 t}{2} + \frac{\cos\omega_1 t}{4} + \frac{\cos 3\omega_1 t}{4}\right) + A_2^3\left(\frac{3\cos\omega_2 t}{4} + \frac{\cos 3\omega_2 t}{4}\right)\right] \\ + A_1^2A_2\left[\frac{3}{2}\cos\omega_2 t + \frac{3}{4}\cos(2\omega_1-\omega_2)t + \frac{3}{4}\cos(2\omega_1-\omega_2)t\right] \\ + A_2^2A_1\left[\frac{3}{2}\cos\omega_1 t + \frac{3}{4}\cos(2\omega_2+\omega_1)t + \frac{3}{4}\cos(2\omega_2-\omega_1)t\right]\end{array}\right\} \end{aligned} \qquad \text{(Eq 4)}$$

off-channel-signal power necessary to degrade a system's SNR by a specified value.

Multiple signals need not be present for gain compression to occur. If only one signal were present, the ratio of gain with distortion to the network's idealized (linear) gain would be:

$$A_1' = \frac{k_1 + k_3\left(\frac{3}{4}A_1^2\right)}{k_1} \qquad \text{(Eq 6)}$$

This is referred to as the *single-tone gain-compression factor*. Fig 1 shows how the k_3 term causes a network's gain to deviate from the ideal. The point at which a network's power gain is down 1 dB from the ideal for a single signal is a figure of merit known as the *1-dB compression point* (P_{-1dB}). Many networks—including many receiving and low-level transmitting circuits, such as low-noise amplifiers, mixers and IF amplifiers—are usually operated under small-signal conditions, at levels sufficiently below P_{-1dB} to maintain high linearity. As we'll see, however, some networks—including power amplifiers for wireless systems—may be operated under large-signal conditions near or in compression to achieve optimum efficiency at some specified level of linearity. Fig 2 shows what happens when a digital emission that uses amplitude to convey information is subjected to amplitude compression.

Intermodulation

The new signals produced through IMD can profoundly affect the performance of systems even when they are operated well below gain compression (Fig 3). IMD products of significant power can appear at frequencies remote from, in or near the system passband, resulting in demodulation errors in reception and interference to other communication in transmission. Where an IMD product appears relative to the passband depends on the passband width and center frequency, the frequencies of the signals present at the system input and the order of the nonlinearity involved. These factors also determine the strength of an IMD product relative to the desired signal.

Second-order IMD (IM_2) results, for an input consisting of two signals ω_1 and ω_2, in the production of new signals at $\omega_1 + \omega_2$ and $\omega_1 - \omega_2$. Third-order IMD (IM_3) results, for an input consisting of two signals ω_1 and ω_2, in the production of new signals at $2\omega_1 \pm \omega_2$ and $2\omega_2 \pm \omega_1$.

Under small-signal conditions well below compression, the power of an IM_2 product varies by 2 dB, and the power of an IM_3 product varies by 3 dB, per decibel change in input power level. This allows us to derive a network figure of merit, the *intermodulation intercept point* (*IP*), for a given intermodulation order. We can do so by extrapolating a network's linear and intermodu-lation responses to their point of intersection (Fig 4). This is the point at which their powers would be equal if compression did not occur.

Because of the system noise and/or intermodulation distortion products, there is a minimum discernible signal (*MDS*) that limits the dynamic range at the lower end. Theoretically, Fig 4 should show a noise floor or IMD-spur floor for a given input signal that represents a lower limit below which signals cannot be detected. The intercept point for a given intermodulation order n can be expressed, and should always be characterized, relative to input power ($IP_{n,in}$) or output power ($IP_{n,out}$). The IP_{in} and IP_{out} values differ by the network's linear gain. For equal-level test tones, $IP_{n,in}$ can be determined by:

$$IP_{n,in} = \frac{nP_A - P_{IM_n}}{n-1} \qquad \text{(Eq 7)}$$

where n is the order, P_A is the input power (of one tone), P_{IMn} is the power of the intermodulation product, and *IP*

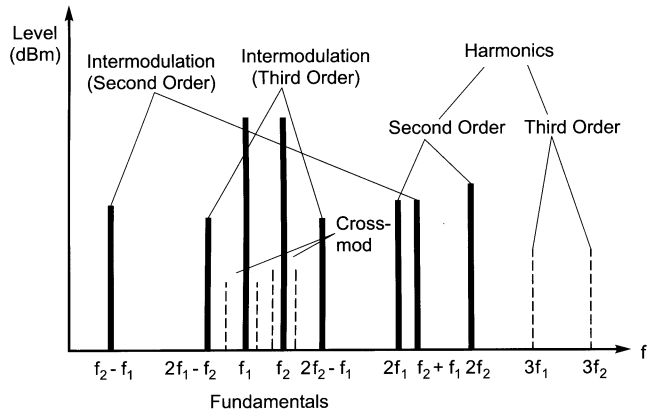


Fig 3—Relationships between fundamental and spurious signals, including harmonics and products of intermodulation.

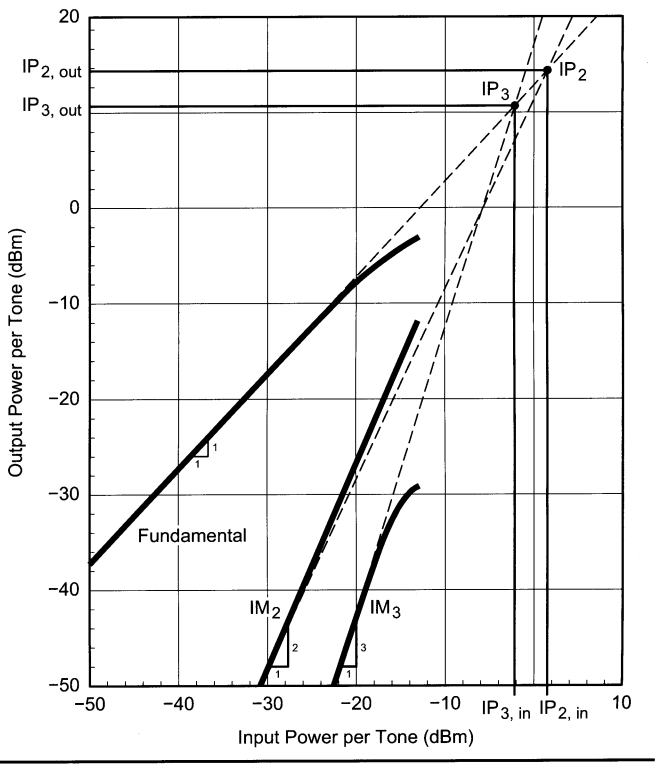


Fig 4 (right)—The level at which the power of one of a network's intermodulation products equals that of the network's linear output is a figure of merit known as the *intermodulation intercept point* (*IP*). The intercept point for a given intermodulation order, *n*, can be expressed, and should always be characterized, relative to input power ($IP_{n,in}$) or output power ($IP_{n,out}$); the IP_{in} and IP_{out} values differ by the network's linear gain. For the amplifier simulated here, $IP_{2,in} \approx 1.5$ dBm, $IP_{2,out} \approx 14.5$ dBm, $IP_{3,in} \approx -2.3$ dBm and $IP_{3,out} \approx 10.7$ dBm. Each curve depicts the power in one tone of the response evaluated.

