[5] M. Soyuer et al., "A 3-V 4-GHz nMOS Voltage-Controlled Oscillator with Integrated Resonator," *IEEE J. Solid-State Circuits,* vol. 31, pp. 2042–2045, Dec. 1996.
[6] B. Kleveland et al., "Monolithic CMOS Distributed Amplifier and Oscillator," *ISSCC Dig. Tech. Papers,* pp. 70–71, Feb. 1999.
[7] H. Wang, "A 50-GHz VCO in 0.25-$\mu$m CMOS," *ISSCC Dig. Tech. Papers,* pp. 372–373, Feb. 2001.
[8] L. Franca-Neto, R. Bishop, and B. Bloechel, "64 GHz and 100 GHz VCOs in 90 nm CMOS Using Optimum Pumping Method," ISSCC Dig. Tech. Papers, pp. 444–445, Feb. 2004.
[9] E. Seok et al., "A 410GHz CMOS Push-Push Oscillator with an On-Chip Patch Antenna" *ISSCC Dig. Tech. Papers,* pp. 472–473, Feb. 2008.
[10] B. Razavi, "A 300-GHz Fundamental Oscillator in 65-nm CMOS Technology," *Symposium on VLSI Circuits Dig. Of Tech. Papers,* pp. 113–114, June 2010.



CHAPTER

2

# BASIC CONCEPTS IN RF DESIGN

RF design draws upon many concepts from a variety of fields, including signals and systems, electromagnetics and microwave theory, and communications. Nonetheless, RF design has developed its own analytical methods and its own language. For example, while the nonlinear behavior of analog circuits may be characterized by "harmonic distortion," that of RF circuits is quantified by very different measures.

This chapter deals with general concepts that prove essential to the analysis and design of RF circuits, closing the gaps with respect to other fields such as analog design, microwave theory, and communication systems. The outline is shown below.

| Nonlinearity | Noise | Impedance Transformation |
|---|---|---|
| ▪ Harmonic Distortion | ▪ Noise Spectrum | ▪ Series–Parallel Conversion |
| ▪ Compression | ▪ Device Noise | ▪ Matching Networks |
| ▪ Intermodulation | ▪ Noise in Circuits | ▪ S–Parameters |
| ▪ Dynamic Nonlinear Systems | | |

## 2.1 GENERAL CONSIDERATIONS

### 2.1.1 Units in RF Design

RF design has traditionally employed certain units to express gains and signal levels. It is helpful to review these units at the outset so that we can comfortably use them in our subsequent studies.

The voltage gain, $V_{out}/V_{in}$, and power gain, $P_{out}/P_{in}$, are expressed in decibels (dB):

$$A_V|_{\mathrm{dB}} = 20\log\frac{V_{out}}{V_{in}} \tag{2.1}$$

$$A_P|_{\mathrm{dB}} = 10\log\frac{P_{out}}{P_{in}}. \tag{2.2}$$

These two quantities are equal (in dB) only if the input and output voltages appear across *equal* impedances. For example, an amplifier having an input resistance of $R_0$ (e.g., 50 Ω) and driving a load resistance of $R_0$ satisfies the following equation:

$$A_P|_{\text{dB}} = 10 \log \frac{\frac{V_{out}^2}{R_0}}{\frac{V_{in}^2}{R_0}} \tag{2.3}$$

$$= 20 \log \frac{V_{out}}{V_{in}} \tag{2.4}$$

$$= A_V|_{\text{dB}}, \tag{2.5}$$

where $V_{out}$ and $V_{in}$ are rms values. In many RF systems, however, this relationship does not hold because the input and output impedances are not equal.

The absolute signal levels are often expressed in dBm rather than in watts or volts. Used for power quantities, the unit dBm refers to "dB's above 1 mW." To express the signal power, $P_{sig}$, in dBm, we write

$$P_{sig}|_{\text{dBm}} = 10 \log \left( \frac{P_{sig}}{1\text{ mW}} \right). \tag{2.6}$$

**Example 2.1**

An amplifier senses a sinusoidal signal and delivers a power of 0 dBm to a load resistance of 50 Ω. Determine the peak-to-peak voltage swing across the load.

**Solution:**

Since 0 dBm is equivalent to 1 mW, for a sinusoidal having a peak-to-peak amplitude of $V_{pp}$ and hence an rms value of $V_{pp}/(2\sqrt{2})$, we write

$$\frac{V_{pp}^2}{8R_L} = 1\text{ mW}, \tag{2.7}$$

where $R_L = 50\ \Omega$. Thus,

$$V_{pp} = 632\text{ mV}. \tag{2.8}$$

This is an extremely useful result, as demonstrated in the next example.

**Example 2.2**

A GSM receiver senses a narrowband (modulated) signal having a level of −100 dBm. If the front-end amplifier provides a voltage gain of 15 dB, calculate the peak-to-peak voltage swing at the output of the amplifier.



**Example 2.2** (*Continued*)

**Solution:**

Since the amplifier output *voltage* swing is of interest, we first convert the received signal level to voltage. From the previous example, we note that −100 dBm is 100 dB below 632 $\text{mV}_{\text{pp}}$. Also, 100 dB for voltage quantities is equivalent to $10^5$. Thus, −100 dBm is equivalent to 6.32 $\mu\text{V}_{\text{pp}}$. This input level is amplified by 15 dB (≈ 5.62), resulting in an output swing of 35.5 $\mu\text{V}_{\text{pp}}$.

The reader may wonder why the output *voltage* of the amplifier is of interest in the above example. This may occur if the circuit following the amplifier does not present a 50-Ω input impedance, and hence the power gain and voltage gain are not equal in dB. In fact, the next stage may exhibit a purely *capacitive* input impedance, thereby requiring no signal "power." This situation is more familiar in analog circuits wherein one stage drives the gate of the transistor in the next stage. As explained in Chapter 5, in most integrated RF systems, we prefer voltage quantities to power quantities so as to avoid confusion if the input and output impedances of cascade stages are unequal or contain negligible real parts.

The reader may also wonder why we were able to assume 0 dBm is equivalent to 632 $\text{mV}_{\text{pp}}$ in the above example even though the signal is not a pure sinusoid. After all, only for a sinusoid can we assume that the rms value is equal to the peak-to-peak value divided by $2\sqrt{2}$. Fortunately, for a narrowband 0-dBm signal, it is still possible to approximate the (average) peak-to-peak swing as 632 mV.

Although dBm is a unit of power, we sometimes use it at interfaces that do not necessarily entail power transfer. For example, consider the case shown in Fig. 2.1(a), where the LNA drives a purely-capacitive load with a 632-$\text{mV}_{\text{pp}}$ swing, delivering no average power. We mentally attach an ideal voltage buffer to node $X$ and drive a 50-Ω load [Fig. 2.1(b)]. We then say that the signal at node $X$ has a level of 0 dBm, tacitly meaning that *if* this signal were applied to a 50-Ω load, *then* it would deliver 1 mW.

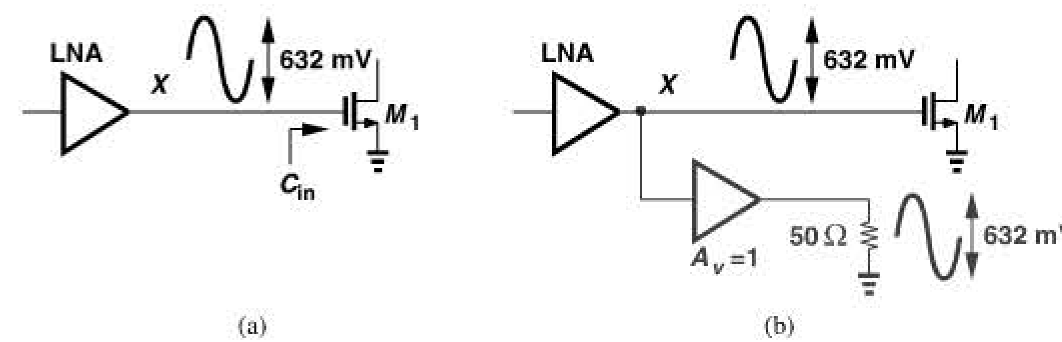



**Figure 2.1** *(a) LNA driving a capacitive impedance, (b) use of fictitious buffer to visualize the signal level in dBm.*

### 2.1.2 Time Variance

A system is linear if its output can be expressed as a linear combination (superposition) of responses to individual inputs. More specifically, if the outputs in response to inputs $x_1(t)$

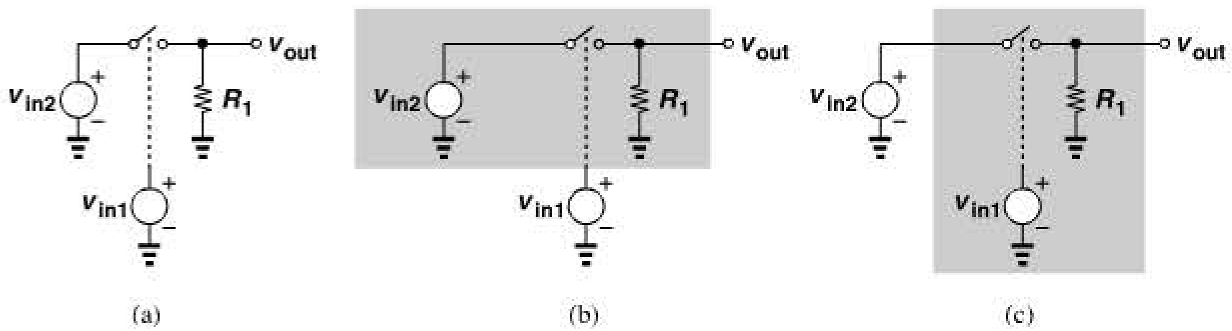



**Figure 2.2** *(a) Simple switching circuit, (b) system with $V_{in1}$ as the input, (c) system with $V_{in2}$ as the input.*

and $x_2(t)$ can be respectively expressed as

$$y_1(t) = f[x_1(t)] \tag{2.9}$$
$$y_2(t) = f[x_2(t)], \tag{2.10}$$

then,

$$ay_1(t) + by_2(t) = f[ax_1(t) + bx_2(t)], \tag{2.11}$$

for arbitrary values of $a$ and $b$. Any system that does not satisfy this condition is nonlinear. Note that, according to this definition, nonzero initial conditions or dc offsets also make a system nonlinear, but we often relax the rule to accommodate these two effects.

Another attribute of systems that may be confused with nonlinearity is time variance. A system is time-invariant if a time shift in its input results in the same time shift in its output. That is, if $y(t) = f[x(t)]$, then $y(t-\tau) = f[x(t-\tau)]$ for arbitrary $\tau$.

As an example of an RF circuit in which time variance plays a critical role and must not be confused with nonlinearity, let us consider the simple switching circuit shown in Fig. 2.2(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 depicted in Fig. 2.2(b), the input of interest is $v_{in1}$ (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 independent of its amplitude. This system is also time-variant because the output depends on $v_{in2}$. For example, if $v_{in1}$ is constant and positive, then $v_{out}(t) = v_{in2}(t)$, and if $v_{in1}$ is constant and negative, then $v_{out}(t) = 0$ (why?).

Now consider the case shown in Fig. 2.2(c), where the input of interest is $v_{in2}$ (while $v_{in1}$ remains part of the system and still equal to $A_1 \cos\omega_1 t$). This system is linear with respect to $v_{in2}$. For example, doubling the amplitude of $v_{in2}$ directly doubles that of $v_{out}$. The system is also time-variant due to the effect of $v_{in1}$.

### Example 2.3

Plot the output waveform of the circuit in Fig. 2.2(a) if $v_{in1} = A_1 \cos\omega_1 t$ and $v_{in2} = A_2 \cos(1.25\omega_1 t)$.



### Example 2.3 (Continued)

**Solution:**

As shown in Fig. 2.3, $v_{out}$ tracks $v_{in2}$ if $v_{in1} > 0$ and is pulled down to zero by $R_1$ if $v_{in1} < 0$. That is, $v_{out}$ is equal to the product of $v_{in2}$ and a square wave toggling between 0 and 1.

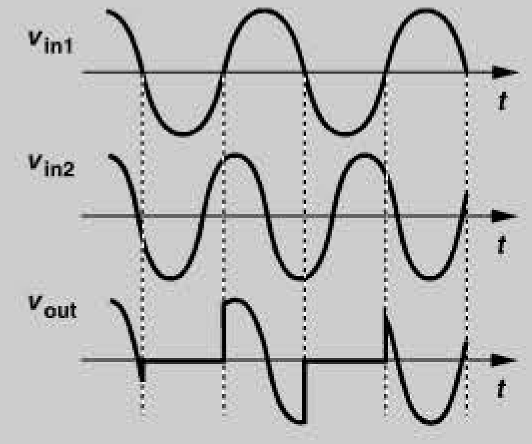



**Figure 2.3** *Input and output waveforms.*

The circuit of Fig. 2.2(a) is an example of RF "mixers." We will study such circuits in Chapter 6 extensively, but it is important to draw several conclusions from the above study. First, statements such as "switches are nonlinear" are ambiguous. Second, a linear system *can* generate frequency components that do not exist in the input signal—the system only need be time-variant. From Example 2.3,

$$v_{out}(t) = v_{in2}(t) \cdot S(t), \tag{2.12}$$

where $S(t)$ denotes a square wave toggling between 0 and 1 with a frequency of $f_1 = \omega_1/(2\pi)$. The output spectrum is therefore given by the convolution of the spectra of $v_{in2}(t)$ and $S(t)$. Since the spectrum of a square wave is equal to a train of impulses whose amplitudes follow a sinc envelope, we have

$$V_{out}(f) = V_{in2}(f) * \sum_{n=-\infty}^{+\infty} \frac{\sin(n\pi/2)}{n\pi}\delta\left(f - \frac{n}{T_1}\right) \tag{2.13}$$
$$= \sum_{n=-\infty}^{+\infty} \frac{\sin(n\pi/2)}{n\pi} V_{in2}\left(f - \frac{n}{T_1}\right), \tag{2.14}$$

where $T_1 = 2\pi/\omega_1$. This operation is illustrated in Fig. 2.4 for a $V_{in2}$ spectrum located around zero frequency.[1]

1. It is helpful to remember that, for $n = 1$, each impulse in the above summation has an area of $1/\pi$ and the corresponding sinusoid, a *peak amplitude* of $2/\pi$.

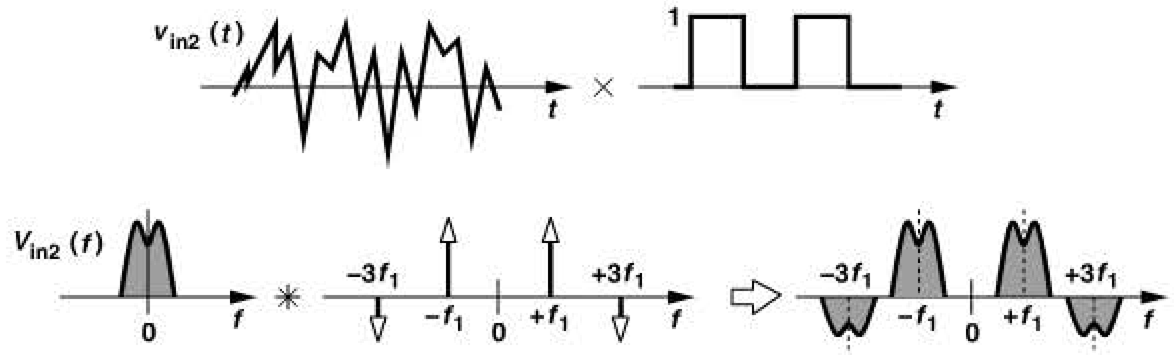



**Figure 2.4** *Multiplication in the time domain and corresponding convolution in the frequency domain.*

### 2.1.3 Nonlinearity

A system is called "memoryless" or "static" if its output does not depend on the past values of its input (or the past values of the output itself). For a memoryless linear system, the input/output characteristic is given by

$$y(t) = \alpha x(t), \tag{2.15}$$

where $\alpha$ is a function of time if the system is time-variant [e.g., Fig. 2.2(c)]. For a memoryless nonlinear system, the input/output characteristic can be approximated with a polynomial,

$$y(t) = \alpha_0 + \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) + \cdots, \tag{2.16}$$

where $\alpha_j$ may be functions of time if the system is time-variant. Figure 2.5 shows a common-source stage as an example of a memoryless nonlinear circuit (at low frequencies). If $M_1$ operates in the saturation region and can be approximated as a square-law device, then

$$V_{out} = V_{DD} - I_D R_D \tag{2.17}$$

$$= V_{DD} - \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{in} - V_{TH})^2 R_D. \tag{2.18}$$

In this idealized case, the circuit displays only second-order nonlinearity.

