provide both higher levels of integration and lower overall cost, as demonstrated in complex circuits such as frequency synthesizers. In fact, all building blocks of typical transceivers are available in silicon bipolar technologies from many manufacturers.

The third contender is CMOS technology. Supported by the enormous momentum of the digital market, CMOS devices have achieved high transit frequencies, e.g., tens of gigahertz in the 0.35-μm generation. As we will see in this book, "RF CMOS" has suddenly become the topic of active research. CMOS technology must nevertheless resolve a number of practical issues: substrate coupling of signals that differ in amplitude by 100 dB, parameter variation with temperature and process, and device modeling for RF operation.

REFERENCES

[1] C. Marshall et al., "2.7 V GSM Transceiver ICs With On-Chip Filtering," *ISSCC Dig. Tech. Papers,* pp. 148–149, February 1995.

[2] B. Sklar, *Digital Communications*, Englewood Cliffs, NJ: Prentice Hall, 1988.

[3] K. Feher, *Wireless Digital Communications,* Englewood Cliffs, NJ: Prentice Hall, 1995.

[4] R. Steele, ed. *Mobile Radio Communications*, Piscataway, NJ: IEEE Press, 1992.

[5] R. Schneiderman, "GaAs Continues to Gain in Wireless Applications," *Wireless Systems Design,* pp. 14–16, March 1997.

2

BASIC CONCEPTS IN RF DESIGN

RF designers draw upon many concepts that originate from the theory of signals and systems. In this chapter, we describe these concepts and define the terminology used in RF electronics so as to prepare the reader for the material in the following chapters.

Beginning with nonlinear systems, we describe effects such as harmonic distortion, gain compression, cross modulation, and intermodulation. We then briefly study intersymbol interference and Nyquist signaling, review random processes and noise, and introduce approaches to representing noise in circuits. Finally, we describe passive impedance transformation.

2.1 NONLINEARITY AND TIME VARIANCE

A system is linear if its output can be expressed as a linear combination (superposition) of responses to individual inputs. More accurately, if for inputs $x_1(t)$ and $x_2(t)$

$$x_1(t) \rightarrow y_1(t), \quad x_2(t) \rightarrow y_2(t), \tag{2.1}$$

where the arrow denotes the operation of the system, then

$$ax_1(t) + bx_2(t) \rightarrow ay_1(t) + by_2(t), \tag{2.2}$$

for all values of the constants a and b. Any system that does not satisfy this condition is nonlinear. Note that according to this definition, we consider a system nonlinear if it has nonzero initial conditions or finite "offsets."

A system is time invariant if a time shift in its input results in the same time shift in its output. That is, if $x(t) \to y(t)$, then $x(t-\tau) \to y(t-\tau)$, for all values of τ. A system is time variant if it does not satisfy this condition.

While nonlinearity and time variance are intuitively obvious concepts, they may be confused with *each other* in some cases. For example, consider the switching circuit shown in Fig. 2.1(a). The control terminal of the switch is driven by $v_{in1}(t) = A_1 \cos\omega_1 t$ and the input terminal by $v_{in2}(t) = A_2 \cos\omega_2 t$. We assume the switch is on if $v_{in1} > 0$ and off otherwise. Is this system nonlinear or time variant? If, as shown in Fig. 2.1(b), the path of interest is from v_{in1} to v_{out}, (while v_{in2} is part of the system and still equal to $A_2 \cos\omega_2 t$), then the system is nonlinear because the control is only sensitive to the polarity of v_{in1}, and time variant because v_{out} also depends on v_{in2}. On the other hand, if, as shown in Fig. 2.1(c), the path of interest is from v_{in2} to v_{out} (while v_{in1} is part of the system and still equal to $A_1 \cos\omega_1 t$), then the system is *linear* [Eq. (2.2)] but time variant. Thus, general statements such as "switches are nonlinear" are ambiguous. As we will see in Chapter 6, these distinctions are critical in the design of mixers.

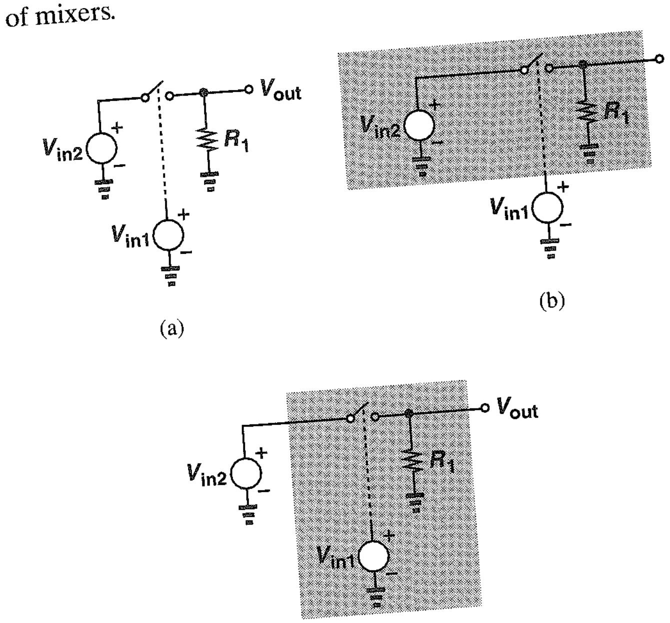


Figure 2.1 (a) Simple switching circuit, (b) nonlinear time-variant system, (c) linear time-variant system.