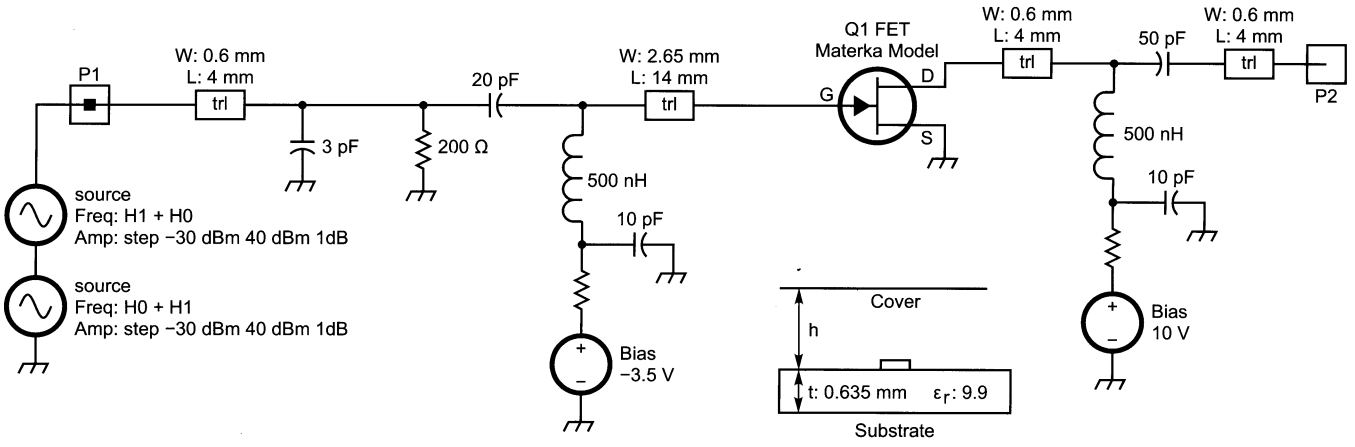


Fig 5—Circuit of a single-stage GaAs-FET amplifier operating at 120 mA used to demonstrate IP_2 and IP_3 for CAD analysis.

is the intercept point. The intercept point for cascaded networks can be determined from:

$$IP_{2,\text{in}} = \frac{1}{\left(\frac{1}{\sqrt{IP1}} + \frac{G}{\sqrt{IP2}}\right)^2} \quad \text{(Eq 8)}$$

for IP_2 and from:

$$IP_{3,\text{in}} = \frac{1}{\frac{1}{IP1} + \frac{G}{IP2}} \quad \text{(Eq 9)}$$

for IP_3. In both equations, $IP1$ is the input intercept of Stage 1 in watts, $IP2$ is the input intercept of Stage 2 in watts and G is the gain of Stage 1 (as a numerical ratio, *not* in decibels). Both equations assume the worst-case condition, in which the distortion products of both stages add in-phase.

The ratio of the signal power to the intermodulation-product power, the *distortion ratio*, can be expressed as:

$$R_{dn} = (n-1)\left[IP_{n(\text{in})} - P_{(\text{in})}\right] \quad \text{(Eq 10)}$$

where n is the order, R_{dn} is the distortion ratio, $IP_{n(\text{in})}$ is the input intercept point and $P_{(\text{in})}$ is the input power of one tone.[2]

Discussions of IMD have traditionally downplayed the importance of IM_2. This is true because the incidental distributed filtering of the tuned circuitry, once common in radio communication systems, was usually enough to render out-of-passband IM_2 products caused by in-band signals and in-band IM_2 products caused by out-of-passband signals vanishingly weak compared to fundamental and IM_3 signals. In broadband systems that operate at bandwidths of an octave or more, however, in-passband signals may produce significantly strong in-passband IM_2 and second-harmonic products. In such applications, balanced circuit structures (such as push-pull amplifiers and balanced mixers) can be used to minimize IM_2 and other even-order nonlinear products.

As with IM_2, the importance of specific IM_3 products depends on the spacing of the signals involved and the relative width of the system passband. If ω_1 and ω_2 are of approximately the same frequency, the additive products

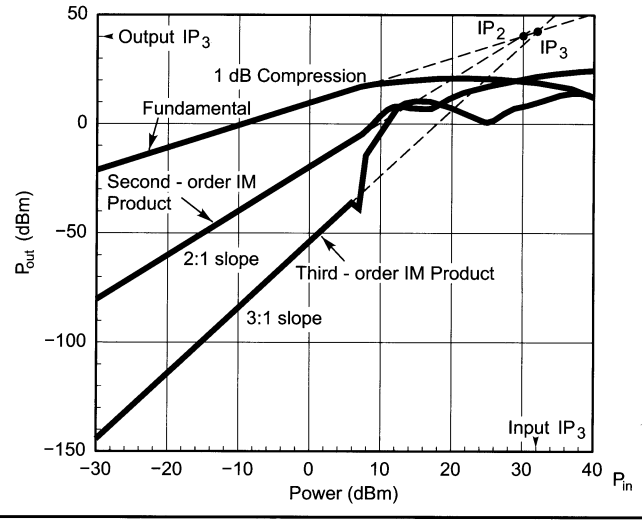


Fig 6—A simulation plot showing second and third-order IMD behavior for the one-stage amplifier shown in Fig 5.

$2\omega_1 + \omega_2$ and $2\omega_2 + \omega_1$ would be outside the passband of a narrow-bandwidth system. The subtractive products $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$, however, will likely appear near or within the system passband. The IM_3 performance of any network subjected to multiple signals is therefore of critical importance. Consequently, an array of IM_3-related, sometimes application-specific, figures of merit has evolved.

Example of a One-Stage FET Amplifier

The simulation of second- and third-order IMD products, based on the circuit shown in Fig 5, is shown in Fig 6. This gives a lot of insight into performance. The circuit in question is a simple single FET amplifier operating at about 120 mA dc. The top curve shows the fundamental and demonstrates 10-dB gain. At about 8 dBm of input, the 1-dB compression point occurs, and the fundamental amplitude actually gets smaller as the input power increases. The slope underneath is the fundamental referenced to the second-order IMD products. The crossover with the fundamental is called the second-order intercept point. A few decibels below the 1-dB compression point, the curve for the second-order IMD drastically deviates from the straight line; this is

due to clipping of harmonics. This curve has a 2:1 slope, meaning that with a 1 dB increase of the fundamental the products increase by 2 dB.

The final line starting at close to –150 dBm relates to the third-order IMD products. At about 7-dBm input, the IP_3 curve shows a dent. This is caused by the cancellation of the second-order harmonics, but then higher-level distortion is responsible for all kinds of deviation from linearity. Contrary to typical textbook plots, this curve shows a real, live amplifier performance.

Dynamic Range

As we have seen, thermal noise sets the lower limit of the power span over which a network can operate. Distortion—that is, degradation by distortion of the signal's ability to convey information—sets the upper limit of a network's power span. Because the power level at which distortion becomes intolerable varies with signal type and application, a generic definition has evolved: The upper limit of a network's power span is the level at which the power of one intermodulation product of a specified order is equal in power to the network's noise floor. The ratio of the noise-floor power to the upper-limit signal power is referred to as the network's *dynamic range (DR)*. This is often more carefully characterized as *two-tone IMD dynamic range*, which, when evaluated with equal-power test tones, is a figure of merit commonly used to characterize receivers. The *MDS* relative to the input, as already defined, is $MDS_{in} = kTB + 3\ dB + NF$.

When $IP_{(n)in}$ and *MDS* are known, *IMD DR* can be determined from:

$$DR_n = \frac{(n-1)[IP_{n(in)} - MDS_{in}]}{n} \qquad \text{(Eq 11)}$$

where *DR* is the dynamic range in decibels, *n* is the order, $IP_{(in)}$ is the input intercept power in dBm and *MDS* is the minimum detectable signal power in dBm. The so-called *spurious-free dynamic range* (*SFDR* or DR_{SF}) is calculated from:

$$DR_{SF} = \frac{2}{3}(IP_3 - 174\ \text{dBm} + NF + 3\ \text{dB}) \qquad \text{(Eq 12)}$$

This equation allows us to determine how to measure the spurious-free dynamic range. This is done by applying the two-tone signals (as in the case of IP_3) and increasing the two signals to the point where the signal-to-noise ratio deteriorates by 3 dB; or, if the measurement were done relative to *MDS*, where the noise floor rises by 3 dB. The "2/3" factor is derived from the fact that the levels of IM_3 outputs increase 3 dB for 1 dB of input increase. This definition of dynamic range now is referenced to a noise figure rather than a minimum level (in dBm) and is therefore independent of bandwidth. By choosing smaller bandwidths (1 kHz instead of 10 kHz), a dynamic-range measurement can be made to look better. Basing the specification on noise figure directly avoids this problem.