The system described by Eq. (2.16) has "odd symmetry" if $y(t)$ is an odd function of $x(t)$, i.e., if the response to $-x(t)$ is the negative of that to $+x(t)$. This occurs if $\alpha_j = 0$ for even $j$. Such a system is sometimes called "balanced," as exemplified by the differential

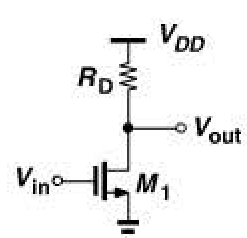



**Figure 2.5** *Common-source stage.*



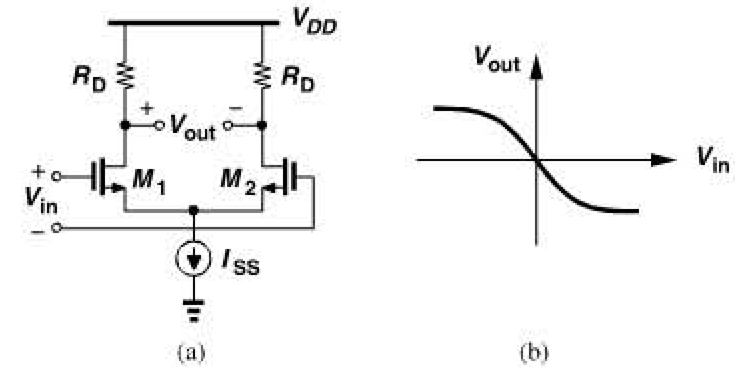



**Figure 2.6** *(a) Differential pair and (b) its input/output characteristic.*

pair shown in Fig. 2.6(a). Recall from basic analog design that by virtue of symmetry, the circuit exhibits the characteristic depicted in Fig. 2.6(b) if the differential input varies from very negative values to very positive values.

#### Example 2.4

For square-law MOS transistors operating in saturation, the characteristic of Fig. 2.6(b) can be expressed as [1]

$$V_{out} = -\frac{1}{2}\mu_n C_{ox}\frac{W}{L}V_{in}\sqrt{\frac{4I_{SS}}{\mu_n C_{ox}\frac{W}{L}} - V_{in}^2}\,R_D. \tag{2.19}$$

If the differential input is small, approximate the characteristic by a polynomial.

**Solution:**

Factoring $4I_{SS}/(\mu_n C_{ox} W/L)$ out of the square root and assuming

$$V_{in}^2 \ll \frac{4I_{SS}}{\mu_n C_{ox}\frac{W}{L}}, \tag{2.20}$$

we use the approximation $\sqrt{1-\epsilon} \approx 1 - \epsilon/2$ to write

$$V_{out} \approx -\sqrt{\mu_n C_{ox}\frac{W}{L}I_{SS}}\,V_{in}\left(1 - \frac{\mu_n C_{ox}\frac{W}{L}}{8I_{SS}}V_{in}^2\right)R_D \tag{2.21}$$

$$\approx -\sqrt{\mu_n C_{ox}\frac{W}{L}I_{SS}}\,R_D V_{in} + \frac{\left(\mu_n C_{ox}\frac{W}{L}\right)^{3/2}}{8\sqrt{I_{SS}}}R_D V_{in}^3. \tag{2.22}$$

The first term on the right-hand side represents linear operation, revealing the small-signal voltage gain of the circuit ($-g_m R_D$). Due to symmetry, even-order nonlinear terms are absent. Interestingly, square-law devices yield a *third-order* characteristic in this case. We return to this point in Chapter 5.

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,

$$y(t) = h(t) * x(t), \tag{2.23}$$

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 $\delta(t)$ yields $h(t)$, then $\delta(t-\tau)$ produces $h(t, \tau)$. Thus,

$$y(t) = h(t, \tau) * x(t). \tag{2.24}$$

Finally, if a system is both nonlinear and dynamic, then its impulse response can be approximated by a Volterra series. This is described in Section 2.8.

## 2.2 EFFECTS OF NONLINEARITY

While analog and RF circuits can be approximated by a linear model for small-signal operation, nonlinearities often lead to interesting and important phenomena that are not predicted by small-signal models. In this section, we study these phenomena for memoryless systems whose input/output characteristic can be approximated by[2]

$$y(t) \approx \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t). \tag{2.25}$$

The reader is cautioned, however, that the effect of storage elements (dynamic nonlinearity) and higher-order nonlinear terms must be carefully examined to ensure (2.25) is a plausible representation. Section 2.7 deals with the case of dynamic nonlinearity. We may consider $\alpha_1$ as the small-signal gain of the system because the other two terms are negligible for small input swings. For example, $\alpha_1 = -\sqrt{\mu_n C_{ox}(W/L)I_{SS}}R_D$ in Eq. (2.22).

The nonlinearity effects described in this section primarily arise from the third-order term in Eq. (2.25). The second-order term too manifests itself in certain types of receivers and is studied in Chapter 4.

### 2.2.1 Harmonic Distortion

If a sinusoid is applied to a nonlinear system, the output generally exhibits frequency components that are integer multiples ("harmonics") of the input frequency. In Eq. (2.25), if $x(t) = A\cos\omega t$, then

$$y(t) = \alpha_1 A\cos\omega t + \alpha_2 A^2\cos^2\omega t + \alpha_3 A^3\cos^3\omega t \tag{2.26}$$

$$= \alpha_1 A\cos\omega t + \frac{\alpha_2 A^2}{2}(1+\cos 2\omega t) + \frac{\alpha_3 A^3}{4}(3\cos\omega t + \cos 3\omega t) \tag{2.27}$$

$$= \frac{\alpha_2 A^2}{2} + \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4}\right)\cos\omega t + \frac{\alpha_2 A^2}{2}\cos 2\omega t + \frac{\alpha_3 A^3}{4}\cos 3\omega t. \tag{2.28}$$

2. Note that this expression should be considered as a fit across the signal swings of interest rather than as a Taylor expansion in the vicinity of $x = 0$. These two views may yield slightly different values for $\alpha_j$.



In Eq. (2.28), the first term on the right-hand side is a dc quantity arising from second-order nonlinearity, the second is called the "fundamental," the third is the second harmonic, and the fourth is the third harmonic. We sometimes say that even-order nonlinearity introduces dc offsets.

From the above expansion, we make two observations. First, even-order harmonics result from $\alpha_j$ with even $j$, and vanish if the system has odd symmetry, i.e., if it is fully differential. In reality, however, random mismatches corrupt the symmetry, yielding finite even-order harmonics. Second, in (2.28) the amplitudes of the second and third harmonics are proportional to $A^2$ and $A^3$, respectively, i.e., we say the $n$th harmonic grows in proportion to $A^n$.

In many RF circuits, harmonic distortion is unimportant or an irrelevant indicator of the effect of nonlinearity. For example, an amplifier operating at 2.4 GHz produces a second harmonic at 4.8 GHz, which is greatly suppressed if the circuit has a narrow bandwidth. Nonetheless, harmonics must always be considered carefully before they are dismissed. The following examples illustrate this point.

**Example 2.5**

An analog multiplier "mixes" its two inputs as shown in Fig. 2.7, ideally producing $y(t) = kx_1(t)x_2(t)$, where $k$ is a constant.[3] Assume $x_1(t) = A_1\cos\omega_1 t$ and $x_2(t) = A_2\cos\omega_2 t$.

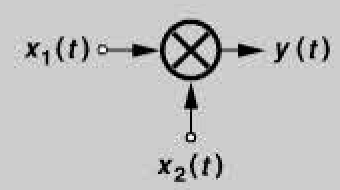



**Figure 2.7** *Analog multiplier.*

(a) If the mixer is ideal, determine the output frequency components.
(b) If the input port sensing $x_2(t)$ suffers from third-order nonlinearity, determine the output frequency components.

**Solution:**

(a) We have

$$y(t) = k(A_1\cos\omega_1 t)(A_2\cos\omega_2 t) \tag{2.29}$$

$$= \frac{kA_1A_2}{2}\cos(\omega_1+\omega_2)t + \frac{kA_1A_2}{2}\cos(\omega_1-\omega_2)t. \tag{2.30}$$

The output thus contains the sum and difference frequencies. These may be considered "desired" components.

*(Continues)*

3. The factor $k$ is necessary to ensure a proper dimension for $y(t)$.

**Example 2.5** (*Continued*)

(b) Representing the third harmonic of $x_2(t)$ by $(\alpha_3 A_2^3/4)\cos 3\omega_2 t$, we write

$$y(t) = k(A_1 \cos \omega_1 t)\left(A_2 \cos \omega_2 t + \frac{\alpha_3 A_2^3}{4} \cos 3\omega_2 t\right) \tag{2.31}$$

$$= \frac{kA_1A_2}{2}\cos(\omega_1 + \omega_2)t + \frac{kA_1A_2}{2}\cos(\omega_1 - \omega_2)t + \frac{k\alpha_3 A_1 A_2^3}{8}\cos(\omega_1 + 3\omega_2)t + \frac{k\alpha_3 A_1 A_2^3}{8}\cos(\omega_1 - 3\omega_2)t. \tag{2.32}$$

The mixer now produces two "spurious" components at $\omega_1 + 3\omega_2$ and $\omega_1 - 3\omega_2$, one or both of which often prove problematic. For example, if $\omega_1 = 2\pi \times (850 \text{ MHz})$ and $\omega_2 = 2\pi \times (900 \text{ MHz})$, then $|\omega_1 - 3\omega_2| = 2\pi \times (1850 \text{ MHz})$, an "undesired" component that is difficult to filter because it lies close to the desired component at $\omega_1 + \omega_2 = 2\pi \times (1750 \text{ MHz})$.

**Example 2.6**

The transmitter in a 900-MHz GSM cellphone delivers 1 W of power to the antenna. Explain the effect of the harmonics of this signal.

**Solution:**

The second harmonic falls within another GSM cell phone band around 1800 MHz and must be sufficiently small to negligibly impact the other users in that band. The third, fourth, and fifth harmonics do not coincide with any popular bands but must still remain below a certain level imposed by regulatory organizations in each country. The sixth harmonic falls in the 5-GHz band used in wireless local area networks (WLANs), e.g., in laptops. Figure 2.8 summarizes these results.

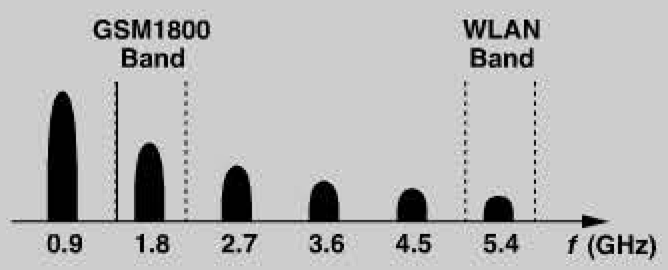



**Figure 2.8** *Summary of harmonic components.*

### 2.2.2 Gain Compression

The small-signal gain of circuits is usually obtained with the assumption that harmonics are negligible. However, our formulation of harmonics, as expressed by Eq. (2.28), indicates



that the gain experienced by $A\cos\omega t$ is equal to $\alpha_1 + 3\alpha_3 A^2/4$ and hence varies appreciably as $A$ becomes larger.[4] We must then ask, do $\alpha_1$ and $\alpha_3$ have the same sign or opposite signs? Returning to the third-order polynomial in Eq. (2.25), we note that if $\alpha_1\alpha_3 > 0$, then $\alpha_1 x + \alpha_3 x^3$ overwhelms $\alpha_2 x^2$ for large $x$ regardless of the sign of $\alpha_2$, yielding an "expansive" characteristic [Fig. 2.9(a)]. For example, an ideal bipolar transistor operating in the forward active region produces a collector current in proportion to $\exp(V_{BE}/V_T)$, exhibiting expansive behavior. On the other hand, if $\alpha_1\alpha_3 < 0$, the term $\alpha_3 x^3$ "bends" the characteristic for sufficiently large $x$ [Fig. 2.9(b)], leading to "compressive" behavior, i.e., a decreasing gain as the input amplitude increases. For example, the differential pair of Fig. 2.6(a) suffers from compression as the second term in (2.22) becomes comparable with the first. Since most RF circuits of interest are compressive, we hereafter focus on this type.

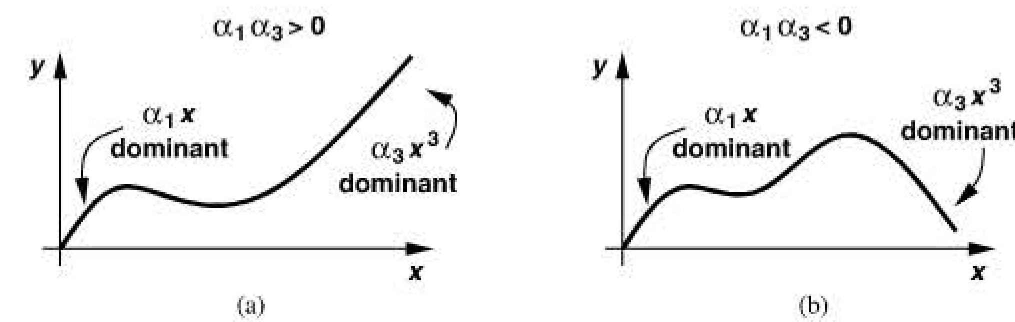



**Figure 2.9** *(a) Expansive and (b) compressive characteristics.*

With $\alpha_1\alpha_3 < 0$, the gain experienced by $A\cos\omega t$ in Eq. (2.28) falls as $A$ rises. We quantify this effect by the "1-dB compression point," defined as the input signal level that causes the gain to drop by 1 dB. If plotted on a log-log scale as a function of the input level, the output level, $A_{out}$, falls below its ideal value by 1 dB at the 1-dB compression point, $A_{in,1dB}$ (Fig. 2.10). Note that (a) $A_{in}$ and $A_{out}$ are voltage quantities here, but compression can also be expressed in terms of power quantities; (b) 1-dB compression may also be specified in terms of the output level at which it occurs, $A_{out,1dB}$. The input and output compression points typically prove relevant in the receive path and the transmit path, respectively.

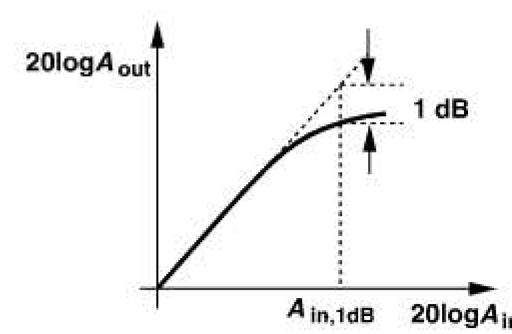



**Figure 2.10** *Definition of 1-dB compression point.*

4. This effect is akin to the fact that nonlinearity can also be viewed as variation of the *slope* of the input/output characteristic with the input level.