Another interesting result of the above observation is that a linear system *can* generate frequency components that do not exist in the input signal. This ... system is time variant, for example, Fig. 2.1(c). Since in this

circuit, v_{out} can be considered as the product of v_{in2} and a square wave toggling between 0 and 1, the output spectrum is given by

$$V_{out}(f) = V_{in2}(f) * \sum_{n=-\infty}^{+\infty} \frac{\sin(n\pi/2)}{n\pi} \delta\left(f - \frac{n}{T_1}\right) \tag{2.3}$$

$$= \sum_{n=-\infty}^{+\infty} \frac{\sin(n\pi/2)}{n\pi} V_{in2}\left(f - \frac{n}{T_1}\right), \tag{2.4}$$

where $*$ denotes convolution, $\delta(\cdot)$ is the Dirac delta function, and $T_1 = 2\pi/\omega_1$. Thus, the output consists of vertically scaled replicas of $V_{in2}(f)$ shifted by n/T_1.

A system is called "memoryless" if its output does not depend on the past values of its input. For a memoryless linear system,

$$y(t) = \alpha x(t), \tag{2.5}$$

where α is a function of time if the system is time variant [e.g., Fig. 2.1(c)]. For a memoryless nonlinear system, the input-output relationship can be approximated with a polynomial,

$$y(t) = \alpha_0 + \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) + \cdots, \tag{2.6}$$

where α_j are in general functions of time if the system is time variant. Fig. 2.2(a) illustrates an example where the input signal is applied to the base of Q_1 while Q_2 and Q_3 are periodically switched by means of a square wave. For ideal bipolar transistors, the circuit can be viewed as in Fig. 2.2(b) and

$$v_{out}(t) = \left(I_{S1} \exp\frac{v_{in}}{V_T}\right) s(t) \cdot R, \tag{2.7}$$

where I_{S1} represents the saturation current of Q_1, $V_T = kT/q$, and $s(t)$ is a square wave toggling between -1 and $+1$.

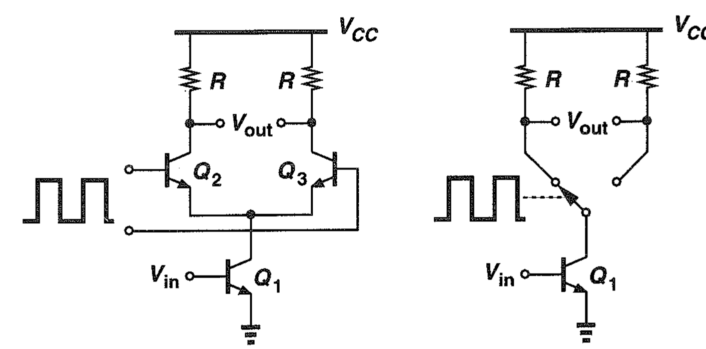
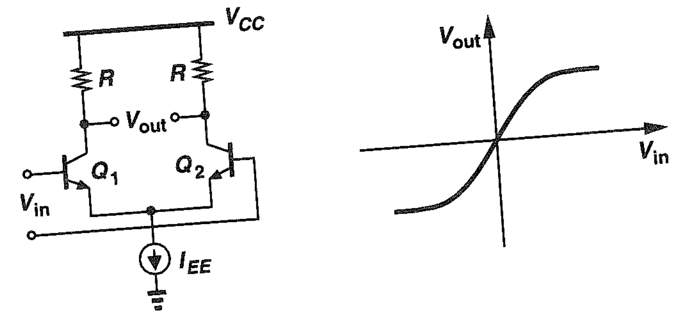


Figure 2.2 (a) Switching differential pair with tail current source driven by a signal, (b) equivalent circuit of (a).

The system described by (2.6) has "odd symmetry" if its response to $-x(t)$ is the negative of that to $x(t)$. This occurs if $\alpha_j = 0$ for even j. A circuit having odd symmetry is called differential or "balanced." For example, the bipolar differential pair of Fig. 2.3 exhibits the following input-output characteristic:

$$v_{out} = RI_{EE} \tanh \frac{v_{in}}{2V_T}, \tag{2.8}$$

which is an odd function.



(a) (b)

Figure 2.3 Bipolar differential pair along with its input-output characteristic.

A system is called "dynamic" if its output depends on the past values of its input(s) or output(s). For a linear, time-invariant, dynamic system,

$$y(t) = h(t) * x(t), \tag{2.9}$$

where $h(t)$ denotes the impulse response. If a dynamic system is linear but time variant, its impulse response depends on the time origin; if $\delta(t) \rightarrow h(t)$, then $\delta(t-\tau) \rightarrow h(t,\tau)$. Thus,

$$y(t) = h(t,\tau) * x(t). \tag{2.10}$$

Finally, if a system is both nonlinear and dynamic, then its impulse response can be approximated with a Volterra series [1, 2], a topic beyond the scope of this book.

2.1.1 Effects of Nonlinearity

While many analog and RF circuits can be approximated with a linear model to obtain their response to small signals, nonlinearities often lead to interesting and important phenomena. For simplicity, we limit our analysis to memoryless, time-variant systems and assume

$$y(t) \approx \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t). \tag{2.11}$$

The reader is cautioned, however, that the effect of storage elements and higher-order nonlinear terms must be carefully examined to ensure (2.11) is a plausible representation.