Dynamic Measure: A New Receiver Figure of Merit

Taken by themselves, the IP_n and *DR* figures of merit can be misleading in the sense that broadband, resistive attenuation directly improves them by the amount of attenuation added while directly degrading *NF* by the same amount. The *NF* and *MDS* figures of merit can mislead us because they convey no sense of suitability to task. This can be significant when we interconnect networks and systems having different noise floors and IMD dynamic ranges—as we do, for example, when connecting a receiver to an antenna. Relative to noise figure and intercept point, a noise measure and a third-order-intercept (IP_3) measure have already been defined. Both these measures refer to a "minimum value." In the case of noise, it addresses the minimum noise figure obtainable in a multistage arrangement. The IP_3 measure addresses the total resulting third-order products that result in a minimum value slightly less than that of the first stage, assuming that all stages are identical. As a means of better evaluating a receiver's front-end performance, I propose a new figure of merit, *dynamic measure*, defined as:

$$DM = \left[(IP_{3,in} + Att) - (NF_r + Att + NF_{ant})\right]\frac{NF_r}{NF_r + Att} \qquad \text{(Eq 13)}$$

where *DM* is the dynamic measure, a dimensionless number; $IP_{3,in}$ is the receiver's third-order input intercept in dBμV (instead of dBm) without any added attenuation; NF_r is the receiver noise figure in decibels without any added attenuation; NF_{ant} is the antenna-system noise figure in decibels; and *Att* is the decibel value of any added attenuation, which is sometimes used to shift the upper and lower limits of a system's dynamic range to higher absolute signal levels, increasing $IP_{3,in}$ at the expense of making the system more "deaf." Our dynamic measure figure of merit considers this trade-off to allow more-straightforward comparisons of systems that have the same dynamic range but different sensitivities. Setting NF_{ant} equal to zero allows evaluation of receiver performance under laboratory conditions. Setting NF_{ant} equal to the measured or expected *NF* of the antenna system to which the receiver will be connected allows evaluation of receiver performance in a system sense. Rather than express IP_3 in dBm, we are going to express it in dBμV. This is necessary because for negative values of the difference between IP_3 and the added noise figure, the multiplier term will make the result larger when its numerical value is less than one.

The usefulness of the dynamic measure figure of merit is evident when we consider several examples. Example 1 evaluates a receiver that exhibits the characteristics $IP_{3,in}$ = 16 dBm (123 dBμV), NF_r = 8 dB and NF_{ant} = 10 dB. We calculate its laboratory DM by setting NF_{ant} equal to 0:

$$DM = [123 - (8+0)]\frac{8}{8+0} = [115]1 = 115 \qquad \text{(Eq 14)}$$

Setting NF_{ant} equal to 10 returns the receiver's system DM:

$$DM = [123 - (8+10)]\frac{8}{8+0} = [105]1 = 105 \qquad \text{(Eq 15)}$$

Example 2 consists of the Example 1 receiver with a 10-dB pad switched in, resulting in the characteristics $IP_{3,in}$ = 26 dBm, NF_r = 18 dB and NF_{ant} = 10 dB. Setting NF_{ant} equal to 0, we calculate its laboratory DM as:

$$DM = [133 - (8+10+0)]\frac{8}{8+10} = [115]0.4444 = 51.1 \qquad \text{(Eq 16)}$$

A comparison with Eq 14 shows that although

Table 1—Eight Dynamic Measure Cases Compared

Ex.	System	Base $IP_{3,in}$ (dBm)	Base $IP_{3,in}$ (dBμV)	Atten (dB)	Res. $IP_{3,in}$ (dBm)	Res. $IP_{3,in}$ (dBμV)	Base NF_R (dB)	Res. NF_R (dB)	NF_{ant} (dB)	DM_{lab}	DM_{system}
1	HF receiver	16.0	123.0	0.0	16.0	123.0	8.0	8.0	10.0	115.0	105.0
2	HF receiver with 10-dB pad	16.0	123.0	10.0	26.0	133.0	8.0	18.0	10.0	51.1	46.7
3	HF receiver with 2-dB pad	16.0	123.0	2.0	18.0	125.0	8.0	10.0	10.0	92.0	84.0
4	Rohde & Schwarz XK 2100 shortwave transceiver	40.0	147.0	0.0	40.0	147.0	10.0	10.0	10.0	137.0	127.0
5	XK 2100 with 10 dB of unnecessary attenuation	40.0	147.0	10.0	50.0	157.0	10.0	20.0	10.0	68.5	63.5
6	Satellite receiver	−2.2	104.8	0.0	−2.2	104.8	0.9	0.9	0.2	103.9	103.7
7	Wireless front end without duplexer/filter losses	−12.0	95.0	0.0	−12.0	95.0	2.6	2.6	2.0	92.4	90.4
8	Wireless front end with 3 dB duplexer/filter losses	−12.0	95.0	3.0	−9.0	98.0	2.6	5.6	2.0	42.9	42.0

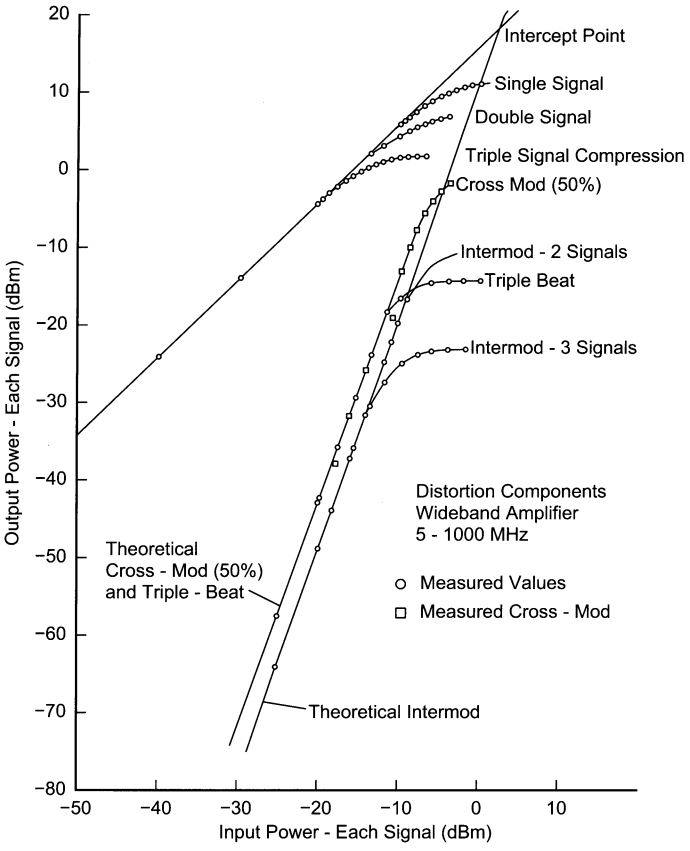


Fig 7—Measured distortion components in a wide-bandwidth (5 to 1000 MHz) amplifier.

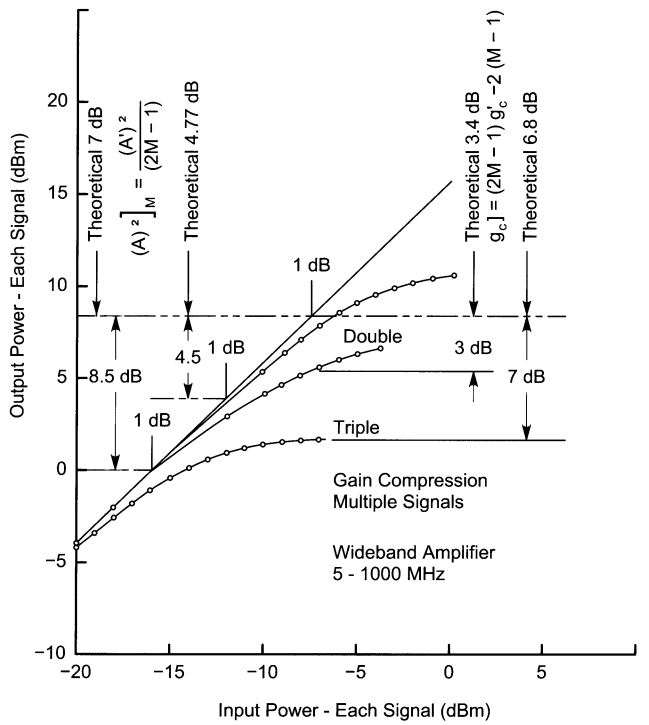


Fig 8—Measured multiple-signal gain compression of the 5- to 1000-MHz amplifier.

inserting the 10-dB pad has shifted $IP_{3,in}$ from 16 dBm (123 dBμV) to 26 dBm (133 dBμV), it has also increased the resulting *NF* from 8 to 18 dB, reducing the DM from 115 to 51.1. Inserting the pad also degrades the receiver's system DM compared to Example 1, as we see by setting NF_{ant} equal to 10:

$$DM = [133 - (8 + 10 + 10)]\frac{8}{8+10} = [105]0.4444 = 46.7 \qquad \text{(Eq 17)}$$