To calculate the input 1-dB compression point, we equate the compressed gain, $\alpha_1 + (3\alpha_3/4)A^2_{in,1dB}$, to 1 dB less than the ideal gain, $\alpha_1$:

$$20 \log \left| \alpha_1 + \frac{3}{4}\alpha_3 A^2_{in,1dB} \right| = 20 \log |\alpha_1| - 1 \text{ dB}. \tag{2.33}$$

It follows that

$$A_{in,1dB} = \sqrt{0.145 \left| \frac{\alpha_1}{\alpha_3} \right|}. \tag{2.34}$$

Note that Eq. (2.34) gives the *peak* value (rather than the peak-to-peak value) of the input. Also denoted by $P_{1dB}$, the 1-dB compression point is typically in the range of $-20$ to $-25$ dBm (63.2 to 35.6 $\text{mV}_{pp}$ in 50-Ω system) at the input of RF receivers. We use the notations $A_{1dB}$ and $P_{1dB}$ interchangeably in this book. Whether they refer to the input or the output will be clear from the context or specified explicitly. While gain compression by 1 dB seems arbitrary, the 1-dB compression point represents a 10% reduction in the gain and is widely used to characterize RF circuits and systems.

Why does compression matter? After all, it appears that if a signal is so large as to reduce the gain of a receiver, then it must lie well above the receiver noise and be easily detectable. In fact, for some modulation schemes, this statement holds and compression of the receiver would seem benign. For example, as illustrated in Fig. 2.11(a), a frequency-modulated signal carries no information in its amplitude and hence tolerates compression (i.e., amplitude limiting). On the other hand, modulation schemes that contain information in the amplitude are distorted by compression [Fig. 2.11(b)]. This issue manifests itself in both receivers and transmitters.

Another adverse effect arising from compression occurs if a large *interferer* accompanies the received signal [Fig. 2.12(a)]. In the time domain, the small desired signal is superimposed on the large interferer. Consequently, the receiver gain is reduced by the large excursions produced by the interferer even though the desired signal itself is small

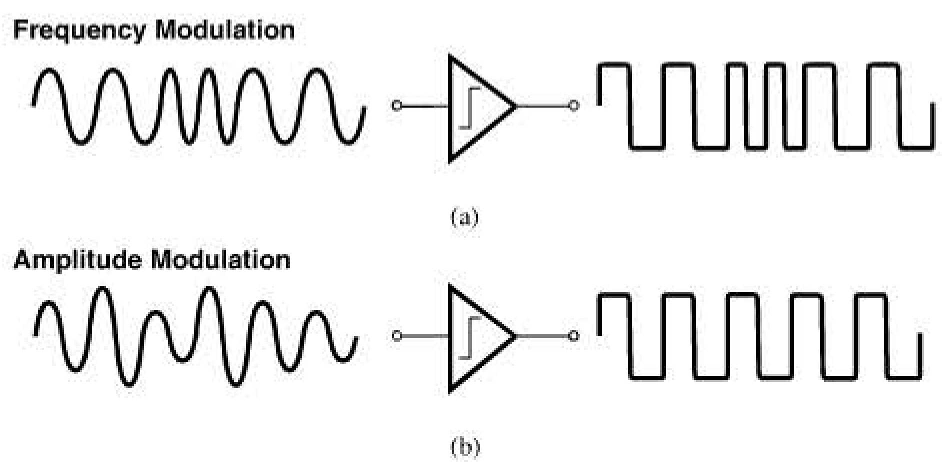



**Figure 2.11** *Effect of compressive nonlinearity on (a) FM and (b) AM waveforms.*



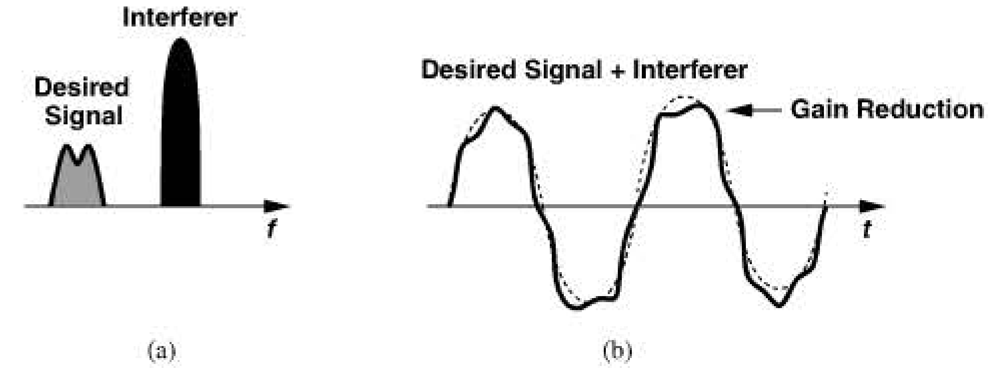



**Figure 2.12** *(a) Interferer accompanying signal, (b) effect in time domain.*

[Fig. 2.12(b)]. Called "desensitization," this phenomenon lowers the signal-to-noise ratio (SNR) at the receiver output and proves critical even if the signal contains no amplitude information.

To quantify desensitization, let us assume $x(t) = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t$, where the first and second terms represent the desired component and the interferer, respectively. With the third-order characteristic of Eq. (2.25), the output appears as

$$y(t) = \left( \alpha_1 + \frac{3}{4}\alpha_3 A_1^2 + \frac{3}{2}\alpha_3 A_2^2 \right) A_1 \cos \omega_1 t + \cdots. \tag{2.35}$$

Note that $\alpha_2$ is absent in compression. For $A_1 \ll A_2$, this reduces to

$$y(t) = \left( \alpha_1 + \frac{3}{2}\alpha_3 A_2^2 \right) A_1 \cos \omega_1 t + \cdots. \tag{2.36}$$

Thus, the gain experienced by the desired signal is equal to $\alpha_1 + 3\alpha_3 A_2^2/2$, a decreasing function of $A_2$ if $\alpha_1 \alpha_3 < 0$. In fact, for sufficiently large $A_2$, the gain drops to zero, and we say the signal is "blocked." In RF design, the term "blocking signal" or "blocker" refers to interferers that desensitize a circuit even if they do not reduce the gain to zero. Some RF receivers must be able to withstand blockers that are 60 to 70 dB greater than the desired signal.

### Example 2.7

A 900-MHz GSM transmitter delivers a power of 1 W to the antenna. By how much must the second harmonic of the signal be suppressed (filtered) so that it does not desensitize a 1.8-GHz receiver having $P_{1dB} = -25$ dBm? Assume the receiver is 1 m away (Fig. 2.13) and the 1.8-GHz signal is attenuated by 10 dB as it propagates across this distance.

*(Continues)*

**Example 2.7** (*Continued*)

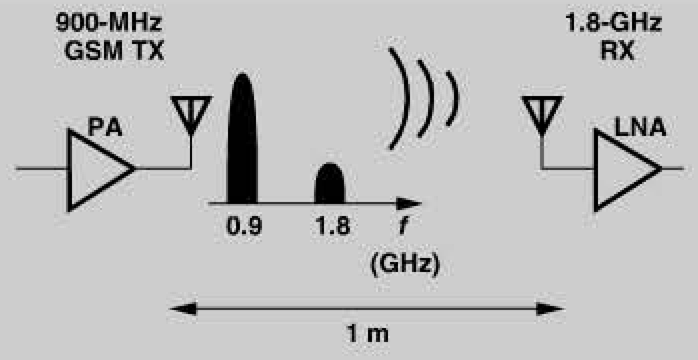



**Figure 2.13** *TX and RX in a cellular system.*

**Solution:**

The output power at 900 MHz is equal to +30 dBm. With an attenuation of 10 dB, the second harmonic must not exceed −15 dBm at the transmitter antenna so that it is below $P_{1dB}$ of the receiver. Thus, the second harmonic must remain at least 45 dB below the fundamental at the TX output. In practice, this interference must be another several dB lower to ensure the RX does not compress.

### 2.2.3 Cross Modulation

Another phenomenon that occurs when a weak signal and a strong interferer pass through a nonlinear system is the *transfer* of modulation from the interferer to the signal. Called "cross modulation," this effect is exemplified by Eq. (2.36), where variations in $A_2$ affect the amplitude of the signal at $\omega_1$. For example, suppose that the interferer is an amplitude-modulated signal, $A_2(1 + m\cos\omega_m t)\cos\omega_2 t$, where $m$ is a constant and $\omega_m$ denotes the modulating frequency. Equation (2.36) thus assumes the following form:

$$y(t) = \left[\alpha_1 + \frac{3}{2}\alpha_3 A_2^2\left(1 + \frac{m^2}{2} + \frac{m^2}{2}\cos 2\omega_m t + 2m\cos\omega_m t\right)\right] A_1\cos\omega_1 t + \cdots. \quad (2.37)$$

In other words, the desired signal at the output suffers from amplitude modulation at $\omega_m$ and $2\omega_m$. Figure 2.14 illustrates this effect.

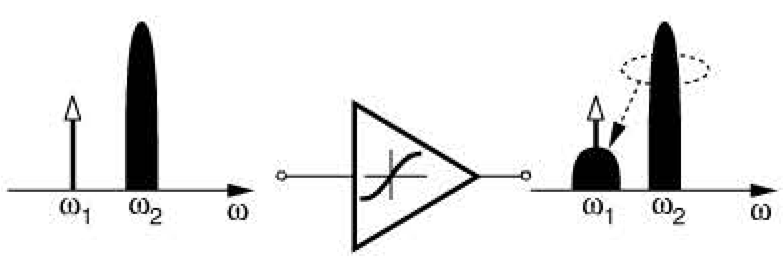



**Figure 2.14** *Cross modulation.*



**Example 2.8**

Suppose an interferer contains phase modulation but not amplitude modulation. Does cross modulation occur in this case?

**Solution:**

Expressing the input as $x(t) = A_1\cos\omega_1 t + A_2\cos(\omega_2 t + \phi)$, where the second term represents the interferer ($A_2$ is constant but $\phi$ varies with time), we use the third-order polynomial in Eq. (2.25) to write

$$y(t) = \alpha_1[A_1\cos\omega_1 t + A_2\cos(\omega_2 t + \phi)] + \alpha_2[A_1\cos\omega_1 t + A_2\cos(\omega_2 t + \phi)]^2 + \alpha_3[A_1\cos\omega_1 t + A_2\cos(\omega_2 t + \phi)]^3. \quad (2.38)$$

We now note that (1) the second-order term yields components at $\omega_1 \pm \omega_2$ but not at $\omega_1$; (2) the third-order term expansion gives $3\alpha_3 A_1\cos\omega_1 t A_2^2\cos^2(\omega_2 t + \phi)$, which, according to $\cos^2 x = (1 + \cos 2x)/2$, results in a component at $\omega_1$. Thus,

$$y(t) = \left(\alpha_1 + \frac{3}{2}\alpha_3 A_2^2\right) A_1\cos\omega_1 t + \cdots. \quad (2.39)$$

Interestingly, the desired signal at $\omega_1$ does not experience cross modulation. That is, phase-modulated interferers do not cause cross modulation in *memoryless* (static) nonlinear systems. Dynamic nonlinear systems, on the other hand, may not follow this rule.

Cross modulation commonly arises in amplifiers that must simultaneously process many independent signal channels. Examples include cable television transmitters and systems employing "orthogonal frequency division multiplexing" (OFDM). We examine OFDM in Chapter 3.

### 2.2.4 Intermodulation

Our study of nonlinearity has thus far considered the case of a single signal (for harmonic distortion) or a signal accompanied by one large interferer (for desensitization). Another scenario of interest in RF design occurs if *two* interferers accompany the desired signal. Such a scenario represents realistic situations and reveals nonlinear effects that may not manifest themselves in a harmonic distortion or desensitization test.

If two interferers at $\omega_1$ and $\omega_2$ are applied to a nonlinear system, the output generally exhibits components that are not harmonics of these frequencies. Called "intermodulation" (IM), this phenomenon arises from "mixing" (multiplication) of the two components as their sum is raised to a power greater than unity. To understand how Eq. (2.25) leads to intermodulation, assume $x(t) = A_1\cos\omega_1 t + A_2\cos\omega_2 t$. Thus,

$$y(t) = \alpha_1(A_1\cos\omega_1 t + A_2\cos\omega_2 t) + \alpha_2(A_1\cos\omega_1 t + A_2\cos\omega_2 t)^2 + \alpha_3(A_1\cos\omega_1 t + A_2\cos\omega_2 t)^3. \quad (2.40)$$

Expanding the right-hand side and discarding the dc terms, harmonics, and components at $\omega_1 \pm \omega_2$, we obtain the following "intermodulation products":

$$\omega = 2\omega_1 \pm \omega_2: \quad \frac{3\alpha_3 A_1^2 A_2}{4}\cos(2\omega_1 + \omega_2)t + \frac{3\alpha_3 A_1^2 A_2}{4}\cos(2\omega_1 - \omega_2)t \tag{2.41}$$

$$\omega = 2\omega_2 \pm \omega_1: \quad \frac{3\alpha_3 A_1 A_2^2}{4}\cos(2\omega_2 + \omega_1)t + \frac{3\alpha_3 A_1 A_2^2}{4}\cos(2\omega_2 - \omega_1)t \tag{2.42}$$

and these fundamental components:

$$\begin{aligned}\omega = \omega_1,\ \omega_2: &\left(\alpha_1 A_1 + \frac{3}{4}\alpha_3 A_1^3 + \frac{3}{2}\alpha_3 A_1 A_2^2\right)\cos\omega_1 t \\ &+ \left(\alpha_1 A_2 + \frac{3}{4}\alpha_3 A_2^3 + \frac{3}{2}\alpha_3 A_2 A_1^2\right)\cos\omega_2 t\end{aligned} \tag{2.43}$$

Figure 2.15 illustrates the results. Among these, the third-order IM products at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ are of particular interest. This is because, if $\omega_1$ and $\omega_2$ are close to each other, then $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ appear in the vicinity of $\omega_1$ and $\omega_2$. We now explain the significance of this statement.

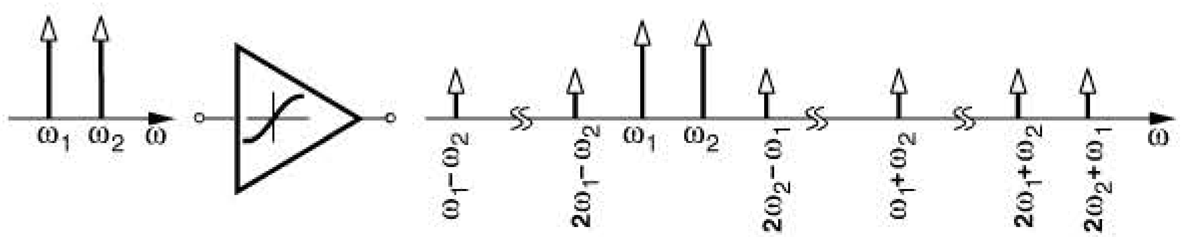



**Figure 2.15** *Generation of various intermodulation components in a two-tone test.*

Suppose an antenna receives a small desired signal at $\omega_0$ along with two large interferers at $\omega_1$ and $\omega_2$, providing this combination to a low-noise amplifier (Fig. 2.16). Let us assume that the interferer frequencies happen to satisfy $2\omega_1 - \omega_2 = \omega_0$. Consequently, the intermodulation product at $2\omega_1 - \omega_2$ falls onto the desired channel, corrupting the signal.

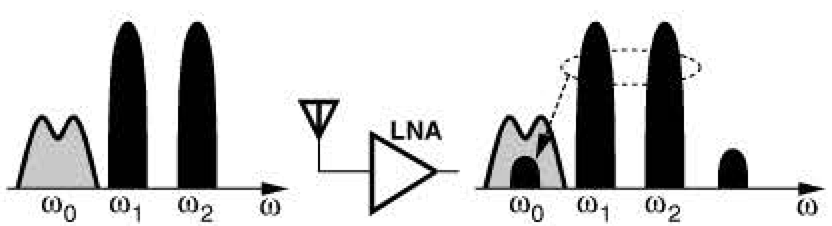



**Figure 2.16** *Corruption due to third-order intermodulation.*

**Example 2.9**

Suppose four Bluetooth users operate in a room as shown in Fig. 2.17. User 4 is in the receive mode and attempts to sense a weak signal transmitted by User 1 at 2.410 GHz.