Harmonics If a sinusoid is applied to a nonlinear system, the output generally exhibits frequency components that are integer multiples of the input frequency. In Eq. (2.11), if $x(t) = A\cos\omega t$, then

$$y(t) = \alpha_1 A\cos\omega t + \alpha_2 A^2 \cos^2\omega t + \alpha_3 A^3 \cos^3 t \tag{2.12}$$

$$= \alpha_1 A\cos\omega t + \frac{\alpha_2 A^2}{2}(1+\cos 2\omega t) + \frac{\alpha_3 A^3}{4}(3\cos\omega t + \cos 3\omega t) \tag{2.13}$$

$$= \frac{\alpha_2 A^2}{2} + \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4}\right)\cos\omega t + \frac{\alpha_2 A^2}{2}\cos 2\omega t + \frac{\alpha_3 A^3}{4}\cos 3\omega t. \tag{2.14}$$

In Eq. (2.14), the term with the input frequency is called the "fundamental" and the higher-order terms the "harmonics."

From the above expansion, we can make two observations. First, even-order harmonics result from α_j with even j and vanish if the system has odd symmetry, i.e., if it is fully differential. In reality, however, mismatches corrupt the symmetry, yielding finite even-order harmonics. Second, in (2.14) the amplitude of the nth harmonic consists of a term proportional to A^n and other terms proportional to higher powers of A. Neglecting the latter for small A, we can assume the nth harmonic grows approximately in proportion to A^n.

Gain Compression The small-signal gain of a circuit is usually obtained with the assumption that harmonics are negligible. For example, if in (2.14), $\alpha_1 A$ is much greater than all the other factors that contain A, then the small-signal gain is equal to α_1. This quantity can be seen in the familiar differential pair of Fig. 2.3 to be equal to

$$\frac{v_{out}}{v_{in}} = \frac{I_{EE}R}{2V_T}. \tag{2.15}$$

However, as the signal amplitude increases, the gain begins to vary. In fact, nonlinearity can be viewed as variation of the small-signal gain with the input level. This is evident from the term $3\alpha_3 A^3/4$ added to $\alpha_1 A$ in (2.14), as well as the input-output characteristic shown in Fig. 2.3.

In most circuits of interest, the output is a "compressive" or "saturating" function of the input; that is, the gain approaches zero for sufficiently high input levels. In (2.14) this occurs if $\alpha_3 < 0$. Written as $\alpha_1 + 3\alpha_3 A^2/4$, the gain is therefore a decreasing function of A. In RF circuits, this effect is quantified by the "1-dB compression point," defined as the input signal level that causes the small-signal gain to drop by 1 dB. If plotted on a log-log scale as a function of the input level, the output level falls below its ideal value by 1 dB at the 1-dB compression point (Fig. 2.4).

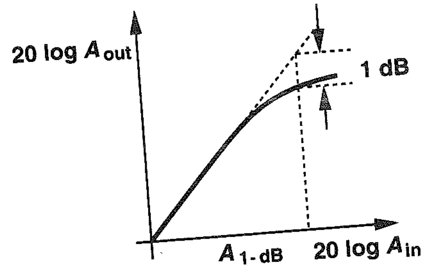


Figure 2.4 Definition of the 1-dB compression point.

To calculate the 1-dB compression point, we can write from (2.14)

$$20 \log \left|\alpha_1 + \frac{3}{4}\alpha_3 A_{1\text{-dB}}^2\right| = 20 \log |\alpha_1| - 1 \text{ dB}. \tag{2.16}$$

That is,

$$A_{1\text{-dB}} = \sqrt{0.145 \left|\frac{\alpha_1}{\alpha_3}\right|}. \tag{2.17}$$

A measure of the maximum input range of the circuit, the 1-dB compression point occurs around −20 to −25 dBm (63.2 to 35.6 mV_{pp} in a 50-Ω system) in typical front-end RF amplifiers.

Desensitization and Blocking Circuits with compressive characteristics exhibit an interesting effect when they process a weak, desired signal along with a strong interferer. Since a large signal tends to reduce the "average" gain of the circuit, the weak signal may experience a vanishingly small gain. Called "desensitization," this effect can be analyzed for the characteristics of (2.11) by assuming $x(t) = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t$. The output is

$$y(t) = \left(\alpha_1 A_1 + \frac{3}{4}\alpha_3 A_1^3 + \frac{3}{2}\alpha_3 A_1 A_2^2\right) \cos \omega_1 t + \cdots, \tag{2.18}$$

which, for $A_1 \ll A_2$, reduces to

$$y(t) = \left(\alpha_1 + \frac{3}{2}\alpha_3 A_2^2\right) A_1 \cos \omega_1 t + \cdots. \tag{2.19}$$

Thus, the gain for the desired signal is equal to $(\alpha_1 + 3\alpha_3 A_2^2/2)$, a decreasing function of A_2 if $\alpha_3 < 0$. For sufficiently large A_2, the gain drops to zero, and ... is "blocked." In RF design, the term "blocking signal" usually

refers to interferers that desensitize a circuit even if the gain does not fall to zero. Many RF receivers must be able to withstand blocking signals 60 to 70 dB greater than the wanted signal.