Example 3: On the other hand, had we inserted 2 dB of attenuation rather than 10 dB ($IP_{3,in}$ = 18 dBm [125 dBμV], NF_r = 10 dB and NF_{ant} = 10 dB), the laboratory DM would have been:

$$DM = [125 - (8 + 2 + 0)]\frac{8}{8+2} = [115]0.8 = 92 \qquad \text{(Eq 18)}$$

and the system DM would have been:

$$DM = [125 - (8 + 2 + 10)]\frac{8}{8+2} = [105]0.8 = 84 \qquad \text{(Eq 19)}$$

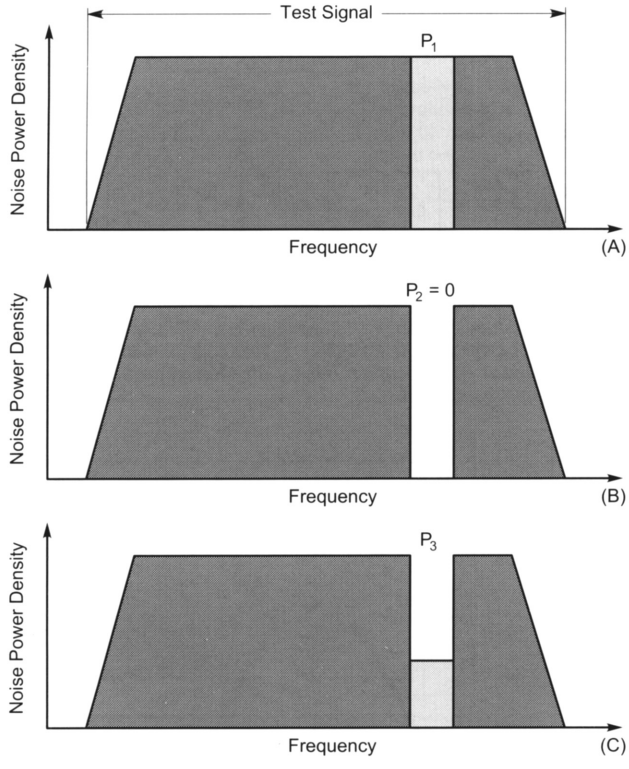


Fig 9—Determining a network's *noise-power ratio (NPR)* involves the application of a test signal consisting of thermal noise. The reference measurement-channel noise power, P_1, is then measured (at A). Next, a stop-band filter is placed between the noise generator and network under test to keep the test signal out of the measurement channel (at B). Assuming sufficient filter attenuation, if the network were absolutely noiseless and linear, the ideal noise power in the measurement channel, P_2, would then be zero. In practice, the network's own thermal noise and intermodulation between noise components outside the measurement channel result in an actual measurement-channel noise power (P_3) greater than zero. The noise-power ratio equals P_1/P_3.

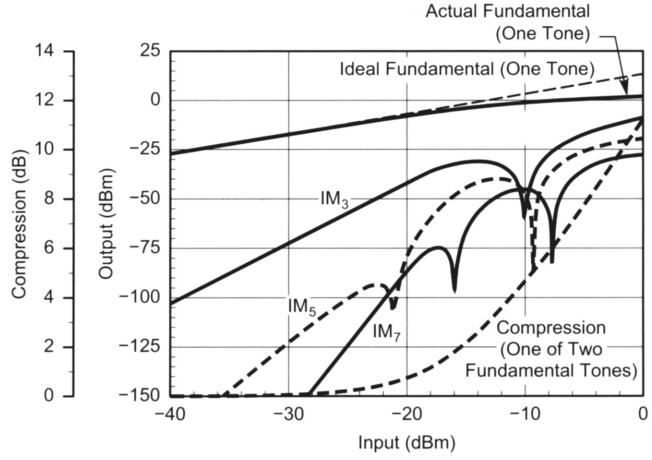


Fig 10—As a network is driven into compression, intermodulation products at odd orders higher than three become significant, and phase shifts in power-dependent device capacitances cause curves and dips in the intermodulation characteristics. The onset of these departures from intermodulation-response linearity occurs at generally lower input-power levels for higher intermodulation orders. Their severity and their positions on the intermodulation curves differ among the various products of a given order and varies with network topology and tone spacing. Figures of merit based on straight-line intermodulation responses fail to usefully predict nonlinear network behavior under these conditions. This graph shows the simulated performance of a single-BJT broadband amplifier driven by two equal-amplitude tones at 10 and 11 MHz.

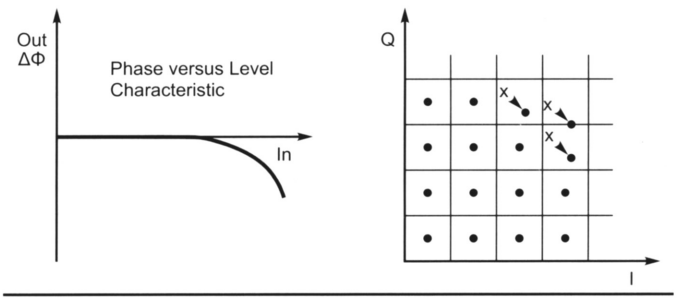


Fig 12—The influence of differential phase error (AM-to-PM conversion) on a QAM constellation.

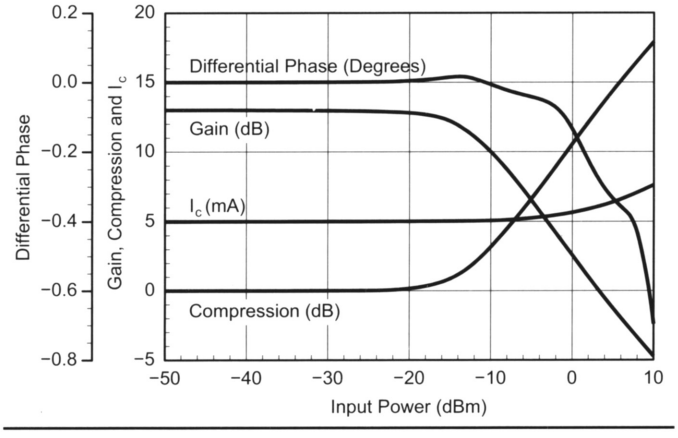


Fig 11—Driving a network into compression and saturation shifts the bias point(s) of its active device(s), changing their drive-dependent reactances and shifting the phase of the output signal relative to its value at input levels below compression. This graph shows the simulated performance of a single-BJT broadband amplifier driven by a single tone at 10 MHz.

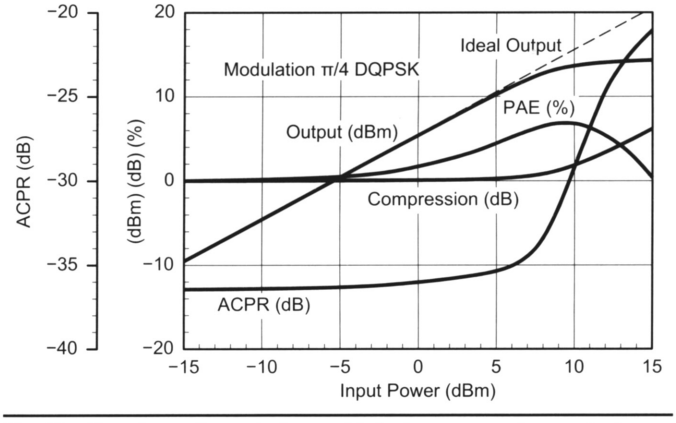


Fig 13—Keeping adjacent-channel interference under control can involve a critical trade-off between a wireless transmitter's power-added-efficiency (PAE) and ACPR, as shown in this graph showing the simulated behavior of a 1-GHz MESFET power amplifier. As the amplifier is driven into compression, a peak occurs in the PAE response—in this case, at P_{-1dB}—and ACPR rises sharply.

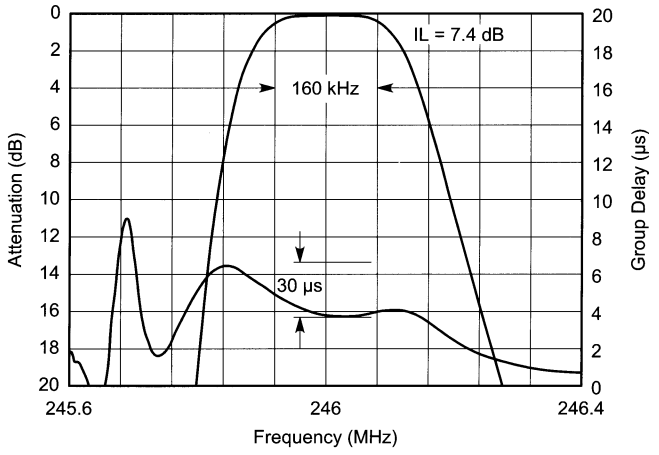


Fig 14—Close-in amplitude and group-delay responses for a 246-MHz SAW filter designed for GSM applications. This filter is well within its 3.0 µs differential-group-delay specification across its passband (160 kHz at –3 dB); the peaks just outside the passband limits are characteristic of a network's transition-band phase response.