**Example 2.9** *(Continued)*

At the same time, Users 2 and 3 transmit at 2.420 GHz and 2.430 GHz, respectively. Explain what happens.

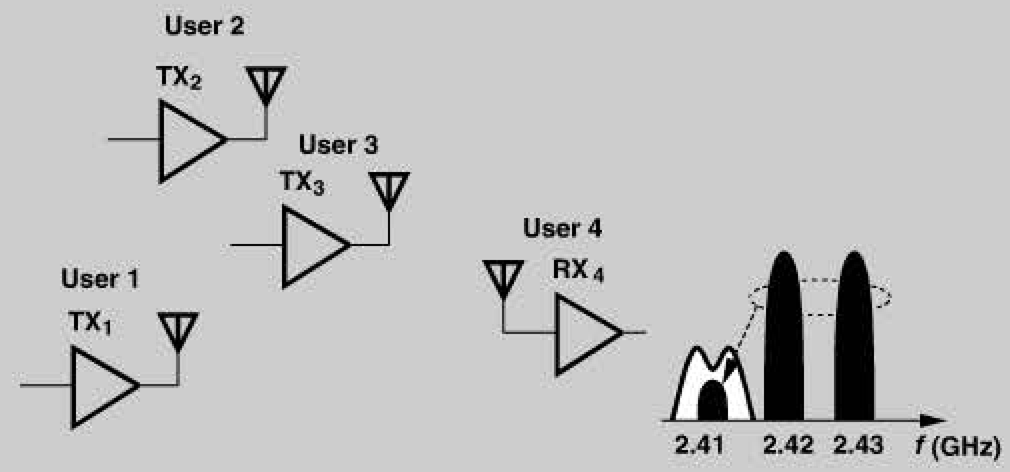



**Figure 2.17** *Bluetooth RX in the presence of several transmitters.*

**Solution:**

Since the frequencies transmitted by Users 1, 2, and 3 happen to be equally spaced, the intermodulation in the LNA of $RX_4$ corrupts the desired signal at 2.410 GHz.

The reader may raise several questions at this point: (1) In our analysis of intermodulation, we represented the interferers with pure (unmodulated) sinusoids (called "tones") whereas in Figs. 2.16 and 2.17, the interferers are modulated. Are these consistent? (2) Can gain compression and desensitization ($P_{1dB}$) also model intermodulation, or do we need other measures of nonlinearity? (3) Why can we not simply remove the interferers by filters so that the receiver does not experience intermodulation? We answer the first two here and address the third in Chapter 4.

For narrowband signals, it is sometimes helpful to "condense" their energy into an impulse, i.e., represent them with a tone of equal power [Fig. 2.18(a)]. This approximation must be made judiciously: if applied to study gain compression, it yields reasonably accurate results; on the other hand, if applied to the case of cross modulation, it fails. In intermodulation analyses, we proceed as follows: (a) approximate the interferers with tones, (b) calculate the level of intermodulation products at the output, and (c) mentally convert the intermodulation tones back to modulated components so as to see the corruption.[5] This thought process is illustrated in Fig. 2.18(b).

We now deal with the second question: if the gain is not compressed, then can we say that intermodulation is negligible? The answer is no; the following example illustrates this point.

5. Since a tone contains no randomness, it generally does not corrupt a signal. But a tone appearing in the spectrum of a signal may make the detection difficult.

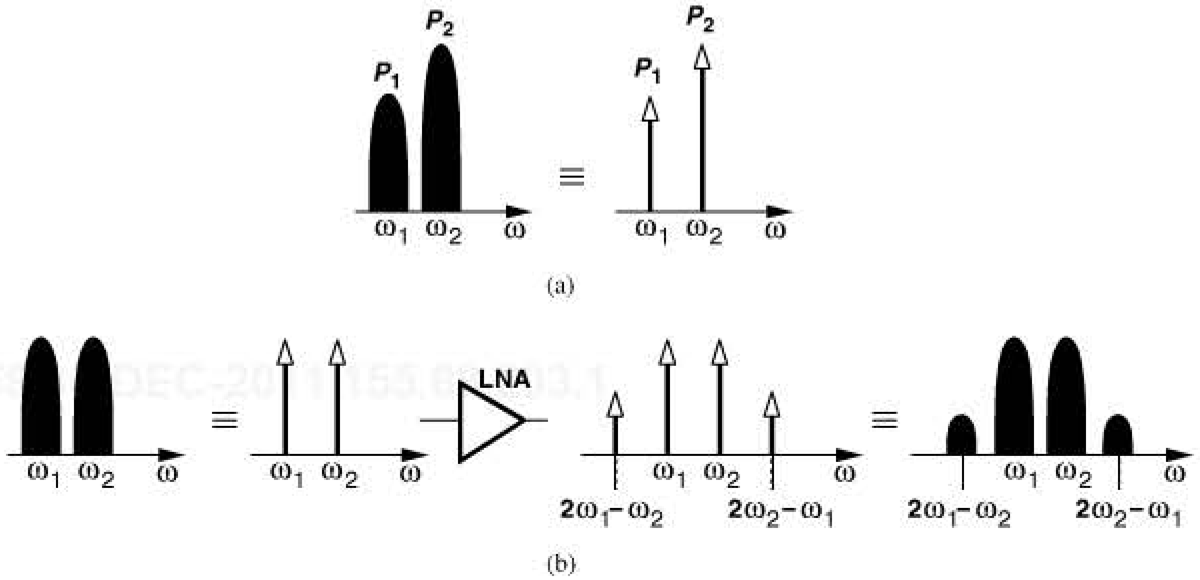



**Figure 2.18** *(a) Approximation of modulated signals by impulses, (b) application to intermodulation.*

### Example 2.10

A Bluetooth receiver employs a low-noise amplifier having a gain of 10 and an input impedance of 50 Ω. The LNA senses a desired signal level of −80 dBm at 2.410 GHz and two interferers of equal levels at 2.420 GHz and 2.430 GHz. For simplicity, assume the LNA drives a 50-Ω load.

(a) Determine the value of $\alpha_3$ that yields a $P_{1dB}$ of −30 dBm.

(b) If each interferer is 10 dB below $P_{1dB}$, determine the corruption experienced by the desired signal at the LNA output.

**Solution:**

(a) Noting that $-30\,\text{dBm} = 20\,\text{mV}_{\text{pp}} = 10\,\text{mV}_{\text{p}}$, from Eq. (2.34), we have $\sqrt{0.145|\alpha_1/\alpha_3|} = 10\,\text{mV}_{\text{p}}$. Since $\alpha_1 = 10$, we obtain $\alpha_3 = 14,500\,\text{V}^{-2}$.

(b) Each interferer has a level of −40 dBm ($=6.32\,\text{mV}_{\text{pp}}$). Setting $A_1 = A_2 = 6.32\,\text{mV}_{\text{pp}}/2$ in Eq. (2.41), we determine the amplitude of the IM product at 2.410 GHz as

$$\frac{3\alpha_3 A_1^2 A_2}{4} = 0.343\,\text{mV}_{\text{p}} = -59.3\,\text{dBm}. \tag{2.44}$$

The desired signal is amplified by a factor of $\alpha_1 = 10 = 20\,\text{dB}$, emerging at the output at a level of −60 dBm. Unfortunately, the IM product is as large as the signal itself even though the LNA does not experience significant compression.

The two-tone test is versatile and powerful because it can be applied to systems with arbitrarily narrow bandwidths. A sufficiently small difference between the two tone frequencies ensures that the IM products also fall within the band, thereby providing a



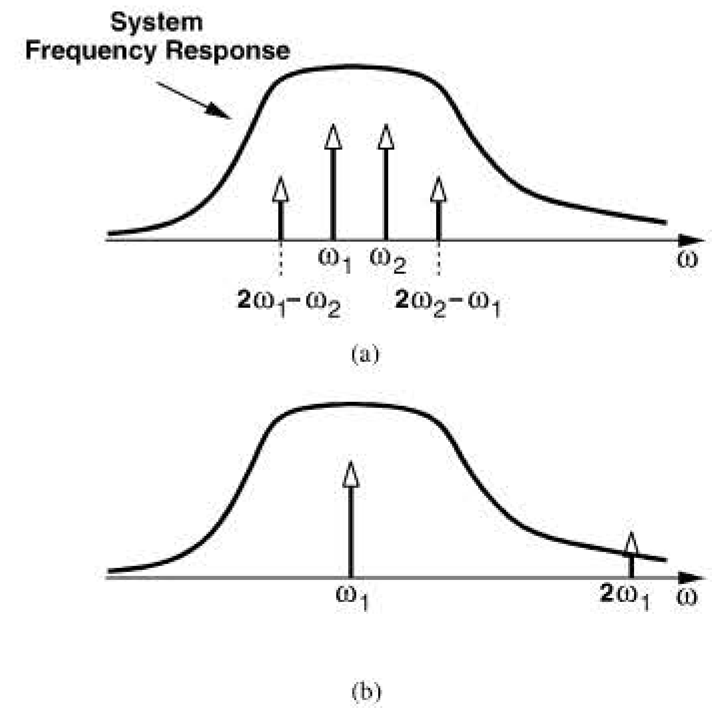



**Figure 2.19** *(a) Two-tone and (b) harmonic tests in a narrowband system.*

meaningful view of the nonlinear behavior of the system. Depicted in Fig. 2.19(a), this attribute stands in contrast to harmonic distortion tests, where higher harmonics lie so far away in frequency that they are heavily filtered, making the system appear quite linear [Fig. 2.19(b)].

**Third Intercept Point** Our thoughts thus far indicate the need for a measure of intermodulation. A common method of IM characterization is the "two-tone" test, whereby two pure sinusoids of *equal* amplitudes are applied to the input. The amplitude of the output IM products is then normalized to that of the fundamentals at the output. Denoting the peak amplitude of each tone by $A$, we can write the result as

$$\text{Relative IM} = 20\log\left(\frac{3}{4}\frac{\alpha_3}{\alpha_1}A^2\right)\ \text{dBc}, \tag{2.45}$$

where the unit dBc denotes decibels with respect to the "carrier" to emphasize the normalization. Note that, if the amplitude of each input tone increases by 6 dB (a factor of two), the amplitude of the IM products ($\propto A^3$) rises by 18 dB and hence the *relative* IM by 12 dB.[6]

The principal difficulty in specifying the relative IM for a circuit is that it is meaningful only if the value of $A$ is given. From a practical point of view, we prefer a *single* measure that captures the intermodulation behavior of the circuit with no need to know the input level at which the two-tone test is carried out. Fortunately, such a measure exists and is called the "third intercept point" ($\text{IP}_3$).

The concept of $\text{IP}_3$ originates from our earlier observation that, if the amplitude of each tone rises, that of the output IM products increases more sharply ($\propto A^3$). Thus, if we continue to raise $A$, the amplitude of the IM products eventually becomes *equal* to that

6. It is assumed that no compression occurs so that the output fundamental tones also rise by 6 dB.

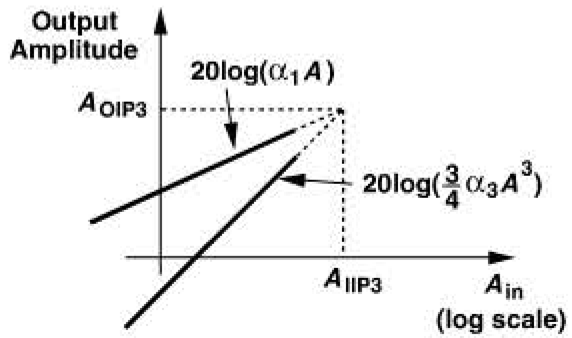



**Figure 2.20** *Definition of IP$_3$ (for voltage quantities).*

of the fundamental tones at the output. As illustrated in Fig. 2.20 on a log-log scale, the input level at which this occurs is called the "input third intercept point" (IIP$_3$). Similarly, the corresponding output is represented by OIP$_3$. In subsequent derivations, we denote the input amplitude as $A_{IIP3}$.

To determine the IIP$_3$, we simply equate the fundamental and IM amplitudes:

$$|\alpha_1 A_{IIP3}| = \left|\frac{3}{4}\alpha_3 A_{IIP3}^3\right|, \tag{2.46}$$

obtaining

$$A_{IIP3} = \sqrt{\frac{4}{3}\left|\frac{\alpha_1}{\alpha_3}\right|}. \tag{2.47}$$

Interestingly,

$$\frac{A_{IIP3}}{A_{1dB}} = \sqrt{\frac{4}{0.435}} \tag{2.48}$$

$$\approx 9.6\,\text{dB}. \tag{2.49}$$

This ratio proves helpful as a sanity check in simulations and measurements.[7] We sometimes write IP$_3$ rather than IIP$_3$ if it is clear from the context that the input is of interest.

Upon further consideration, the reader may question the consistency of the above derivations. If the IP$_3$ is 9.6 dB *higher* than $P_{1dB}$, is the gain not heavily compressed at $A_{in} = A_{IIP3}$?! If the gain is compressed, why do we still express the amplitude of the fundamentals at the output as $\alpha_1 A$? It appears that we must instead write this amplitude as $[\alpha_1 + (9/4)\alpha_3 A^2]A$ to account for the compression.

In reality, the situation is even more complicated. The value of IP$_3$ given by Eq. (2.47) may *exceed* the supply voltage, indicating that higher-order nonlinearities manifest themselves as $A_{in}$ approaches $A_{IIP3}$ [Fig. 2.21(a)]. In other words, the IP$_3$ is not a directly measureable quantity.

In order to avoid these quandaries, we measure the *IP$_3$* as follows. We begin with a very low input level so that $\alpha_1 + (9/4)\alpha_3 A_{in}^2 \approx \alpha_1$ (and, of course, higher order nonlinearities are also negligible). We increase $A_{in}$, plot the amplitudes of the fundamentals and the IM products on a log-log scale, and *extrapolate* these plots according to their slopes (one and three, respectively) to obtain the IP$_3$ [Fig. 2.21(b)]. To ensure that the signal levels remain well below compression and higher-order terms are negligible, we must observe a 3-dB rise in the IM products for every 1-dB increase in $A_{in}$. On the other hand, if $A_{in}$ is excessively small, then the output IM components become comparable with the noise floor of the circuit (or the noise floor of the simulated spectrum), thus leading to inaccurate results.

7. Note that this relationship holds for a third-order system and not necessarily if higher-order terms manifest themselves.



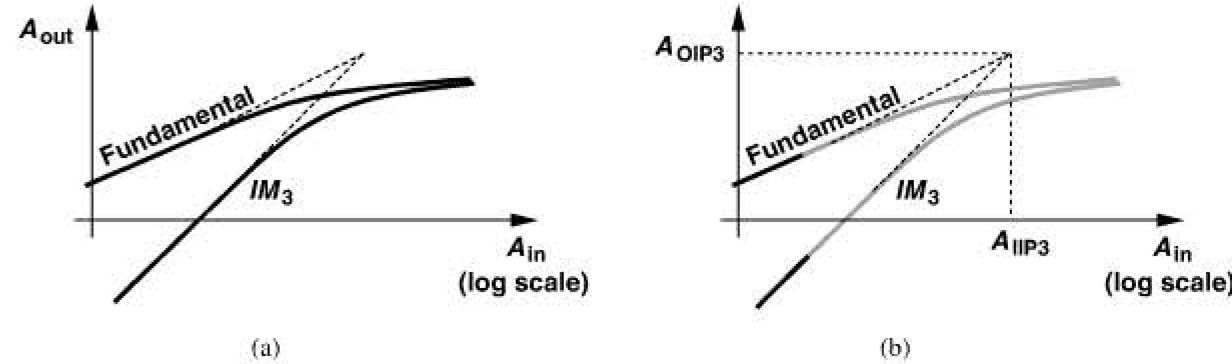



**Figure 2.21** *(a) Actual behavior of nonlinear circuits, (b) definition of IP$_3$ based on extrapolation.*

### Example 2.11

A low-noise amplifier senses a −80-dBm signal at 2.410 GHz and two −20-dBm interferers at 2.420 GHz and 2.430 GHz. What IIP$_3$ is required if the IM products must remain 20 dB below the signal? For simplicity, assume 50-Ω interfaces at the input and output.