Cross Modulation Another phenomenon that occurs when a weak signal and a strong interferer pass through a nonlinear system is the transfer of modulation (or noise) on the amplitude of the interferer to the amplitude of the weak signal. Called "cross modulation," this effect is evident from Eq. (2.19), where variations in A_2 affect the amplitude of the output component at ω_1. For example, if the amplitude of the interferer is modulated by a sinusoid $A_2(1 + m \cos \omega_m t) \cos \omega_2 t$, where m is the modulation index and less than unity, then (2.19) assumes the following form:

$$y(t) = \left[\alpha_1 A_1 + \frac{3}{2}\alpha_3 A_1 A_2^2 \left(1 + \frac{m^2}{2} + \frac{m^2}{2} \cos 2\omega_m t + 2m \cos \omega_m t\right)\right] \cos \omega_1 t + \cdots. \tag{2.20}$$

Thus, the desired signal at the output contains amplitude modulation at ω_m and $2\omega_m$.

A common case of cross modulation arises in amplifiers that must *simultaneously* process many independent signal channels, e.g., in cable television transmitters. Nonlinearities in the amplifier corrupt each signal with the amplitude variations in other channels.

Intermodulation While harmonic distortion is often used to describe nonlinearities of analog circuits, certain cases require other measures of nonlinear behavior. For example, suppose the nonlinearity of an active low-pass filter is to be evaluated. If, as depicted in Fig. 2.5, the input sinusoid frequency is chosen such that its harmonics fall out of the passband, then the output distortion appears quite small even if the input stage of the filter introduces substantial nonlinearity. Thus, another type of test is required here. Commonly used is the "intermodulation distortion" in a "two-tone" test.

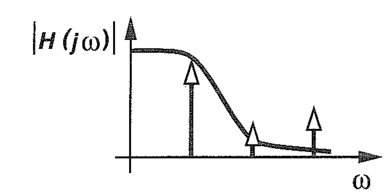


Figure 2.5 Harmonic distortion in a low-pass filter.

When two signals with different frequencies are applied to a nonlinar system, the output in general exhibits some components that are not harmonics of the input frequencies. Called intermodulation (IM), this phenomenon arises from "mixing" (multiplication) of the two signals when their sum is raised to a power greater than unity. To understand how Eq. (2.11) leads to intermodulation, assume $x(t) = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t$. Thus,

$$y(t) = \alpha_1(A_1 \cos \omega_1 t + A_2 \cos \omega_2 t) + \alpha_2(A_1 \cos \omega_1 t + A_2 \cos \omega_2 t)^2 + \alpha_3(A_1 \cos \omega_1 t + A_2 \cos \omega_2 t)^3. \quad (2.21)$$

Expanding the left side and discarding dc terms and harmonics, we obtain the following intermodulation products:

$$\omega = \omega_1 \pm \omega_2 : \alpha_2 A_1 A_2 \cos(\omega_1 + \omega_2)t + \alpha_2 A_1 A_2 \cos(\omega_1 - \omega_2)t \quad (2.22)$$

$$= 2\omega_1 \pm \omega_2 : \frac{3\alpha_3 A_1^2 A_2}{4} \cos(2\omega_1 + \omega_2)t + \frac{3\alpha_3 A_1^2 A_2}{4} \cos(2\omega_1 - \omega_2)t \quad (2.23)$$

$$= 2\omega_2 \pm \omega_1 : \frac{3\alpha_3 A_2^2 A_1}{4} \cos(2\omega_2 + \omega_1)t + \frac{3\alpha_3 A_2^2 A_1}{4} \cos(2\omega_2 - \omega_1)t \quad (2.24)$$

and these fundamental components

$$\omega = \omega_1, \omega_2 : \left(\alpha_1 A_1 + \frac{3}{4}\alpha_3 A_1^3 + \frac{3}{2}\alpha_3 A_1 A_2^2\right) \cos \omega_1 t + \left(\alpha_1 A_2 + \frac{3}{4}\alpha_3 A_2^3 + \frac{3}{2}\alpha_3 A_2 A_1^2\right) \cos \omega_2 t. \quad (2.25)$$

Of particular interest are the third-order IM products at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$, illustrated in Fig. 2.6. The key point here is that if the difference between ω_1 and ω_2 is small, the components at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ appear in the vicinity of ω_1 and ω_2, thus revealing nonlinearities even in cases such as the LPF of Fig. 2.5. In a typical two-tone test, $A_1 = A_2 = A$, and the ratio of the amplitude of the output third-order products to $\alpha_1 A$ defines the IM distortion. For example, if $\alpha_1 A = 1\ V_{pp}$, and $3\alpha_3 A^3/4 = 10\ mV_{pp}$, then we say the IM components are at −40 dBc, where the letter "c" means "with respect to the carrier."

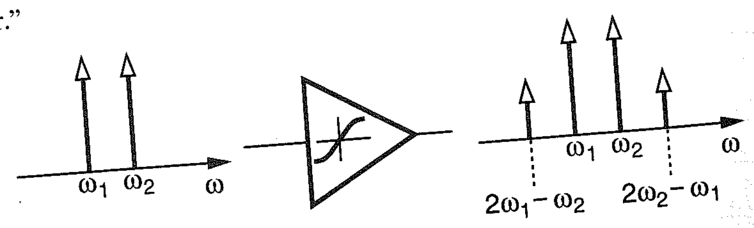


Figure 2.6 Intermodulation in a nonlinear system.