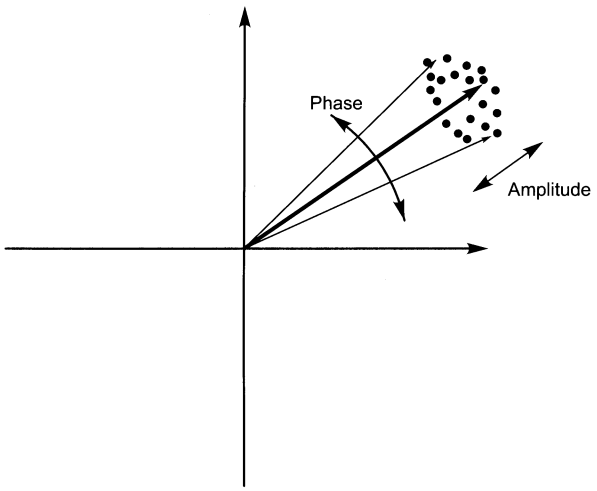


Fig 16—Oscillator noise can be split into amplitude and phase components.

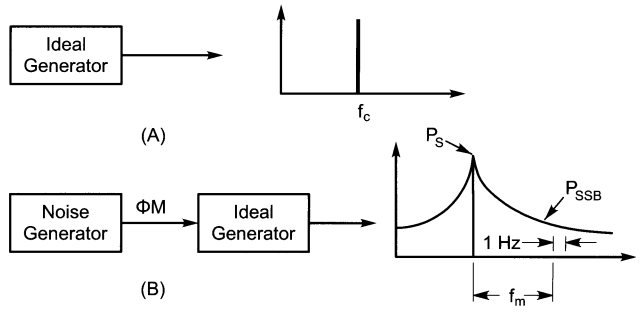
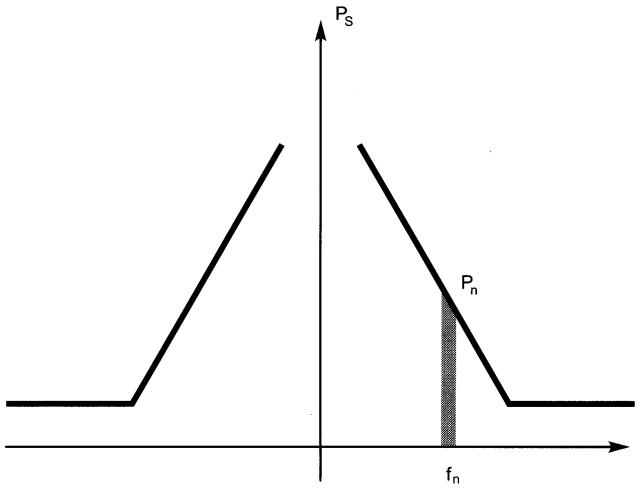


Fig 15—SSB phase noise. An ideal signal generator (A) would produce an absolutely pure carrier. A real signal generator (B), acts like an ideal generator driven by a noise generator, producing a noise-modulated carrier.



$$L(f) = 10\log\frac{P_n}{P_S}$$

P_n = Sideband noise in 1 - Hz bandwidth at offset frequency f_n

P_S = Total signal power

Single - sideband phase noise is expressed in dBc measured in a bandwidth of 1 Hz [dBc(Hz)] at offset frequency f_n

Fig 17—Phase noise calculation.

Table 1 shows these and several additional examples for comparison.

The interested reader will notice the following: The Example 1 receiver (base *NF* 8 dB) was unnecessarily sensitive compared to the antenna *NF* of 10 dB, so adding a 2-dB pad increased the intercept point by 2 dB and matched the antenna noise floor. Yet, the equation rightfully indicates a loss of dynamic measure. The relationship between NF_{ant} and NF_r is not taken into consideration. In addition, we have decided to leave it to the reader to develop a formula that deals with the addition of a preamplifier rather than an attenuator. This will complicate the formula because the preamplifier gain and noise figure are not necessarily related.

Triple-Beat Distortion and Cross Modulation

P_{-1dB} is a single-tone figure of merit; blocking, intercept point and dynamic range evaluate two-tone behavior. For networks that must handle AM and composite (AM and angle modulation) signals very linearly, such as television transmitters and cable-TV distribution systems, a three-tone figure of merit called *triple-beat distortion* has gained acceptance. Signals at ω_1 and ω_2 (closely spaced) and ω_3 (positioned far away from ω_1 and ω_2) are applied to the network under test, at levels, frequencies and spacings that vary with the application. One triple-beat distortion figure of merit is the ratio, expressed in decibels, of the intermodulation product at $\omega_3 + (\omega_2 - \omega_1)$ to one of the network's linear outputs at a specified output level. Alternatively, the triple-beat figure of merit may express the network output level at which a specified triple-beat ratio occurs.

Triple-beat distortion is the mechanism underlying cross-

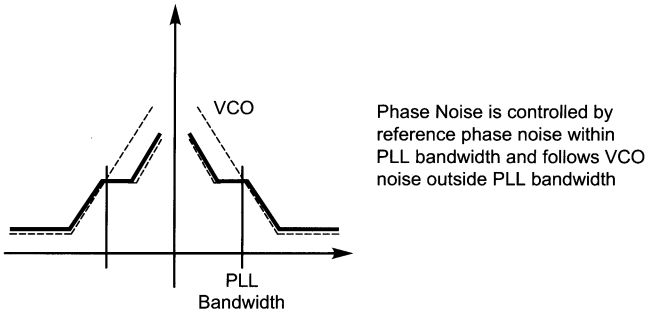


Fig 18—Phase noise of an oscillator controlled by a phase-locked loop.

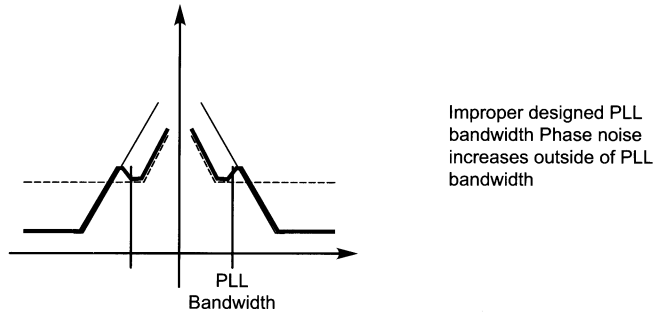


Fig 19—The effect of improper loop-filter design.

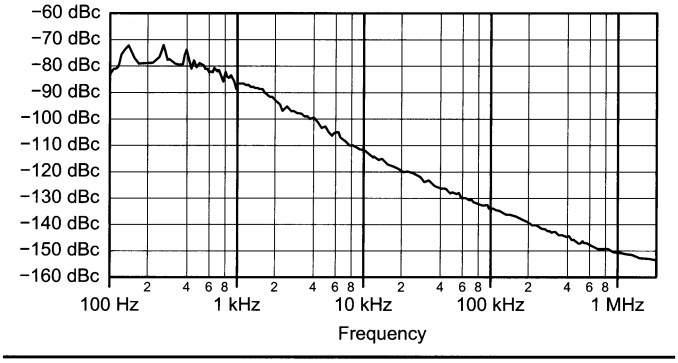


Fig 20—Measured phase noise of the Rohde & Schwarz SMY signal generator at 1 GHz. This signal generator has no provision for digital modulation and therefore shows the best possible phase noise in its class.