**Solution:**

Denoting the peak amplitudes of the signal and the interferers by $A_{sig}$ and $A_{int}$, respectively, we can write at the LNA output:

$$20\log|\alpha_1 A_{sig}| - 20\,\text{dB} = 20\log\left|\frac{3}{4}\alpha_3 A_{int}^3\right|. \tag{2.50}$$

It follows that

$$|\alpha_1 A_{sig}| = \left|\frac{30}{4}\alpha_3 A_{int}^3\right|. \tag{2.51}$$

In a 50-Ω system, the −80-dBm and −20-dBm levels respectively yield $A_{sig} = 31.6\ \mu V_p$ and $A_{int} = 31.6\ mV_p$. Thus,

$$\text{IIP}_3 = \sqrt{\frac{4}{3}\left|\frac{\alpha_1}{\alpha_3}\right|} \tag{2.52}$$

$$= 3.65\ V_p \tag{2.53}$$

$$= +15.2\,\text{dBm}. \tag{2.54}$$

Such an IP$_3$ is extremely difficult to achieve, especially for a complete receiver chain.

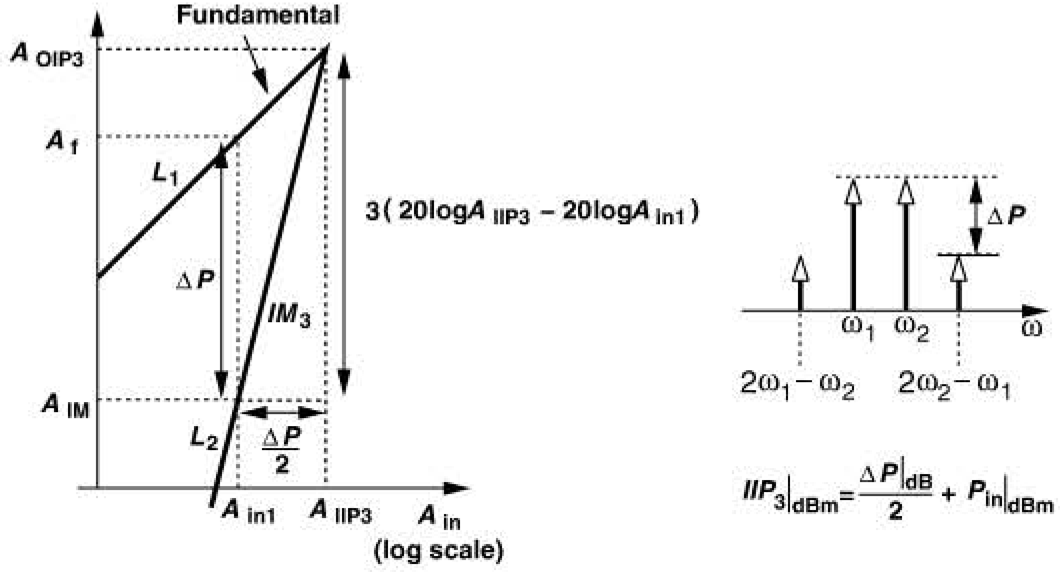



**Figure 2.22** *(a) Relationships among various power levels in a two-tone test, (b) illustration of shortcut technique.*

Since extrapolation proves quite tedious in simulations or measurements, we often employ a shortcut that provides a reasonable initial estimate. As illustrated in Fig. 2.22(a), suppose *hypothetically* that the input is equal to $A_{IIP3}$, and hence the (extrapolated) output IM products are as large as the (extrapolated) fundamental tones. Now, the input is reduced to a level $A_{in1}$. That is, the change in the input is equal to $20\log A_{IIP3} - 20\log A_{in1}$. On a log-log scale, the IM products fall with a slope of 3 and the fundamentals with a slope of unity. Thus, the *difference* between the two plots increases with a slope of 2. We denote $20\log A_f - 20\log A_{IM}$ by $\Delta P$ and write

$$\Delta P = 20\log A_f - 20\log A_{IM} = 2(20\log A_{IIP3} - 20\log A_{in1}), \tag{2.55}$$

obtaining

$$20\log A_{IIP3} = \frac{\Delta P}{2} + 20\log A_{in1}. \tag{2.56}$$

In other words, for a given input level (well below $P_{1dB}$), the $IIP_3$ can be calculated by halving the difference between the output fundamental and IM levels and adding the result to the input level, where all values are expressed as logarithmic quantities. Figure 2.22(b) depicts an abbreviated notation for this rule. The key point here is that the $IP_3$ is measured without extrapolation.

Why do we consider the above result an *estimate*? After all, the derivation assumes third-order nonlinearity. A difficulty arises if the circuit contains *dynamic* nonlinearities, in which case this result may deviate from that obtained by extrapolation. The latter is the standard and accepted method for measuring and reporting the $IP_3$, but the shortcut method proves useful in understanding the behavior of the device under test.



We should remark that *second-order* nonlinearity also leads to a certain type of intermodulation and is characterized by a "second intercept point," ($IP_2$).[8] We elaborate on this effect in Chapter 4.

### 2.2.5 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. The calculation of $P_{1dB}$ for a cascade is outlined in Problem 2.1. Here, we determine the $IP_3$ of a cascade. For the sake of brevity, we hereafter denote the input $IP_3$ by $A_{IP3}$ unless otherwise noted.

Consider two nonlinear stages in cascade (Fig. 2.23). If the input/output characteristics of the two stages are expressed, respectively, as

$$y_1(t) = \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \tag{2.57}$$

$$y_2(t) = \beta_1 y_1(t) + \beta_2 y_1^2(t) + \beta_3 y_1^3(t), \tag{2.58}$$

then

$$\begin{aligned} y_2(t) = {} & \beta_1[\alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t)] + \beta_2[\alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t)]^2 \\ & + \beta_3[\alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t)]^3. \end{aligned} \tag{2.59}$$

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

$$y_2(t) = \alpha_1\beta_1 x(t) + (\alpha_3\beta_1 + 2\alpha_1\alpha_2\beta_2 + \alpha_1^3\beta_3)x^3(t) + \cdots. \tag{2.60}$$

Thus, from Eq. (2.47),

$$A_{IP3} = \sqrt{\frac{4}{3}\left|\frac{\alpha_1\beta_1}{\alpha_3\beta_1 + 2\alpha_1\alpha_2\beta_2 + \alpha_1^3\beta_3}\right|}. \tag{2.61}$$

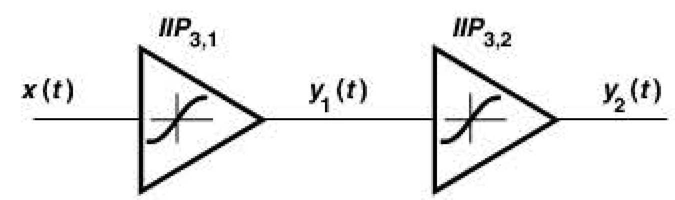



**Figure 2.23** *Cascaded nonlinear stages.*

**Example 2.12**

Two differential pairs are cascaded. Is it possible to select the denominator of Eq. (2.61) such that $IP_3$ goes to infinity?

*(Continues)*

8. As seen in the next section, second-order nonlinearity also affects the $IP_3$ in cascaded systems.

**Example 2.12** (*Continued*)

**Solution:**

With no asymmetries in the cascade, $\alpha_2 = \beta_2 = 0$. Thus, we seek the condition $\alpha_3\beta_1 + \alpha_1^3\beta_3 = 0$, or equivalently,

$$\frac{\alpha_3}{\alpha_1} = -\frac{\beta_3}{\beta_1} \cdot \alpha_1^2. \tag{2.62}$$

Since both stages are compressive, $\alpha_3/\alpha_1 < 0$ and $\beta_3/\beta_1 < 0$. It is therefore impossible to achieve an arbitrarily high $IP_3$.

Equation (2.61) leads to more intuitive results if its two sides are squared and inverted:

$$\frac{1}{A_{IP3}^2} = \frac{3}{4}\left|\frac{\alpha_3\beta_1 + 2\alpha_1\alpha_2\beta_2 + \alpha_1^3\beta_3}{\alpha_1\beta_1}\right| \tag{2.63}$$

$$= \frac{3}{4}\left|\frac{\alpha_3}{\alpha_1} + \frac{2\alpha_2\beta_2}{\beta_1} + \frac{\alpha_1^2\beta_3}{\beta_1}\right| \tag{2.64}$$

$$= \left|\frac{1}{A_{IP3,1}^2} + \frac{3\alpha_2\beta_2}{2\beta_1} + \frac{\alpha_1^2}{A_{IP3,2}^2}\right|, \tag{2.65}$$

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

The key observation in Eq. (2.65) is that to "refer" the $IP_3$ of the second stage to the input of the cascade, we must divide it by $\alpha_1$. Thus, the higher the gain of the first stage, the more nonlinearity is contributed by the second stage.

**IM Spectra in a Cascade** To gain more insight into the above results, let us assume $x(t) = A\cos\omega_1 t + A\cos\omega_2 t$ and identify the IM products in a cascade. With the aid of Fig. 2.24, we make the following observations:[9]

1. The input tones are amplified by a factor of approximately $\alpha_1$ in the first stage and $\beta_1$ in the second. Thus, the output fundamentals are given by $\alpha_1\beta_1 A(\cos\omega_1 t + \cos\omega_2 t)$.
2. The IM 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 amplified by a factor of $\beta_1$ when they appear at the output of the second stage.
3. Sensing $\alpha_1 A(\cos\omega_1 t + \cos\omega_2 t)$ at its input, the second stage produces its own IM components: $(3\beta_3/4)(\alpha_1 A)^3\cos(2\omega_1 - \omega_2)t + (3\beta_3/4)(\alpha_1 A)^3\cos(2\omega_2 - \omega_1)t$.



9. The spectrum of $A\cos\omega t$ consists of two impulses, each with a weight of $A/2$. We drop the factor of 1/2 in the figures for simplicity.



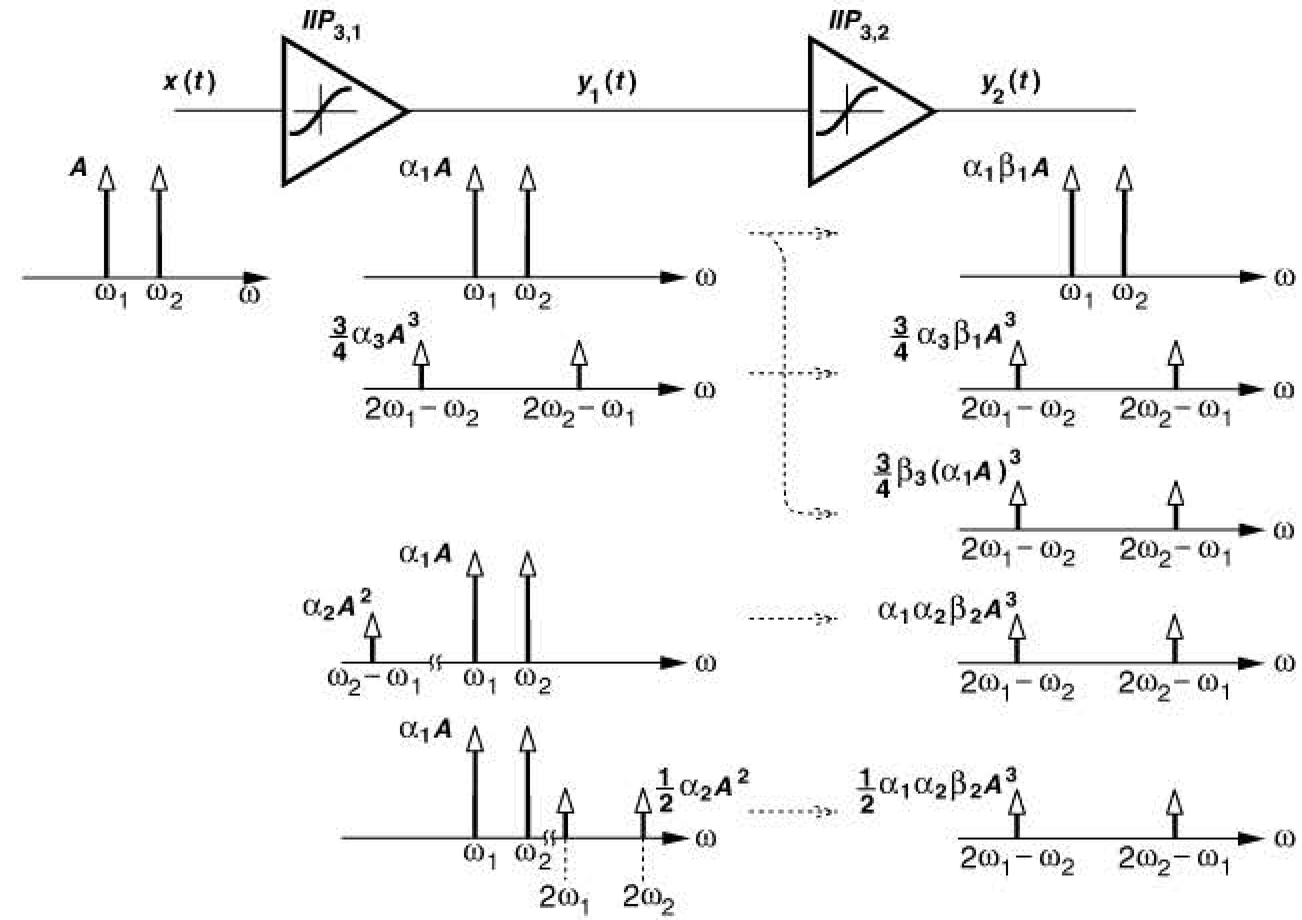



**Figure 2.24** *Spectra in a cascade of nonlinear stages.*

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, these components are mixed with those at $\omega_1$ and $\omega_2$ and translated to $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$. Specifically, as shown in Fig. 2.24, $y_2(t)$ contains terms such as $2\beta_2[\alpha_1 A\cos\omega_1 t \times \alpha_2 A^2\cos(\omega_1 - \omega_2)t]$ and $2\beta_2(\alpha_1 A\cos\omega_1 t \times 0.5\alpha_2 A^2\cos 2\omega_2 t)$. The resulting IM products can be expressed as $(3\alpha_1\alpha_2\beta_2 A^3/2)[\cos(2\omega_1 - \omega_2)t + \cos(2\omega_2 - \omega_1)t]$. Interestingly, the cascade of two *second-order* nonlinearities can produce *third-order* IM products.

Adding the amplitudes of the IM products, we have

$$\begin{aligned} y_2(t) &= \alpha_1\beta_1 A(\cos\omega_1 t + \cos\omega_2 t) \\ &\quad + \left(\frac{3\alpha_3\beta_1}{4} + \frac{3\alpha_1^3\beta_3}{4} + \frac{3\alpha_1\alpha_2\beta_2}{2}\right)A^3[\cos(\omega_1 - 2\omega_2)t \\ &\quad + \cos(2\omega_2 - \omega_1)t] + \cdots, \end{aligned} \tag{2.66}$$

obtaining the same $IP_3$ as above. This result assumes zero phase shift for all components.

Why did we add the amplitudes of the $IM_3$ products in Eq. (2.66) without regard for their phases? Is it possible that phase shifts in the first and second stages allow partial *cancellation* of these terms and hence a higher $IP_3$? Yes, it is possible but uncommon in practice. Since the frequencies $\omega_1$, $\omega_2$, $2\omega_1 - \omega_2$, and $2\omega_2 - \omega_1$ are close to one another, these components experience approximately equal phase shifts.