Intermodulation is a troublesome effect in RF systems. As shown in Fig. 2.7, if a weak signal accompanied by two strong interferers experiences third-order nonlinearity, then one of the IM products falls in the band of interest, corrupting the desired component. While operating on the *amplitude* of the signals, this effect degrades the performance even if the modulation is on the phase (because zero-crossing points are still affected.) Note that this phenomenon cannot be directly quantified by *harmonic* distortion.

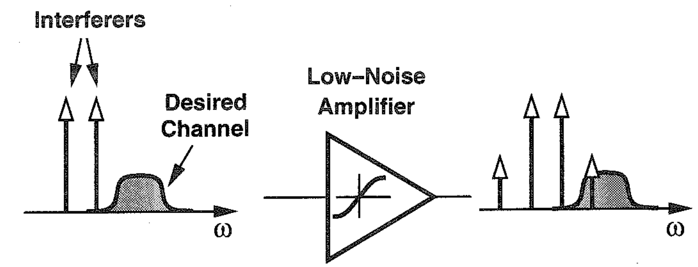


Figure 2.7 Corruption of a signal due to intermodulation between two interferers.

The corruption of signals due to third-order intermodulation of two nearby interferers is so common and so critical that a performance metric has been defined to characterize this behavior. Called the "third intercept point" (IP_3), this parameter is measured by a two-tone test in which A is chosen to be sufficiently small so that higher-order nonlinear terms are negligible and the gain is relatively constant and equal to α_1. From (2.23), (2.24), and (2.25), we note that as A increases, the fundamentals increase in proportion to A, whereas the third-order IM products increase in proportion to A^3 [Fig. 2.8(a)]. Plotted on a logarithmic scale [Fig. 2.8(b)], the magnitude of the IM products grows at three times the rate at which the main components increase. The third-order intercept point is defined to be at the intersection of the two lines. The horizontal coordinate of this point is called the input IP_3 (IIP_3), and the vertical coordinate is called the output IP_3 (OIP_3).

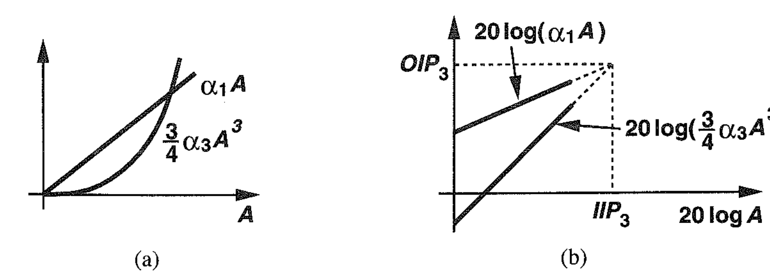


Figure 2.8 Growth of output components in an intermodulation test.

It is important to appreciate the advantage of IP_3 over a simple IM measurement. If the magnitude of IM products (normalized to that of the carrier) is used as a measure of linearity, then the input amplitude with which the test is performed must be specified. The third intercept point, on the other hand, is a unique quantity that by itself can serve as a means of comparing the linearity of different circuits.

From the input-output characteristic of Eq. (2.11), we can derive a simple expression for IP_3. Let $x(t) = A\cos\omega_1 t + A\cos\omega_2 t$. Then,

$$y(t) = \left(\alpha_1 + \frac{9}{4}\alpha_3 A^2\right) A\cos\omega_1 t + \left(\alpha_1 + \frac{9}{4}\alpha_3 A^2\right) A\cos\omega_2 t + \frac{3}{4}\alpha_3 A^3\cos(2\omega_1 - \omega_2)t + \frac{3}{4}\alpha_3 A^3 \cos(2\omega_2 - \omega_1)t + \cdots. \quad (2.26)$$

If $\alpha_1 \gg 9\alpha_3 A^2/4$, the input level for which the output components at ω_1 and ω_2 have the same amplitude as those at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ is given by

$$|\alpha_1| A_{IP3} = \frac{3}{4}|\alpha_3| A_{IP3}^3. \quad (2.27)$$

Thus, the input IP_3 is

$$A_{IP3} = \sqrt{\frac{4}{3}\left|\frac{\alpha_1}{\alpha_3}\right|}, \quad (2.28)$$

and the output IP_3 is equal to $\alpha_1 A_{IP3}$.

The parameter IP_3 characterizes only third-order nonlinearities. In practice, if the input level is increased to reach the intercept point, the assumption $\alpha_1 \gg 9\alpha_3 A^2/4$ no longer holds, the gain drops, and higher-order IM products become significant. In fact, in many circuits the IP_3 is beyond the allowable input range, sometimes even higher than the supply voltage. Thus, the practical method of obtaining the IP_3 is to measure the characteristic of Fig. 2.8(b) for small input amplitudes and use linear extrapolation on a logarithmic scale to find the intercept point.