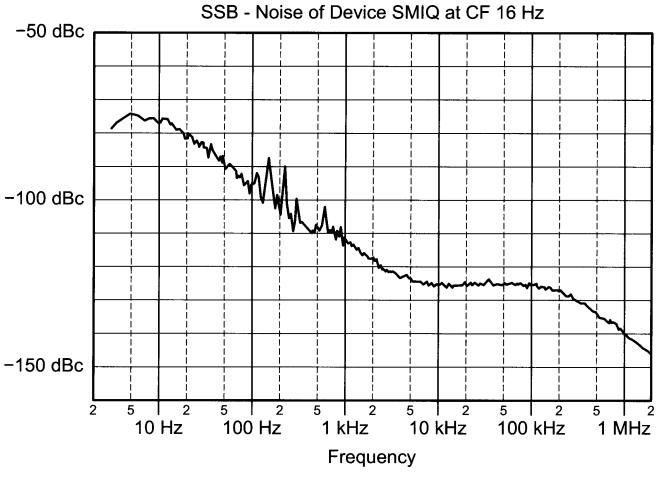


Fig 21—Measured phase noise of the Rohde & Schwarz SMIQ signal generator at 1 GHz. This signal generator is optimized for all digital modulation capabilities and can be configured via appropriate programming. Above 10 kHz, the influence of the wide-bandwidth loop becomes noticeable; above 200 kHz, the resonator *Q* takes over.

modulation—a form of intermodulation—in which one or more AM signals present in a network amplitude-modulate all signals present in the network. Angle-modulation-based wireless systems are largely immune to such effects. Figs 7 and 8 graph the results of gain compression, two-tone intermodulation, cross-modulation and triple-beat testing on a wideband (5 to 1000 MHz) amplifier.

Noise Power Ratio

Triple-beat testing is one way of improving on two-tone testing as a means of evaluating a network's intermodulation behavior in the presence of multiple signals. Another figure of merit—*noise power ratio (NPR)*—uses thermal noise as a test signal. The test measures the introduction, by intermodulation, of noise into a quiet slot created by the insertion of a band-stop filter between the noise generator and the network under test (Fig 9). The filter stopband width is equal to the width of the measurement channel.

Large-Signal Effects

Except for P_{-1dB}, the figures of merit discussed so far evaluate amplitude nonlinearity under small-signal conditions. At input powers that drive a network into gain compression and saturation, intermodulation products of odd orders higher than three become significant. Curves, dips and nulls appear in their characteristics (Fig 10). Phase shifts related to nonlinear (primarily voltage-dependent) capacitances in solid-state devices are one cause of these effects. Under such conditions a network may exhibit hysteresis, with its behavior at any given instant depending not only on the voltage or current applied to it, but also on its recent history.

AM-to-PM Conversion

The nonlinear distortion effects we've discussed so far can be termed *AM-to-AM distortion*. Such distortion, to a degree, depends on the amplitude of the signal(s) applied to the network and results in changes in the network's gain and/or production of signals at new frequencies. AM-to-PM distortion can also occur. As a network nears saturation, part of the driving signal shifts the bias point(s) of the active device(s). This changes their drive-dependent reactances and shifts the phase of the output signal relative to its value at input levels below compression (Figs 11 and 12). This effect, *AM-to-PM conversion*, can cause incidental phase modulation that degrades the performance of digital communication systems.

Spectral Regrowth and Adjacent-Channel Power Ratio

Spectral regrowth occurs largely because of third-, fifth- and seventh-order IMD in power amplifiers operated near or in compression. That is, at power levels where hysteretic intermodulation effects result in poor agreement between measured behavior and predictions based on

small-signal intermodulation figures of merit. We therefore evaluate the impact of spectral regrowth more directly, using a figure of merit called *adjacent-channel power ratio* (ACPR). ACPR measurement techniques that incorporate memory can be used to increase ACPR predictions for networks that exhibit saturation hysteresis. Fig 13 shows the critical relationship between compression, power-added efficiency and ACPR in a MESFET power amplifier.

Phase Response Issues and Figures of Merit

We have already seen how large-signal nonlinear distortion can result in amplitude-dependent phase shifts through AM-to-PM conversion. Because phase linearity is critical at *all* signal levels in PM systems, especially those using digital modulation, we must also consider linear distortion in evaluating networks used in wireless systems.

Differential Group Delay

Every frequency-selective network subjects signals passing through it to some degree of time delay. Ideally, this delay (also known as *group* or *envelope delay*) does not vary with frequency. That is, the network's phase-shift versus frequency response is monotonic and linear. In practice, a network's time delay varies across its passband, transition bands and stopbands, exhibiting curvature, ripple and transition-band peaks (Fig 14). The network's *differential* group delay—its group-delay spread—is therefore of considerable importance. This is especially so in digitally modulated systems, where the resulting phase distortion can cause errors in modulation and demodulation.

Effects of Phase Noise

The phase of an oscillator's output signal is subject to random phase variations (Figs 15 and 16). Called *phase noise*, this effect is often quantified as the decibel ratio of the phase-noise power in a single phase-noise sideband (a 1-Hz bandwidth centered at a specified frequency offset from the oscillator carrier) to the carrier power (Fig 17). Alternatively, it may be specified in degrees RMS. A microwave voltage-controlled oscillator, for instance, might exhibit an SSB phase noise of –95 dBc/Hz at 10 kHz. Oscillator phase noise may manifest itself, through a mechanism known as *reciprocal mixing*, as the emission of unacceptably strong noise outside a transmitter's occupied bandwidth or as an increase in receiver noise floor. Phase noise may also directly introduce phase errors that cause modulation and demodulation errors.

Because the oscillators used for frequency translation in wireless systems are usually embedded in phase-locked loops, their phase-noise characteristics differ from those of "bare" oscillators, as shown in Figs 18 and 19. Figs 20 and 21 show the measured phase noise of the Rohde & Schwarz SMY and SMIQ signal generators. The SMY is a low-cost signal source, while the SMIQ is a very high performance signal generator, which can be programmed for all digital modulations. Therefore, their PLL systems exhibit different phase-noise-versus-frequency responses as the measured results show.

Reciprocal Mixing

In reciprocal mixing, incoming signals mix with LO-sideband energy to produce an IF output (Fig 22). Because one of the two signals is usually noise, the resulting IF output is usually noise. Reciprocal-mixing effects are not limited to noise. Discrete-frequency oscillator sideband components, such as those resulting from crosstalk to, or reference energy on, a VCO control line, or the discrete-frequency spurious signals endemic to direct digital synthesis, can also mix incoming signals to IF. In practice, the resulting noise-floor increase can compromise the receiver's ability to detect weak signals and achieve a high IMD dynamic range. On the test bench, noise from reciprocal mixing may invalidate desensitization,

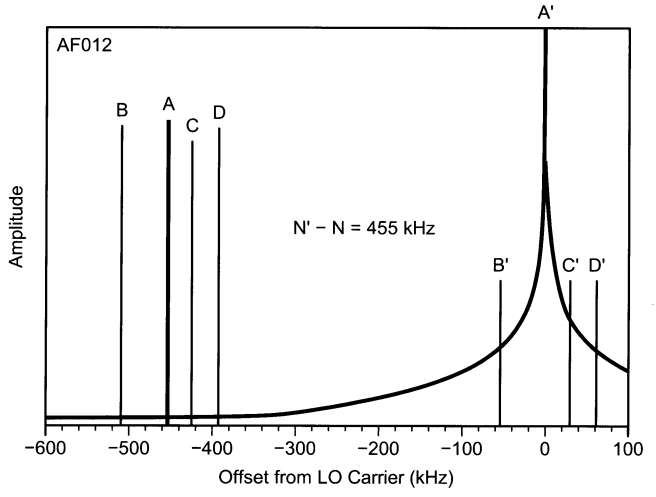


Fig 22—Reciprocal mixing occurs when incoming signals mix energy from an oscillator's sidebands to the IF. In this example, the oscillator is tuned so that its carrier, at A′, heterodynes with the desired signal, A, to the 455-kHz IF as intended. At the same time, the undesired signals B, C and D mix the oscillator noise-sideband energy at B′, C′ and D′, respectively, to the IF. Depending on the levels of the interfering signals and the noise-sideband energy, the result may be a significant rise in the receiver noise floor.