But how about the terms described in the fourth observation? Components such as $\omega_1 - \omega_2$ and $2\omega_1$ may fall well out of the signal band and experience phase shifts different from those in the first three observations. For this reason, we may consider Eqs. (2.65) and (2.66) as the worst-case scenario. Since most RF systems incorporate narrowband circuits, the terms at $\omega_1 \pm \omega_2$, $2\omega_1$, and $2\omega_2$ are heavily attenuated at the output of the first stage. Consequently, the second term on the right-hand side of (2.65) becomes negligible, and

$$\frac{1}{A_{IP3}^2} \approx \frac{1}{A_{IP3,1}^2} + \frac{\alpha_1^2}{A_{IP3,2}^2}. \tag{2.67}$$

Extending this result to three or more stages, we have

$$\frac{1}{A_{IP3}^2} \approx \frac{1}{A_{IP3,1}^2} + \frac{\alpha_1^2}{A_{IP3,2}^2} + \frac{\alpha_1^2\beta_1^2}{A_{IP3,3}^2} + \cdots. \tag{2.68}$$

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 equivalently scaled *down* by the total gain preceding that stage.

**Example 2.13**

A low-noise amplifier having an input $IP_3$ of $-10$ dBm and a gain of 20 dB is followed by a mixer with an input $IP_3$ of $+4$ dBm. Which stage limits the $IP_3$ of the cascade more? Assume the conceptual picture shown in Fig. 2.1(b) to go between volts and dBm's.

**Solution:**

With $\alpha_1 = 20$ dB, we note that

$$A_{IP3,1} = -10\,\text{dBm} \tag{2.69}$$

$$\frac{A_{IP3,2}}{\alpha_1} = -16\,\text{dBm}. \tag{2.70}$$

Since the scaled $IP_3$ of the second stage is lower than the $IP_3$ of the first stage, we say the second stage limits the overall $IP_3$ more.

In the simulation of a cascade, it is possible to determine which stage limits the linearity more. As depicted in Fig. 2.25, we examine the relative IM magnitudes at the output of each stage ($\Delta_1$ and $\Delta_2$, expressed in dB.) If $\Delta_2 \approx \Delta_1$, the second stage contributes negligible nonlinearity. On the other hand, if $\Delta_2$ is substantially less than $\Delta_1$, then the second stage limits the $IP_3$.



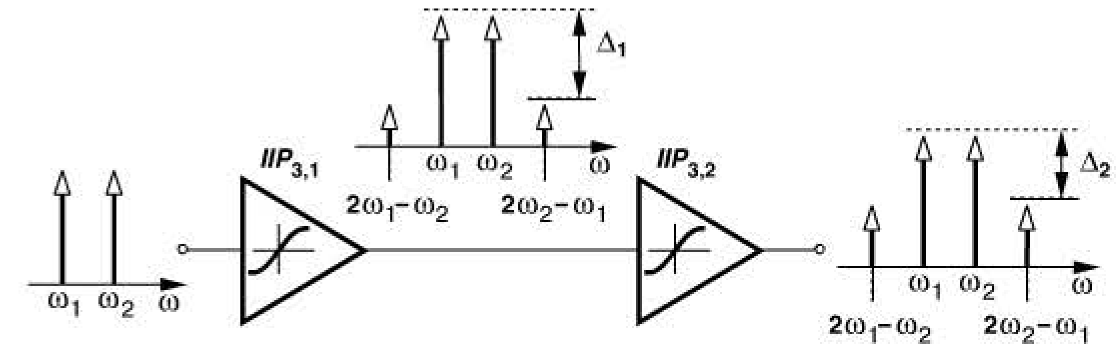



**Figure 2.25** *Growth of IM components along the cascade.*

### 2.2.6 AM/PM Conversion

In some RF circuits, e.g., power amplifiers, amplitude modulation (AM) may be converted to phase modulation (PM), thus producing undesirable effects. In this section, we study this phenomenon.

AM/PM conversion (APC) can be viewed as the dependence of the phase shift upon the signal amplitude. That is, for an input $V_{in}(t) = V_1 \cos\omega_1 t$, the fundamental output component is given by

$$V_{out}(t) = V_2 \cos[\omega_1 t + \phi(V_1)], \tag{2.71}$$

where $\phi(V_1)$ denotes the amplitude-dependent phase shift. This, of course, does not occur in a linear time-invariant system. For example, the phase shift experienced by a sinusoid of frequency $\omega_1$ through a first-order low-pass RC section is given by $-\tan^{-1}(RC\omega_1)$ regardless of the amplitude. Moreover, APC does not appear in a memoryless nonlinear system because the phase shift is zero in this case.

We may therefore surmise that AM/PM conversion arises if a system is both dynamic and nonlinear. For example, if the capacitor in a first-order low-pass RC section is nonlinear, then its "average" value may depend on $V_1$, resulting in a phase shift, $-\tan^{-1}(RC\omega_1)$, that itself varies with $V_1$. To explore this point, let us consider the arrangement shown in Fig. 2.26 and assume

$$C_1 = (1 + \alpha V_{out})C_0. \tag{2.72}$$

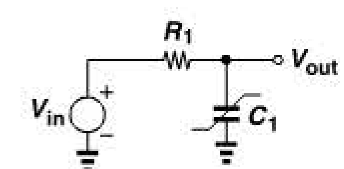



**Figure 2.26** *RC section with nonlinear capacitor.*

This capacitor is considered nonlinear because its value depends on its voltage. An exact calculation of the phase shift is difficult here as it requires that we write $V_{in} = R_1 C_1 dV_{out}/dt + V_{out}$ and hence solve

$$V_1 \cos\omega_1 t = R_1(1 + \alpha V_{out})C_0 \frac{dV_{out}}{dt} + V_{out}. \tag{2.73}$$

We therefore make an approximation. Since the value of $C_1$ varies *periodically* with time, we can express the output as that of a first-order network but with a time-varying capacitance, $C_1(t)$:

$$V_{out}(t) \approx \frac{V_1}{\sqrt{1 + R_1^2 C_1^2(t)\omega_1^2}} \cos\{\omega_1 t - \tan^{-1}[R_1 C_1(t)\omega_1]\}. \tag{2.74}$$

If $R_1 C_1(t)\omega_1 \ll 1$ rad,

$$V_{out}(t) \approx V_1 \cos[\omega_1 t - R_1(1 + \alpha V_{out})C_0\omega_1]. \tag{2.75}$$

We also assume that $(1 + \alpha V_{out})C_0 \approx (1 + \alpha V_1 \cos \omega_1 t)C_0$, obtaining

$$V_{out}(t) \approx V_1 \cos(\omega_1 t - R_1 C_0 \omega_1 - \alpha R_1 C_0 \omega_1 V_1 \cos \omega_1 t). \tag{2.76}$$

Does the output *fundamental* contain an input-dependent phase shift here? No, it does not! The reader can show that the third term inside the parentheses produces only higher *harmonics*. Thus, the phase shift of the fundamental is equal to $-R_1 C_0 \omega_1$ and hence constant.

The above example entails no AM/PM conversion because of the *first-order* dependence of $C_1$ upon $V_{out}$. As illustrated in Fig. 2.27, the average value of $C_1$ is equal to $C_0$ regardless of the output amplitude. In general, since $C_1$ varies periodically, it can be expressed as a Fourier series with a "dc" term representing its average value:

$$C_1(t) = C_{avg} + \sum_{n=1}^{\infty} a_n \cos(n\omega_1 t) + \sum_{n=1}^{\infty} b_n \sin(n\omega_1 t). \tag{2.77}$$

Thus, if $C_{avg}$ is a function of the amplitude, then the phase shift of the fundamental component in the output voltage becomes input-dependent. The following example illustrates this point.

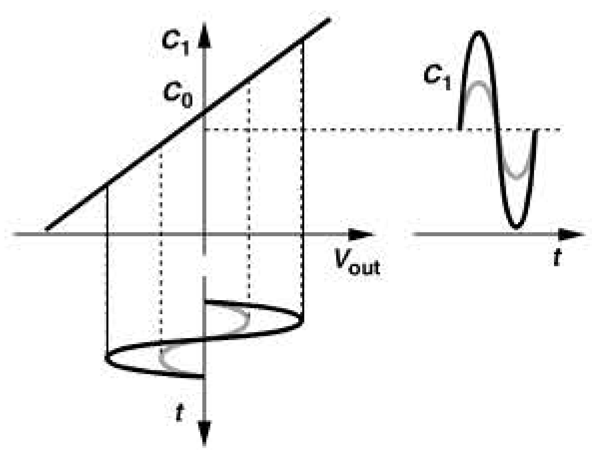



**Figure 2.27** *Time variation of capacitor with first-order voltage dependence for small and large swings.*



**Example 2.14**

Suppose $C_1$ in Fig. 2.26 is expressed as $C_1 = C_0(1 + \alpha_1 V_{out} + \alpha_2 V_{out}^2)$. Study the AM/PM conversion in this case if $V_{in}(t) = V_1 \cos \omega_1 t$.

**Solution:**

Figure 2.28 plots $C_1(t)$ for small and large input swings, revealing that $C_{avg}$ indeed depends on the amplitude. We rewrite Eq. (2.75) as

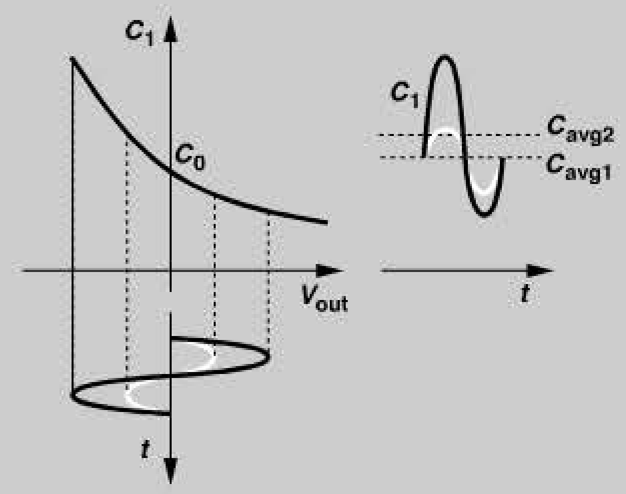



**Figure 2.28** *Time variation of capacitor with second-order voltage dependence for small and large swings.*

$$V_{out}(t) \approx V_1 \cos[\omega_1 t - R_1 C_0 \omega_1(1 + \alpha_1 V_1 \cos \omega_1 t + \alpha_2 V_1^2 \cos^2 \omega_1 t)] \tag{2.78}$$

$$\approx V_1 \cos\left(\omega_1 t - R_1 C_0 \omega_1 - \frac{\alpha_2 R_1 C_0 \omega_1 V_1^2}{2} - \cdots\right). \tag{2.79}$$

The phase shift of the fundamental now contains an input-dependent term, $-(\alpha_2 R_1 C_0 \omega_1 V_1^2)/2$. Figure 2.28 also suggests that AM/PM conversion does not occur if the capacitor voltage dependence is odd-symmetric.

What is the effect of APC? In the presence of APC, amplitude modulation (or amplitude noise) corrupts the phase of the signal. For example, if $V_{in}(t) = V_1(1 + m \cos \omega_m t)\cos \omega_1 t$, then Eq. (2.79) yields a phase corruption equal to $-\alpha_2 R_1 C_0 \omega_1 (2mV_1 \cos \omega_m t + m^2 V_1^2 \cos^2 \omega_m t)/2$. We will encounter examples of APC in Chapters 8 and 12.

## 2.3 NOISE

The performance of RF systems is limited by noise. Without noise, an RF receiver would be able to detect arbitrarily small inputs, allowing communication across arbitrarily long distances. In this section, we review basic properties of noise and methods of calculating noise in circuits. For a more complete study of noise in analog circuits, the reader is referred to [1].

### 2.3.1 Noise as a Random Process

The trouble with noise is that it is random. Engineers who are used to dealing with well-defined, deterministic, "hard" facts often find the concept of randomness difficult to grasp, especially if it must be incorporated mathematically. To overcome this fear of randomness, we approach the problem from an intuitive angle.

By "noise is random," we mean the instantaneous value of noise cannot be predicted. For example, consider a resistor tied to a battery and carrying a current [Fig. 2.29(a)]. Due to the ambient temperature, each electron carrying the current experiences thermal agitation, thus following a somewhat random path while, on the average, moving toward the positive terminal of the battery. As a result, the *average* current remains equal to $V_B/R$ but the instantaneous current displays random values.[10]

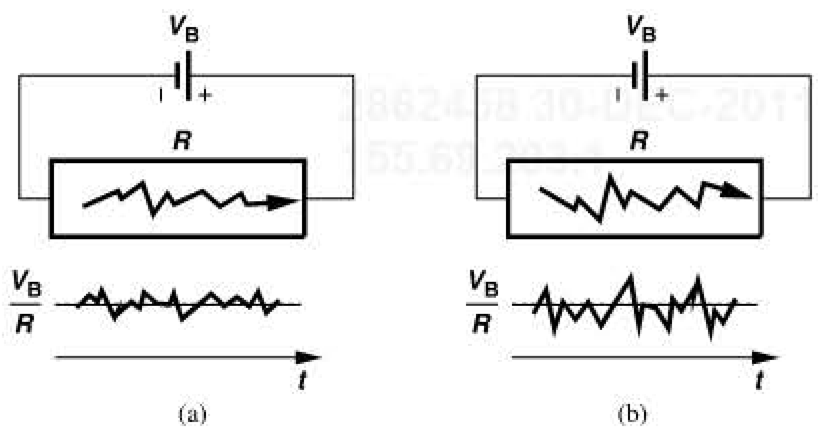



**Figure 2.29** *(a) Noise generated in a resistor, (b) effect of higher temperature.*

Since noise cannot be characterized in terms of instantaneous voltages or currents, we seek other attributes of noise that are predictable. For example, we know that a higher ambient temperature leads to greater thermal agitation of electrons and hence larger fluctuations in the current [Fig. 2.29(b)]. How do we express the concept of larger random swings for a current or voltage quantity? This property is revealed by the *average power* of the noise, defined, in analogy with periodic signals, as

$$P_n = \lim_{T\to\infty} \frac{1}{T}\int_0^T n^2(t)dt, \tag{2.80}$$

where $n(t)$ represents the noise waveform. Illustrated in Fig. 2.30, this definition simply means that we compute the area under $n^2(t)$ for a long time, $T$, and normalize the result to $T$, thus obtaining the average power. For example, the two scenarios depicted in Fig. 2.29 yield different average powers.

10. As explained later, this is true even with a zero average current.



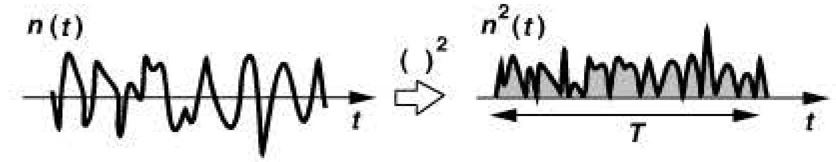



**Figure 2.30** *Computation of noise power.*

If $n(t)$ is random, how do we know that $P_n$ is not?! We are fortunate that noise components in circuits have a constant average power. For example, $P_n$ is known and constant for a resistor at a constant ambient temperature.

How long should $T$ in Eq. (2.80) be? Due to its randomness, noise consists of different frequencies. Thus, $T$ must be long enough to accommodate several cycles of the *lowest* frequency. For example, the noise in a crowded restaurant arises from human voice and covers the range of 20 Hz to 20 kHz, requiring that $T$ be on the order of 0.5 s to capture about 10 cycles of the 20-Hz components.[11]

### 2.3.2 Noise Spectrum

Our foregoing study suggests that the time-domain view of noise provides limited information, e.g., the average power. The frequency-domain view, on the other hand, yields much greater insight and proves more useful in RF design.

The reader may already have some intuitive understanding of the concept of "spectrum." We say the spectrum of human voice spans the range of 20 Hz to 20 kHz. This means that if we somehow measure the frequency content of the voice, we observe all components from 20 Hz to 20 kHz. How, then, do we measure a signal's frequency content, e.g., the strength of a component at 10 kHz? We would need to filter out the remainder of the spectrum and measure the *average power* of the 10-kHz component. Figure 2.31(a) conceptually illustrates such an experiment, where the microphone signal is applied to a band-pass filter having a 1-Hz bandwidth centered around 10 kHz. If a person speaks into the microphone at a steady volume, the power meter reads a constant value.