A quick method of measuring the IP_3 is as follows. Let us denote the input level at each frequency by A_{in}, the amplitude of the output components at ω_1 and ω_2 by $A_{\omega 1,\omega 2}$, and the amplitude of the IM_3 products by A_{IM3}. Then from (2.26), we have

$$\frac{A_{\omega 1,\omega 2}}{A_{IM3}} \approx \frac{|\alpha_1| A_{in}}{3|\alpha_3| A_{in}^3/4} \quad (2.29)$$

$$= \frac{4|\alpha_1|}{3|\alpha_3|}\frac{1}{A_{in}^2}, \quad (2.30)$$

which, in conjunction with (2.28), reduces to

$$\frac{A_{\omega 1,\omega 2}}{A_{IM3}} = \frac{A_{IP3}^2}{A_{in}^2}. \quad (2.31)$$

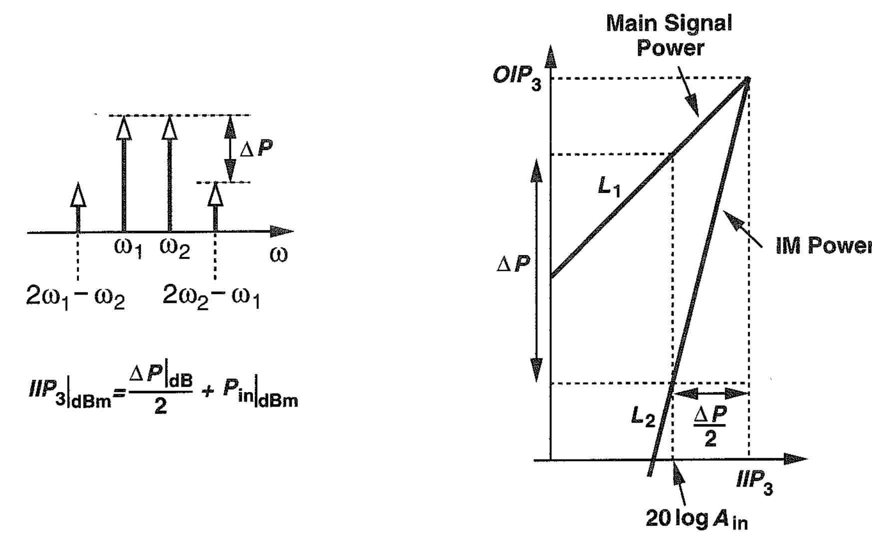
Consequently,

$$20\log A_{\omega 1,\omega 2} - 20\log A_{IM3} = 20\log A_{IP3}^2 - 20\log A_{in}^2, \quad (2.32)$$

and

$$20\log A_{IP3} = \frac{1}{2}(20\log A_{\omega 1,\omega 2} - 20\log A_{IM3}) + 20\log A_{in}. \quad (2.33)$$

Thus, if all the signal levels are expressed in dBm, the *input* third intercept point is equal to half the difference between the magnitudes of the fundamentals and the IM_3 products at the *output* plus the corresponding input level [Fig. 2.9(a)]. The key point here is that IP_3 can be measured with only one input level, obviating the need for extrapolation.



(a) (b)

Figure 2.9 (a) Calculation of IP_3 without extrapolation, (b) graphical interpretation of (a).

Shown in Fig. 2.9(b) is a geometric interpretation of the above relationship. Since line L_1 has a slope equal to unity and line L_2 a slope equal to 3, an input increment $\Delta P/2$ yields an equal increment in L_1 and an increment equal to $3\Delta P/2$ in L_2, reducing the difference between the two lines to zero.

The above approach provides an estimate of IP_3 in initial phases of the design or characterization. The actual value of IP_3, however, must still be obtained through accurate extrapolation to ensure that all nonlinear and frequency-dependent effects are taken into account.

Another measurement method encountered in the literature is to apply a *single* tone, plot the third harmonic magnitude versus the input level, and obtain the intercept point by extrapolation. From the example of Fig. 2.5, we note that this technique does not yield a correct value for IP_3.

To gain a better feeling about the required linearity in typical RF systems, let us calculate the amount of corruption that a 1-μV_{rms} signal experiences by two 1-mV_{rms} interferers in an amplifier having an IIP_3 of 70 mV_{rms} (≈ -10 dBm) (Fig. 2.10). Neglecting desensitization and cross modulation, we can write

$$\frac{A_{\text{sig,out}}}{A_{\text{sig,in}}} \approx \frac{A_{\text{int,out}}}{A_{\text{int,in}}}, \tag{2.34}$$

where A_{sig} denotes the signal amplitude and A_{int} the interferer amplitude. It follows from (2.31) that

$$\frac{A_{\text{sig,out}}}{A_{IM3,\text{out}}} = \frac{A_{\text{sig,in}} \cdot A_{IP3}^2}{A_{\text{int,in}}^3} \tag{2.35}$$

where $A_{\text{sig,in}} = 1\ \mu V_{rms}$, $A_{IP3} = 70$ mV_{rms}, and $A_{\text{int,in}} = 1$ mV_{rms}. Thus, the ratio is equal to $4.9 \approx 13.8$ dB.

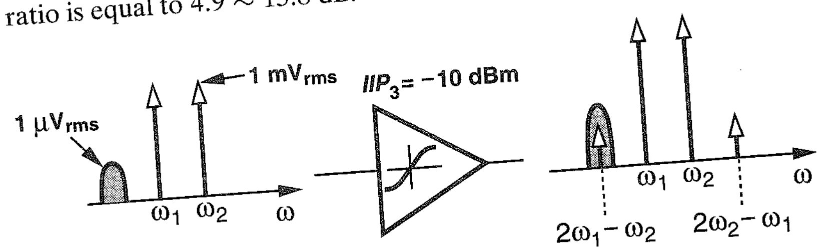


Figure 2.10 Example of achievable SNR in the presence of large interferers.