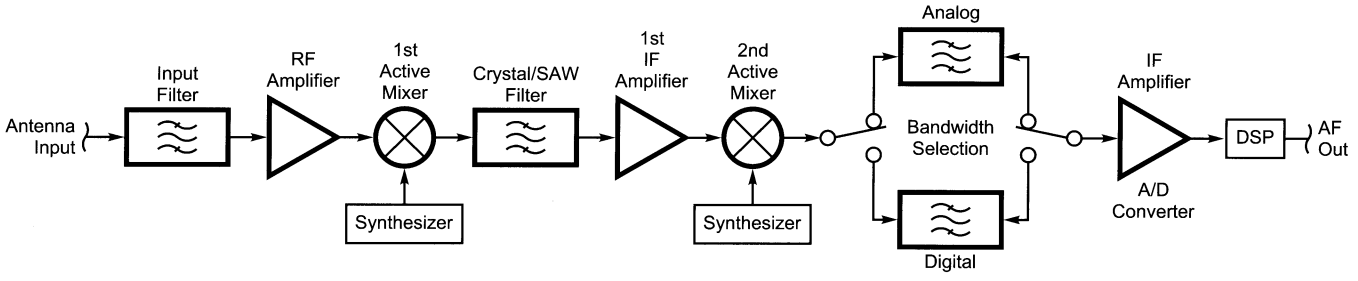


Fig 23—Block diagram of an analog/digital receiver showing the signal path from antenna to audio output. No AGC or other auxiliary circuits are shown. This receiver principle can be used for all types of modulation, since the demodulation is done in the DSP block.

cross-modulation and intermodulation testing by obscuring the weak signals that must be measured in these tests.

Fig 23 shows a typical arrangement of a dual-conversion receiver with local oscillators. The signal coming from the antenna is filtered by an arrangement of tuned circuits referred to as providing *input selectivity*. For a minimum attenuation in the passband, an operating bandwidth must be

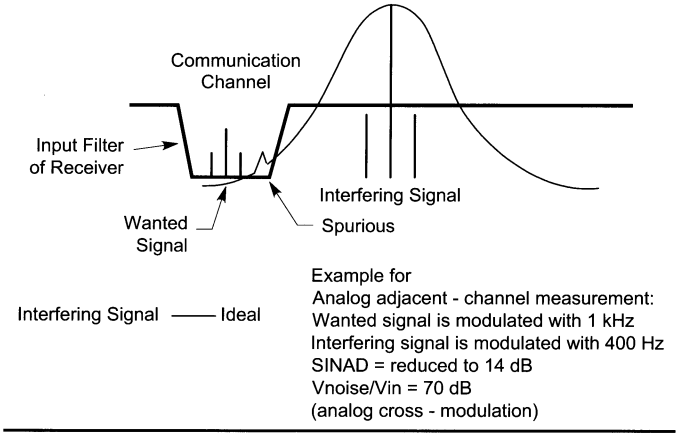
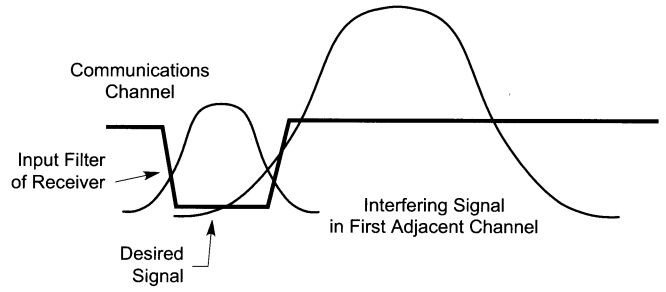


Fig 24—The principle of selectivity-measurement for analog receivers.



ACP Suppression:

NADC	Requirement:
30 kHz from Carrier	<–26 dBc
60 kHz from Carrier	<–45 dBc
90 kHz from Carrier	<–49 dBc
PDC	
50 kHz from Carrier	<–45 dBc
100 kHz from Carrier	<–60 dBc

Fig 25—(right) The principle of selectivity-measurement for digital receivers.

Table 2—Level Diagram of a 2.4-GHz Receiver

Calculation of IP_3

1. Module1: $IP_3=\infty$
2. Module2: IP_3=23.2 dBm = 209 mW
3. Module3: $IP_3=\infty$
4. Module4: IP_3=31–(–0.2 + 3.0 – 0.2) = 31.4 dBm = 1380.384 mW
5. Module5: $IP_3=\infty$
6. Module6: IP_3=140 –(–0.2 + 3.0 – 0.2 – 6.0 – 0.2)= 143.6 dBm
   Input-IP_3 = 2.29086E14 mW

$$IP_{3(2.4\text{GHz\_Receiver})} = \frac{1}{\sum_{i=1}^{i=6}\left(\frac{1}{IP_3(\text{module})}\right)}$$

$$= \frac{1}{\frac{1}{\infty}+\frac{1}{209}+\frac{1}{\infty}+\frac{1}{1380.384}+\frac{1}{\infty}+\frac{1}{2.29086E14}} = \frac{1}{0.0055091} = 181.51$$

$$= 181.51 \text{ mW} = 22.58 \text{dBm}$$

Calculation of NF

1. Module1: NF=0.2 dB: F=1.04712
2. Module2: NF=1 dB: F=1.2589
3. Module3: NF=0.2 dB: F=1.04712
4. Module4: NF=6 dB: F=3.98107
5. Module5: NF=0.2 dB: F=1.04712
6. Module6: NF= 2 dB: F=1.58489

$$F_{\text{Total}} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \frac{F_4 - 1}{G_1 G_2 G_3} + \ldots$$

$$F = 1.04712 + 0.261 + 0.0001205 + 0.06 + 0.01 + 0.054 = 1.4322$$

$$NF = 1.56 \text{ dB}$$

$$BW = \frac{f}{\sqrt{2} \cdot Q_L} \quad \text{(Eq 20)}$$

This approximation formula is valid for the insertion loss of about 1 dB due to loaded Q.

The filter in the first IF is typically either a SAW filter (from 500 MHz to 1 GHz) or a crystal filter (from 45 to 120 MHz) with a typical insertion loss of 6 dB. Since these resonators have significantly higher Qs than LC circuits, the bandwidth for the first IF will vary from ±5 kHz to ±500 kHz. It is now obvious that the first RF filter does not protect the first IF because of its wider bandwidth. For typical communication receivers, IF bandwidths from 150 Hz to 1 MHz are found. For digital modulation, the bandwidth varies roughly from 30 kHz to 1 MHz. Therefore, the IF filter in the second IF must accommodate these bandwidths; other-

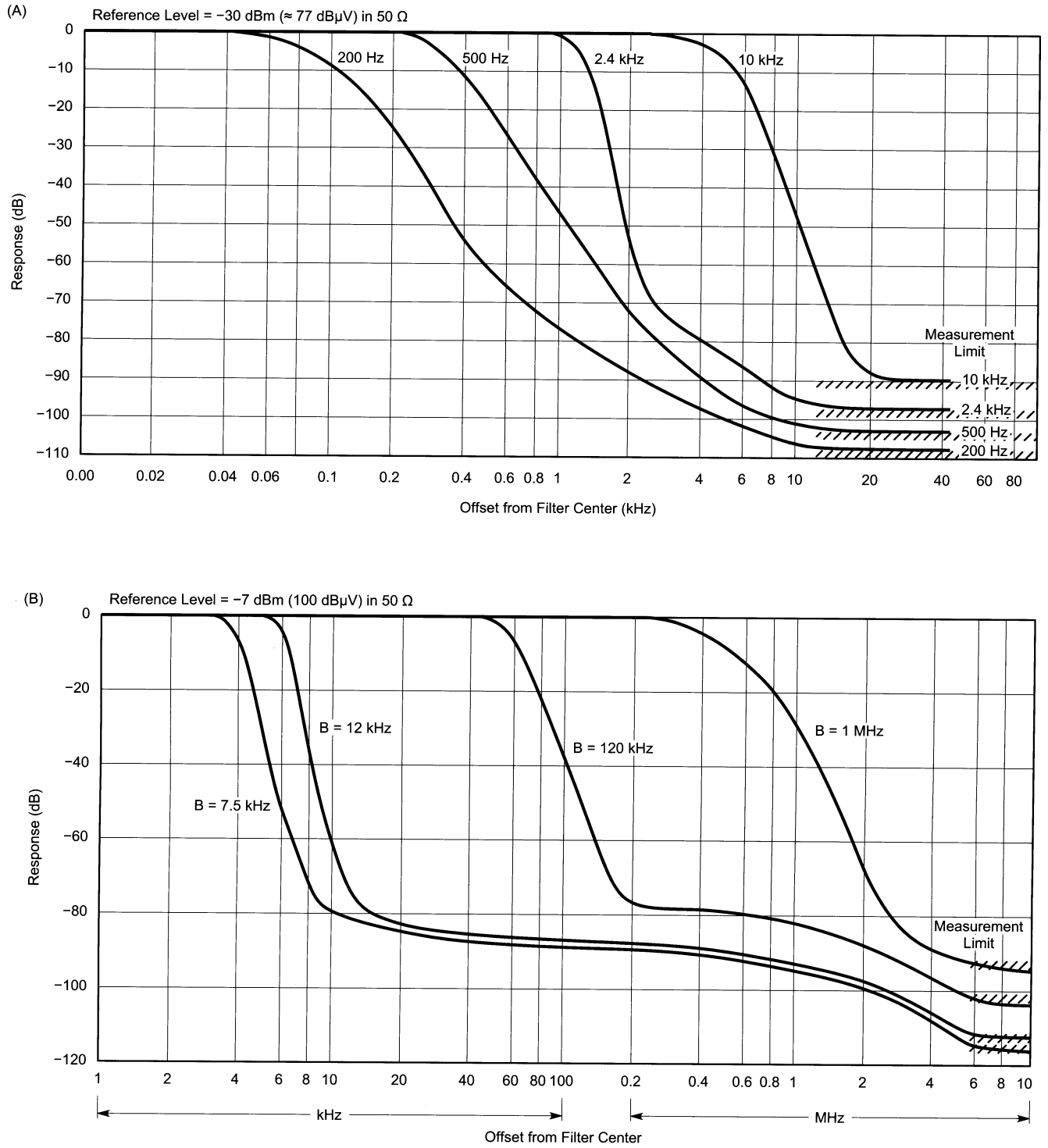
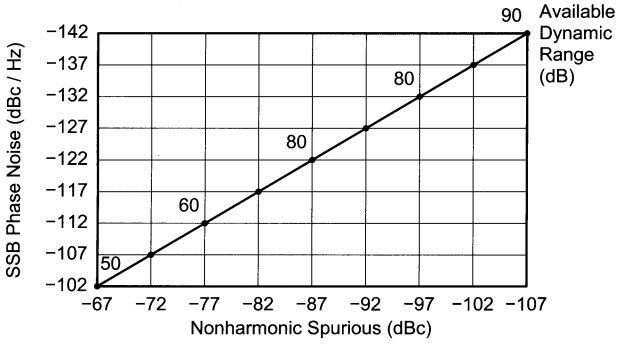