The scheme shown in Fig. 2.31(a) can be readily extended so as to measure the strength of all frequency components. As depicted in Fig. 2.31(b), a bank of 1-Hz band-pass filters centered at $f_1 \cdots f_n$ measures the average power at each frequency.[12] Called the spectrum or the "power spectral density" (PSD) of $x(t)$ and denoted by $S_x(f)$, the resulting plot displays the average power that the voice (or the noise) carries in a 1-Hz bandwidth at different frequencies.[13]

It is interesting to note that the total area under $S_x(f)$ represents the average power carried by $x(t)$:

$$\int_0^\infty S_x(f)df = \lim_{T\to\infty} \frac{1}{T}\int_0^T x^2(t)dt. \tag{2.81}$$

11. In practice, we make a guess for $T$, calculate $P_n$, increase $T$, recalculate $P_n$, and repeat until consecutive values of $P_n$ become nearly equal.
12. This is also the conceptual operation of spectrum analyzers.
13. In the theory of signals and systems, the PSD is defined as the Fourier transform of the autocorrelation of a signal. These two views are equivalent.

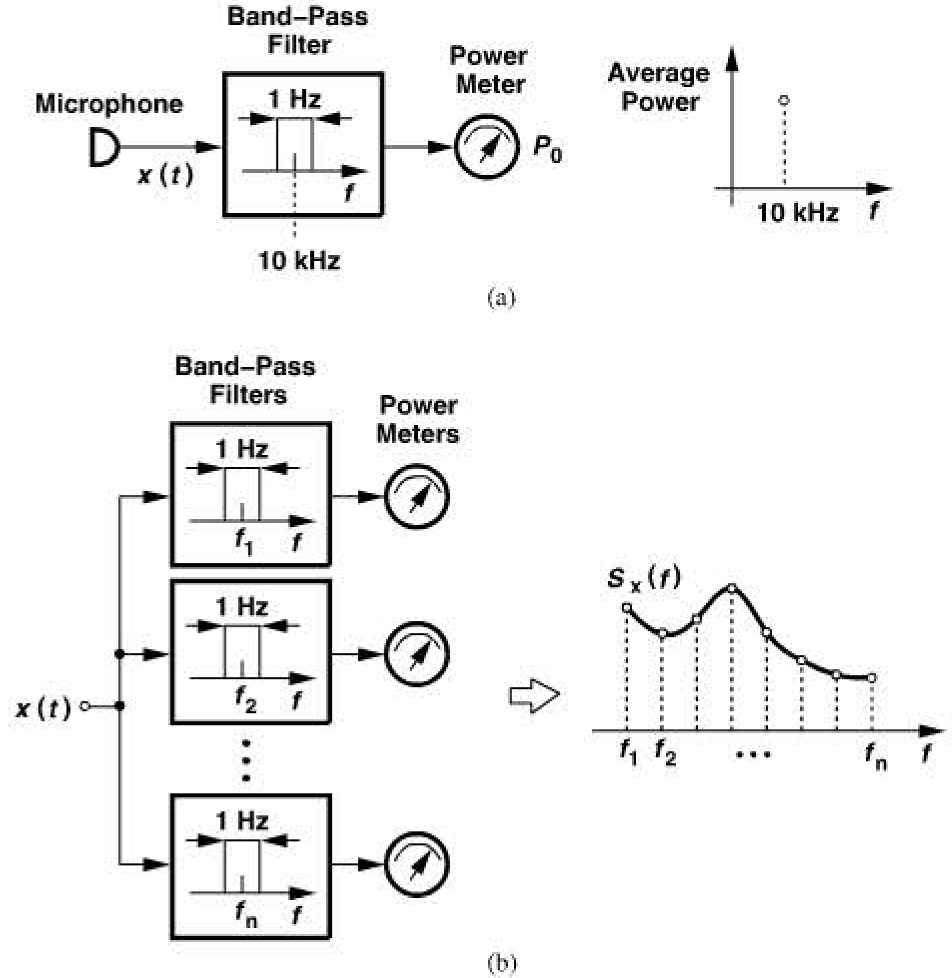



**Figure 2.31** *Measurement of (a) power in 1 Hz, and (b) the spectrum.*

The spectrum shown in Fig. 2.31(b) is called "one-sided" because it is constructed for positive frequencies. In some cases, the analysis is simpler if a "two-sided" spectrum is utilized. The latter is an even-symmetric of the former scaled down vertically by a factor of two (Fig. 2.32), so that the two carry equal energies.

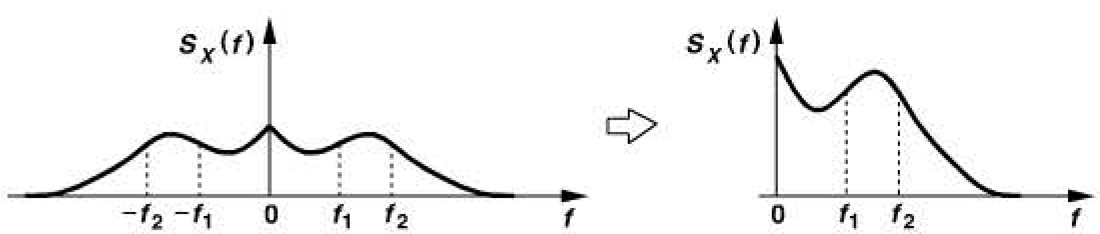



**Figure 2.32** *Two-sided and one-sided spectra.*

**Example 2.15**

A resistor of value $R_1$ generates a noise voltage whose one-sided PSD is given by

$$S_v(f) = 4kTR_1, \tag{2.82}$$



**Example 2.15** *(Continued)*

where $k = 1.38 \times 10^{-23}$ J/K denotes the Boltzmann constant and $T$ the absolute temperature. Such a flat PSD is called "white" because, like white light, it contains all frequencies with equal power levels.

(a) What is the total average power carried by the noise voltage?

(b) What is the dimension of $S_v(f)$?

(c) Calculate the noise voltage for a 50-Ω resistor in 1 Hz at room temperature.

**Solution:**

(a) The area under $S_v(f)$ appears to be infinite, an implausible result because the resistor noise arises from the finite ambient heat. In reality, $S_v(f)$ begins to fall at $f > 1$ THz, exhibiting a finite total energy, i.e., thermal noise is not quite white.

(b) The dimension of $S_v(f)$ is voltage squared per unit bandwidth ($V^2$/Hz) rather than power per unit bandwidth (W/Hz). In fact, we may write the PSD as

$$\overline{V_n^2} = 4kTR, \tag{2.83}$$

where $\overline{V_n^2}$ denotes the average power of $V_n$ in 1 Hz.[14] While some texts express the right-hand side as $4kTR\Delta f$ to indicate the total noise in a bandwidth of $\Delta f$, we omit $\Delta f$ with the understanding that our PSDs always represent power in 1 Hz. We shall use $S_v(f)$ and $\overline{V_n^2}$ interchangeably.

(c) For a 50-Ω resistor at $T = 300$ K,

$$\overline{V_n^2} = 8.28 \times 10^{-19}\,\mathrm{V}^2/\mathrm{Hz}. \tag{2.84}$$

This means that if the noise voltage of the resistor is applied to a 1-Hz band-pass filter centered at any frequency (< 1 THz), then the average measured output is given by the above value. To express the result as a root-mean-squared (rms) quantity and in more familiar units, we may take the square root of both sides:

$$\sqrt{\overline{V_n^2}} = 0.91\,\mathrm{nV}/\sqrt{\mathrm{Hz}}. \tag{2.85}$$

The familiar unit is nV but the strange unit is $\sqrt{\mathrm{Hz}}$. The latter bears no profound meaning; it simply says that the average power in 1 Hz is $(0.91\,\mathrm{nV})^2$.



### 2.3.3 Effect of Transfer Function on Noise

The principal reason for defining the PSD is that it allows many of the frequency-domain operations used with deterministic signals to be applied to random signals as well. For

14. Also called "spot noise."

example, if white noise is applied to a low-pass filter, how do we determine the PSD at the output? As shown in Fig. 2.33, we intuitively expect that the output PSD assumes the shape of the filter's frequency response. In fact, if $x(t)$ is applied to a linear, time-invariant system with a transfer function $H(s)$, then the output spectrum is

$$S_y(f) = S_x(f)|H(f)|^2, \tag{2.86}$$

where $H(f) = H(s = j2\pi f)$ [2]. We note that $|H(f)|$ is squared because $S_x(f)$ is a (voltage or current) squared quantity.

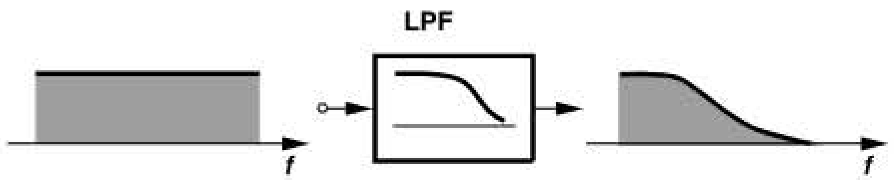



**Figure 2.33** *Effect of low-pass filter on white noise.*

### 2.3.4 Device Noise

In order to analyze the noise performance of circuits, we wish to model the noise of their constituent elements by familiar components such as voltage and current sources. Such a representation allows the use of standard circuit analysis techniques.

**Thermal Noise of Resistors** As mentioned previously, the ambient thermal energy leads to random agitation of charge carriers in resistors and hence noise. The noise can be modeled by a series voltage source with a PSD of $\overline{V_n^2} = 4kTR_1$ [Thevenin equivalent, Fig. 2.34(a)] or a parallel current source with a PSD of $\overline{I_n^2} = \overline{V_n^2}/R_1 = 4kT/R_1$ [Norton equivalent, Fig. 2.34(b)]. The choice of the model sometimes simplifies the analysis. The polarity of the sources is unimportant (but must be kept the same throughout the calculations of a given circuit).

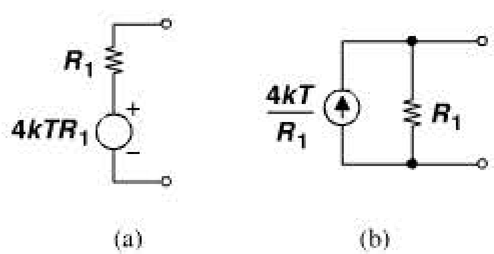



**Figure 2.34** *(a) Thevenin and (b) Norton models of resistor thermal noise.*

**Example 2.16**

Sketch the PSD of the noise voltage measured across the parallel RLC tank depicted in Fig. 2.35(a).



**Example 2.16** *(Continued)*

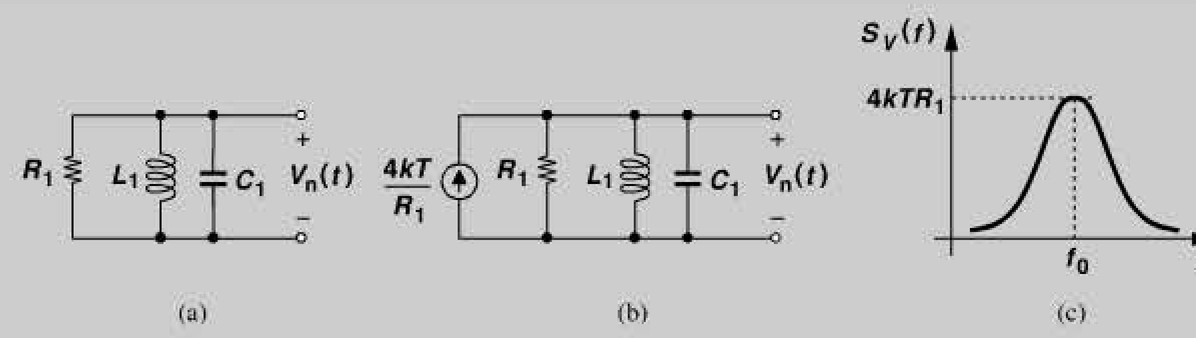



**Figure 2.35** *(a) RLC tank, (b) inclusion of resistor noise, (c) output noise spectrum due to $R_1$.*

**Solution:**

Modeling the noise of $R_1$ by a current source, $\overline{I_{n1}^2} = 4kT/R_1$, [Fig. 2.35(b)] and noting that the transfer function $V_n/I_{n1}$ is, in fact, equal to the impedance of the tank, $Z_T$, we write from Eq. (2.86)

$$\overline{V_n^2} = \overline{I_{n1}^2}|Z_T|^2. \tag{2.87}$$

At $f_0 = (2\pi\sqrt{L_1C_1})^{-1}$, $L_1$ and $C_1$ resonate, reducing the circuit to only $R_1$. Thus, the output noise at $f_0$ is simply equal to $\overline{I_{n1}^2}R_1^2 = 4kTR_1$. At lower or higher frequencies, the impedance of the tank falls and so does the output noise [Fig. 2.35(c)].

If a resistor converts the ambient heat to a noise voltage or current, can we extract energy from the resistor? In particular, does the arrangement shown in Fig. 2.36 deliver energy to $R_2$? Interestingly, if $R_1$ and $R_2$ reside at the same temperature, no net energy is transferred between them because $R_2$ also produces a noise PSD of $4kTR_2$ (Problem 2.8). However, suppose $R_2$ is held at $T = 0$ K. Then, $R_1$ continues to draw thermal energy from its environment, converting it to noise and delivering the energy to $R_2$. The average power transferred to $R_2$ is equal to

$$P_{R2} = \frac{\overline{V_{out}^2}}{R_2} \tag{2.88}$$

$$= \overline{V_n^2}\left(\frac{R_2}{R_1 + R_2}\right)^2 \frac{1}{R_2} \tag{2.89}$$

$$= 4kT\frac{R_1R_2}{(R_1 + R_2)^2}. \tag{2.90}$$

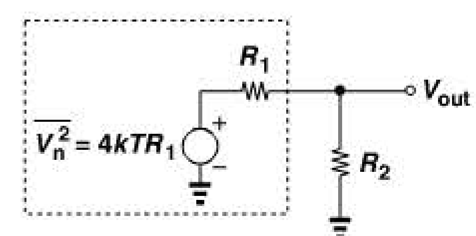



**Figure 2.36** *Transfer of noise from one resistor to another.*

This quantity reaches a maximum if $R_2 = R_1$:

$$P_{R2,max} = kT. \tag{2.91}$$

Called the "available noise power," $kT$ is independent of the resistor value and has the dimension of *power* per unit bandwidth. The reader can prove that $kT = -173.8$ dBm/Hz at $T = 300$ K.

For a circuit to exhibit a thermal noise density of $\overline{V_n^2} = 4kTR_1$, it need not contain an explicit resistor of value $R_1$. After all, Eq. (2.86) suggests that the noise density of a resistor may be transformed to a higher or lower value by the surrounding circuit. We also note that if a passive circuit *dissipates* energy, then it must contain a physical resistance[15] and must therefore *produce* thermal noise. We loosely say "lossy circuits are noisy."