It is also instructive to find the relationship between the 1-dB compression point and the input IP_3 for a third-order nonlinearity. From (2.17) and (2.28), we conclude that the two are related by

$$\frac{A_{1\text{-dB}}}{A_{IP3}} = \frac{\sqrt{0.145}}{\sqrt{4/3}} \tag{2.36}$$

$$\approx -9.6 \text{ dB}. \tag{2.37}$$

2.1.2 Cascaded Nonlinear Stages

Since in RF systems, signals are processed by cascaded stages, it is important to know how the nonlinearity of each stage is referred to the input of the cascade. In particular, it is desirable to calculate an overall input third intercept point in terms of the IP_3 and gain of the individual stages.

Consider two nonlinear stages in cascade (Fig. 2.11). If the input-output characteristics of the two stages are expressed, respectively, as

$$y_1(t) = \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \tag{2.38}$$

$$y_2(t) = \beta_1 y_1(t) + \beta_2 y_1^2(t) + \beta_3 y_1^3(t), \tag{2.39}$$

then

$$\begin{aligned} y_2(t) = {} & \beta_1[\alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t)] \\ & + \beta_2[\alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t)]^2 \\ & + \beta_3[\alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t)]^3. \end{aligned} \tag{2.40}$$

Considering only the first- and third-order terms, we have

$$y_2(t) = \alpha_1\beta_1 x(t) + (\alpha_3\beta_1 + 2\alpha_1\alpha_2\beta_2 + \alpha_1^3\beta_3)x^3(t) + \cdots. \tag{2.41}$$

Thus, from (2.28)

$$A_{IP3} = \sqrt{\frac{4}{3}\left|\frac{\alpha_1\beta_1}{\alpha_3\beta_1 + 2\alpha_1\alpha_2\beta_2 + \alpha_1^3\beta_3}\right|}. \tag{2.42}$$

Interestingly, proper choice of the values and signs of the terms in the denominator can yield an arbitrarily high IP_3. In practice, however, other considerations such as noise, gain, and active device characteristics may not permit this choice. As a worst-case estimate, we add the absolute values of the three terms in the denominator.

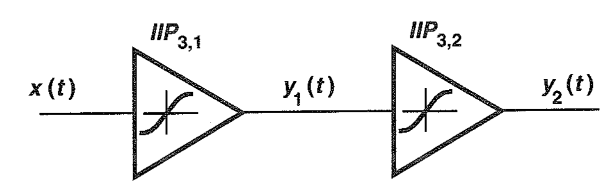


Figure 2.11 Cascaded nonlinear stages.

Equation (2.42) can be simplified if the two sides are inverted and squared:

$$\frac{1}{A_{IP3}^2} = \frac{3}{4}\frac{|\alpha_3\beta_1| + |2\alpha_1\alpha_2\beta_2| + |\alpha_1^3\beta_3|}{|\alpha_1\beta_1|} \tag{2.43}$$

$$= \frac{1}{A_{IP3,1}^2} + \frac{3\alpha_2\beta_2}{2\beta_1} + \frac{\alpha_1^2}{A_{IP3,2}^2}, \tag{2.44}$$

where $A_{IP3,1}$ and $A_{IP3,2}$ represent the input IP_3 points of the first and second stages, respectively. Note that A_{IP3}, $A_{IP3,1}$, and $A_{IP3,2}$ are voltage quantities rather than power quantities.

From the above result, we note that as α_1 increases, the overall IP_3 decreases. This is because with higher gain in the first stage, the second stage senses larger input levels, thereby producing much greater IM_3 products. (Recall that IM_3 products grow with the third power of the input amplitude.)

To gain more insight, we assume $x(t) = A\cos\omega_1 t + A\cos\omega_2 t$ and identify the various IM products. Referring to Fig. 2.12, we make the following observations.[1] (1) The fundamental input components are amplified by approximately α_1 in the first stage and β_1 in the second. Thus the output fundamentals are $\alpha_1\beta_1 A(\cos\omega_1 t + \cos\omega_2 t)$. (2) The IM_3 products generated by the first stage, namely, $(3\alpha_3/4)A^3[\cos(2\omega_1-\omega_2)t+\cos(2\omega_2-\omega_1)t]$, are also amplified by β_1 when they appear at the output of the second stage. (3) The second stage senses $\alpha_1 A(\cos\omega_1 t+\cos\omega_2 t)$ at its input and hence generates these IM_3 products: $(3\beta_3/4)(\alpha_1 A)^3\cos(2\omega_1-\omega_2)t+(3\beta_3/4)(\alpha_1 A)^3\cos(2\omega_2-\omega_1)t$. (4) The second-order nonlinearity in $y_1(t)$ generates components at $\omega_1-\omega_2$, $2\omega_1$, and $2\omega_2$. Upon experiencing a similar nonlinearity in the second stage, such components are translated to $2\omega_1-\omega_2$ and $2\omega_2-\omega_1$. More specifically, as shown in Fig. 2.12, $y_2(t)$ contains terms such as $2\beta_2[\alpha_1 A\cos\omega_1 t\cdot\alpha_2 A^2\cos(\omega_1-\omega_2)t]$ and $2\beta_2(\alpha_1 A\cos\omega_2 t\cdot 0.5\alpha_2 A^2\cos 2\omega_1 t)$. The resulting third-order IM products can be expressed as $(3\alpha_1\alpha_2\beta_2 A^3/2)[\cos(2\omega_1-\omega_2)t+\cos(2\omega_2-\omega_1)t]$.