Fig 26—Dynamic selectivity versus IF bandwidth for (A) the Rohde & Schwarz ESH-2 test receiver (9 kHz to 30 MHz) and (B) the Rohde and Schwarz ESV test receiver (10 MHz to 1 GHz). Reciprocal mixing widens the ESH-2's 2.4 kHz response below −70 dB (−100 dBm) at (A) and the ESV's 7.5, 12 and 120 kHz responses below approximately −80 dB (−87 dBm) at (B).

wise, the second mixer is easily overloaded. This also means that both synthesizer paths must be designed for minimum noise and spurious emissions. The second IF of this arrangement (Fig 23) can be either analog or digital, or it may even be a zero-IF. There are good reasons for using IFs at 50/3 kHz (like the Walkman), with DSP using low-cost modules found in mass-market consumer products.

The following two pictures (Figs 24 and 25) show the principle of selectivity measurement for both analog and digital signals. The main difference is that the occupied bandwidth for the digital system can be significantly wider, and yet both signals can be interfered with either by a noise synthesizer/first LO or a synthesizer that has unwanted spurious frequencies. Such a spurious signal is shown in Fig 24. In the case of Fig 24, the analog adjacent-channel measurement has some of the characteristics of cross-modulation and intermodulation, while in the digital system the problem with the adjacent-channel power suppression in modern terms is more obvious. Rarely has the concept of adjacent-channel power (ACP) been used with analog systems. Also, to meet the standards, requires signal generators that are better, with some headroom, than the required dynamic measurement. We have therefore included in Fig 25 the achievable performance for a practical signal generator—in this case, the Rohde & Schwarz SMHU58.



6 dB Reserve for Errors Less than 1 dB
Carrier - to - Spurious Ratio for 14 dB SINAD = 11 dB
Carrier - to - Noise for 14 dB SINAD = 46 dB

Fig 27—This graph shows the available dynamic range, which is determined by either masking of the unwanted signal by phase noise or by discrete spurii. As far as the culprit synthesizer is concerned, it can be either the local oscillator or the effect of a strong adjacent-channel signal that takes over the function of the local oscillator.

Because reciprocal mixing produces the effect of noise leakage around IF filtering, it plays a role in determining a receiver's dynamic selectivity (Fig 26). There is little value in using IF filters with stopband rejection more than 3 to 10 dB beyond that which results in an acceptable reciprocal mixing level. Although additional RF selectivity can reduce the number of signals that contribute to the noise, improving the LO's spectral purity is the only effective way to reduce reciprocal mixing noise from *all* signals present at a mixer's RF port.

Factoring in the effect of discrete spurious signals with that of oscillator phase noise can give us the useful dynamic range of which an instrument or receiver is capable (Fig 27).

An Example of a Receiver System

Fig 28 shows a level diagram of a high-performance receiver for analyzing the noise and gain distribution. This receiver consists of a band-pass filter centered at 2.4 GHz, an amplifier stage marked "two" (for two-port) with its electrical characteristics attached and another band-pass filter. The signal then is applied to a high-level mixer with a third-order intercept point of 25 dBm. The down-converter signal is filtered in a band-pass filter and applied to a broadband, high-gain amplifier chain. This is a typical front end, as we would expect it at a base station for CDMA applications.

The purpose of this block diagram is to show how the calculation of noise figure and intercept point is done properly. By applying the equations discussed earlier, the complete calculation is shown in Table 2 for the 2.4-GHz receiver. Here are the design steps:

1. Transfer all input intercept points to the system input, subtracting gains and adding losses decibel-for-decibel.
2. Convert intercept points to powers (dBm to milliwatts). We have *IP1*, *IP2*. . ., *IPN* for *N* elements.
3. Assuming all input intercept points are independent and uncorrelated, add powers in "parallel."
4. Convert *IP3* input from power (milliwatts) to dBm.

Notes

[1]U. Rohde and D. Newkirk, *RF/Microwave Circuit Design for Wireless Applications*, (Indianapolis, Indiana: John Wiley & Sons, April 2000) ISBN 0-471-29818-2; **customer@wiley.com**.

[2]P. Vizmuller, *RF Design Guide—Systems, Circuits and Equations*, (Norwood, Massachusetts: Artech House, 1995) ISBN 0-89006-754-6.

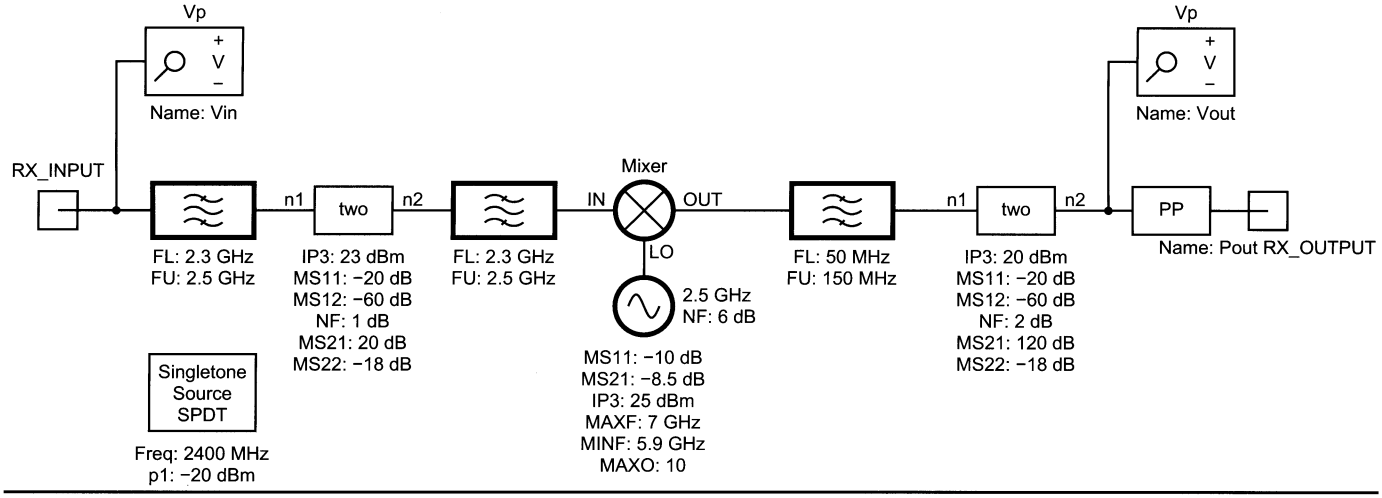


Fig 28—A level diagram of a high-performance 2.4-GHz receiver front-end. □□