From these observations, we can write

$$y_2(t) = \alpha_1\beta_1 A(\cos\omega_1 t + \cos\omega_2 t) + \left(\frac{3\alpha_3\beta_1}{4} + \frac{3\alpha_1^3\beta_3}{4} + \frac{3\alpha_1\alpha_2\beta_2}{2}\right)A^3[\cos(2\omega_1-\omega_2)t + \cos(2\omega_2-\omega_1)t] + \cdots, \tag{2.45}$$

obtaining the same IP_3 as above.

In many RF systems, each stage in a cascade has a narrow frequency band. Thus, the components described in the fourth observation above fall out of the band and are heavily attenuated. Consequently, the second term on the right-hand side of (2.44) becomes negligible, giving

$$\frac{1}{A_{IP3}^2} \approx \frac{1}{A_{IP3,1}^2} + \frac{\alpha_1^2}{A_{IP3,2}^2}. \tag{2.46}$$

This equation readily gives a general expression for three or more stages:

$$\frac{1}{A_{IP3}^2} \approx \frac{1}{A_{IP3,1}^2} + \frac{\alpha_1^2}{A_{IP3,2}^2} + \frac{\alpha_1^2\beta_1^2}{A_{IP3,3}^2} + \cdots, \tag{2.47}$$

where $A_{IP3,3}$ denotes the input IP_3 of the third stage. Thus, if each stage in a cascade has a gain greater than unity, the nonlinearity of the latter stages becomes increasingly more critical because the IP_3 of each stage is effectively scaled *down* by the total gain preceding that stage. We should emphasize that (2.47) is merely an approximation. In practice, more precise calculations or simulations must be performed to predict the overall IP_3.

[1] The spectrum of $A\cos\omega t$ consists of two impulses, each with a weight $A/2$. We drop the factor 1/2 in the figures for simplicity. The end result is still correct.

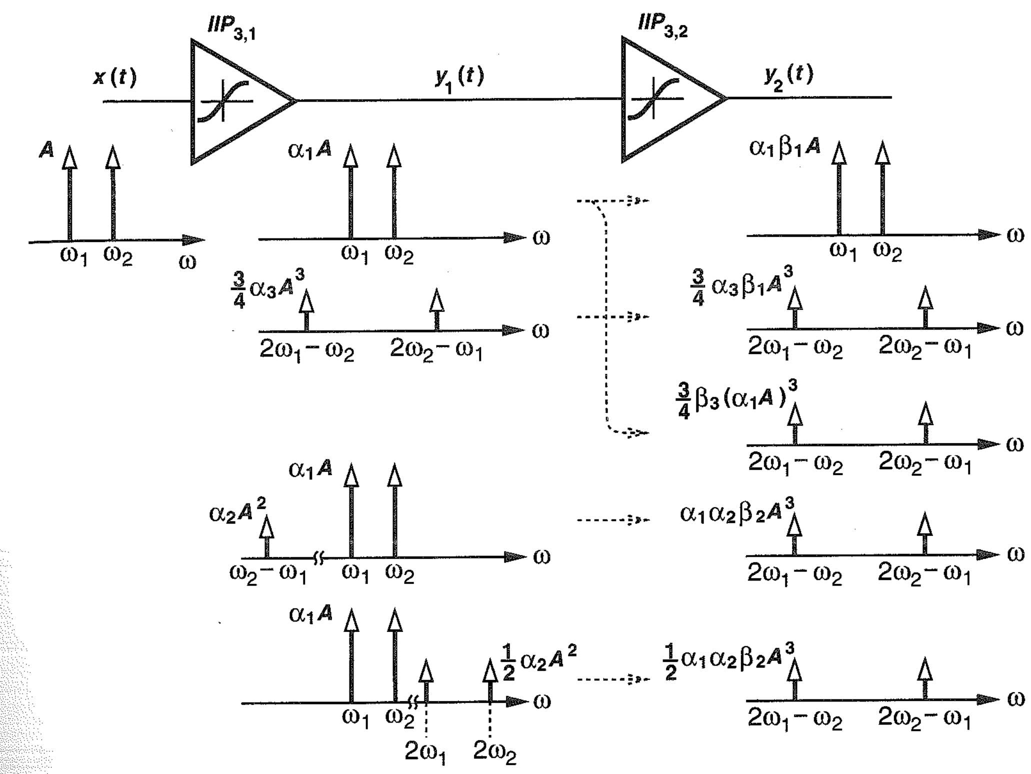


Figure 2.12 Intermodulation mechanisms in cascade of two nonlinear stages.

2.2 INTERSYMBOL INTERFERENCE

Linear time-invariant systems can also "distort" a signal if they do not have sufficient bandwidth. Attenuation of high-frequency components of a periodic square wave in a low-pass filter is a familiar example of such behavior [Fig. 2.13(a)]. However, limited bandwidth has a more detrimental effect on *random* bit streams. To understand the issue, first recall that if a single ideal rectangular pulse is applied to a low-pass filter, the output exhibits an exponential tail that becomes more significant as the filter bandwidth decreases. This occurs fundamentally because a signal cannot be both time limited and bandwidth limited: when the time-limited pulse passes through the band-limited system, the output must extend to infinity in the time domain.