A theorem that consolidates the above observations is as follows: If the real part of the impedance seen between two terminals of a passive (reciprocal) network is equal to $Re\{Z_{out}\}$, then the PSD of the thermal noise seen between these terminals is given by $\overline{V_n^2} = 4kTRe\{Z_{out}\}$ (Fig. 2.37) [8]. This general theorem is not limited to lumped circuits. For example, consider a transmitting antenna that dissipates energy by radiation according to the equation $V_{TX,rms}^2/R_{rad}$, where $R_{rad}$ is the "radiation resistance" [Fig. 2.38(a)]. As a receiving element [Fig. 2.38(b)], the antenna *generates* a thermal noise PSD of[16]

$$\overline{V_{n,ant}^2} = 4kTR_{rad}. \tag{2.92}$$

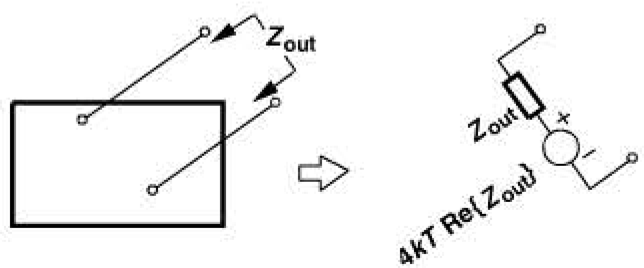



**Figure 2.37** *Output noise of a passive (reciprocal) circuit.*

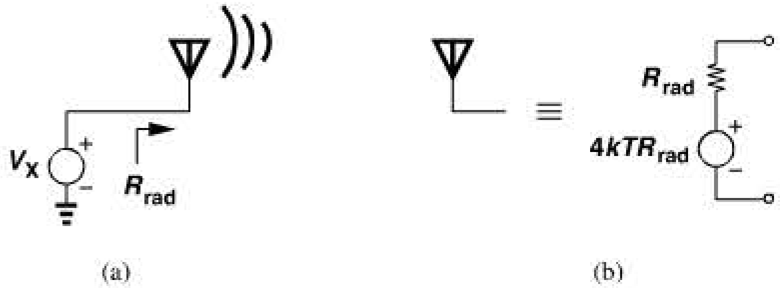



**Figure 2.38** *(a) Transmitting antenna, (b) receiving antenna producing thermal noise.*

15. Recall that ideal inductors and capacitors *store* energy but do not dissipate it.
16. Strictly speaking, this is not correct because the noise of a receiving antenna is in fact given by the "background" noise (e.g., cosmic radiation). However, in RF design, the antenna noise is commonly assumed to be $4kTR_{rad}$.



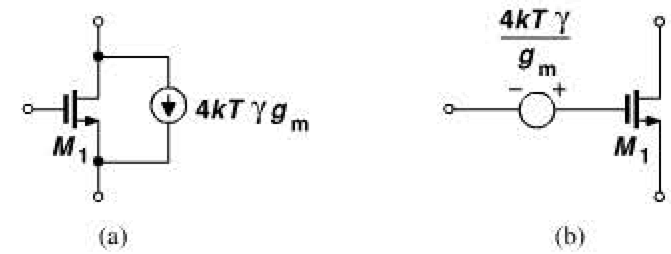



**Figure 2.39** *Thermal channel noise of a MOSFET modeled as a (a) current source, (b) voltage source.*

**Noise in MOSFETs** The thermal noise of MOS transistors operating in the saturation region is approximated by a current source tied between the source and drain terminals [Fig. 2.39(a)]:

$$\overline{I_n^2} = 4kT\gamma g_m, \tag{2.93}$$

where $\gamma$ is the "excess noise coefficient" and $g_m$ the transconductance.[17] The value of $\gamma$ is 2/3 for long-channel transistors and may rise to even 2 in short-channel devices [4]. The actual value of $\gamma$ has other dependencies [5] and is usually obtained by measurements for each generation of CMOS technology. In Problem 2.10, we prove that the noise can alternatively be modeled by a voltage source $\overline{V_n^2} = 4kT\gamma/g_m$ in series with the gate [Fig. 2.39(b)].

Another component of thermal noise arises from the gate resistance of MOSFETs, an effect that becomes increasingly more important as the gate length is scaled down. Illustrated in Fig. 2.40(a) for a device with a width of $W$ and a length of $L$, this resistance amounts to

$$R_G = \frac{W}{L} R_\square, \tag{2.94}$$

where $R_\square$ denotes the sheet resistance (resistance of one square) of the polysilicon gate. For example, if $W = 1\ \mu$m, $L = 45$ nm, and $R_\square = 15\ \Omega$, then $R_G = 333\ \Omega$. Since $R_G$ is *distributed* over the width of the transistor [Fig. 2.40(b)], its noise must be calculated carefully. As proved in [6], the structure can be reduced to a lumped model having an equivalent gate resistance of $R_G/3$ with a thermal noise PSD of $4kTR_G/3$ [Fig. 2.40(c)]. In a good design, this noise must be much less than that of the channel:

$$4kT\frac{R_G}{3} \ll \frac{4kT\gamma}{g_m}. \tag{2.95}$$

The gate and drain terminals also exhibit physical resistances, which are minimized through the use of multiple fingers.

At very high frequencies the thermal noise current flowing through the channel couples to the gate capacitively, thus generating a "gate-induced noise current" [3] (Fig. 2.41). This

17. More accurately, $\overline{I_n^2} = 4kT\gamma g_{d0}$, where $g_{d0}$ is the drain-source conductance in the triode region (even though the noise is measured in saturation) [3].

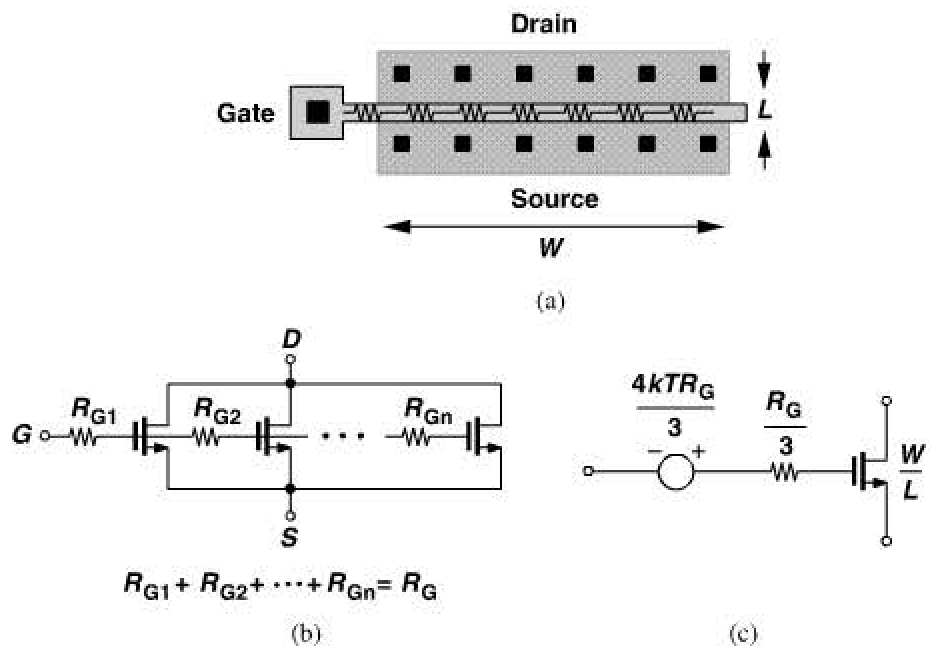



**Figure 2.40** *(a) Gate resistance of a MOSFET, (b) equivalent circuit for noise calculation, (c) equivalent noise and resistance in lumped model.*

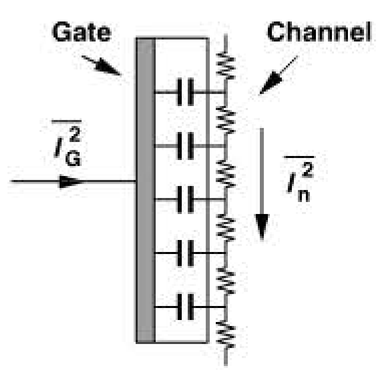



**Figure 2.41** *Gate-induced noise, $\overline{I_G^2}$.*

effect is not modeled in typical circuit simulators, but its significance has remained unclear. In this book, we neglect the gate-induced noise current.

MOS devices also suffer from "flicker" or "$1/f$" noise. Modeled by a voltage source in series with the gate, this noise exhibits the following PSD:

$$\overline{V_n^2} = \frac{K}{WLC_{ox}}\frac{1}{f}, \tag{2.96}$$

where $K$ is a process-dependent constant. In most CMOS technologies, $K$ is lower for PMOS devices than for NMOS transistors because the former carry charge well below the silicon-oxide interface and hence suffer less from "surface states" (dangling bonds) [1]. The $1/f$ dependence means that noise components that vary slowly assume a large amplitude. The choice of the lowest frequency in the noise integration depends on the time scale of interest and/or the spectrum of the desired signal [1].



**Example 2.17**

Can the flicker noise be modeled by a current source?

**Solution:**

Yes, as shown in Fig. 2.42, a MOSFET having a small-signal voltage source of magnitude $V_1$ in series with its gate is equivalent to a device with a current source of value $g_m V_1$ tied between drain and source. Thus,

$$\overline{I_1^2} = g_m^2 \frac{K}{WLC_{ox}}\frac{1}{f}. \tag{2.97}$$

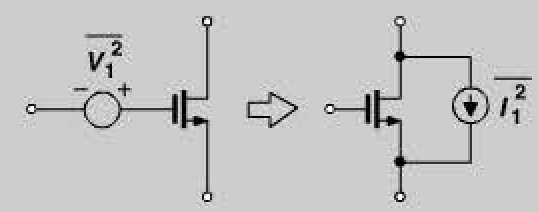



**Figure 2.42** *Conversion of flicker noise voltage to current.*

For a given device size and bias current, the $1/f$ noise PSD intercepts the thermal noise PSD at some frequency, called the "$1/f$ noise corner frequency," $f_c$. Illustrated in Fig. 2.43, $f_c$ can be obtained by converting the flicker noise voltage to current (according to the above example) and equating the result to the thermal noise current:

$$\frac{K}{WLC_{ox}}\frac{1}{f_c}g_m^2 = 4kT\gamma g_m. \tag{2.98}$$

It follows that

$$f_c = \frac{K}{WLC_{ox}}\frac{g_m}{4kT\gamma}. \tag{2.99}$$

The corner frequency falls in the range of tens or even hundreds of megahertz in today's MOS technologies.

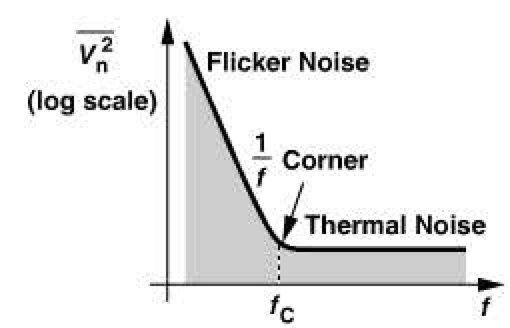



**Figure 2.43** *Flicker noise corner frequency.*

While the effect of flicker noise may seem negligible at high frequencies, we must note that nonlinearity or time variance in circuits such as mixers and oscillators may translate the $1/f$-shaped spectrum to the RF range. We study these phenomena in Chapters 6 and 8.

**Noise in Bipolar Transistors** Bipolar transistors contain physical resistances in their base, emitter, and collector regions, all of which generate thermal noise. Moreover, they also suffer from "shot noise" associated with the transport of carriers across the base-emitter junction. As shown in Fig. 2.44, this noise is modeled by two current sources having the following PSDs:

$$\overline{I_{n,b}^2} = 2qI_B = 2q\frac{I_C}{\beta} \tag{2.100}$$

$$\overline{I_{n,c}^2} = 2qI_C, \tag{2.101}$$

where $I_B$ and $I_C$ are the base and collector bias currents, respectively. Since $g_m = I_C/(kT/q)$ for bipolar transistors, the collector current shot noise is often expressed as

$$\overline{I_{n,c}^2} = 4kT\frac{g_m}{2}, \tag{2.102}$$

in analogy with the thermal noise of MOSFETs or resistors.

In low-noise circuits, the base resistance thermal noise and the collector current shot noise become dominant. For this reason, wide transistors biased at high current levels are employed.

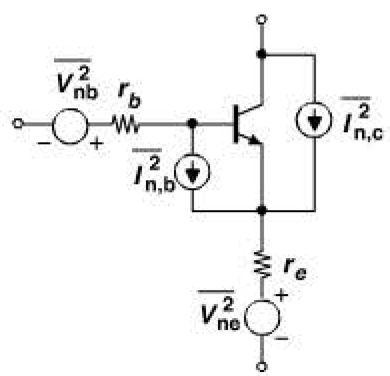



**Figure 2.44** *Noise sources in a bipolar transistor.*

## 2.3.5 Representation of Noise in Circuits

With the noise of devices formulated above, we now wish to develop *measures* of the noise performance of circuits, i.e., metrics that reveal how noisy a given circuit is.

**Input-Referred Noise** How can the noise of a circuit be observed in the laboratory? We have access only to the output and hence can measure only the output noise. Unfortunately, the output noise does not permit a fair comparison between circuits: a circuit may exhibit high output noise because it has a high gain rather than high noise. For this reason, we "refer" the noise to the input.



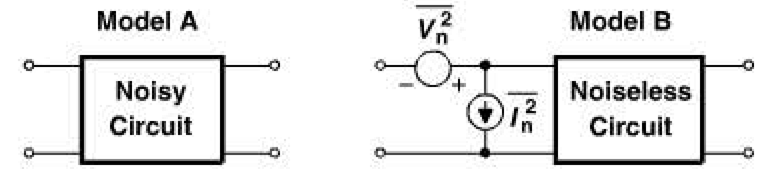



**Figure 2.45** *Input-referred noise.*

In analog design, the input-referred noise is modeled by a series voltage source and a parallel current source (Fig. 2.45) [1]. The former is obtained by shorting the input port of models A and B and equating their output noises (or, equivalently, dividing the output noise by the voltage gain). Similarly, the latter is computed by leaving the input ports open and equating the output noises (or, equivalently, dividing the output noise by the transimpedance gain).

**Example 2.18**

Calculate the input-referred noise of the common-gate stage depicted in Fig. 2.46(a). Assume $I_1$ is ideal and neglect the noise of $R_1$.

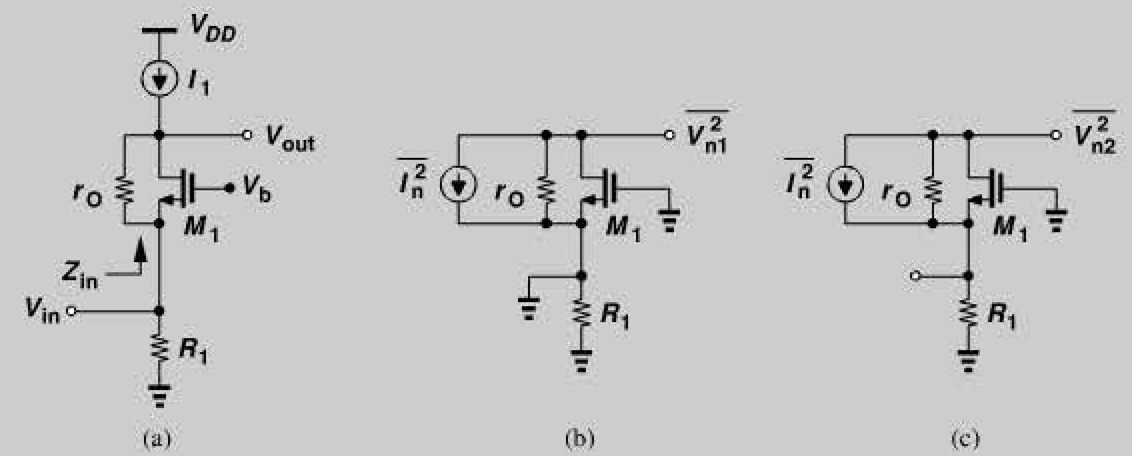



**Figure 2.46** *(a) CG stage, (b) computation of input-referred noise voltage, (c) computation of input-referred noise current.*

**Solution:**

Shorting the input to ground, we write from Fig. 2.46(b),

$$\overline{V_{n1}^2} = \overline{I_n^2} \cdot r_O^2. \tag{2.103}$$

Since the voltage gain of the stage is given by $1 + g_m r_O$, the input-referred noise voltage is equal to

$$\overline{V_{n,in}^2} = \frac{\overline{I_n^2} r_O^2}{(1 + g_m r_O)^2} \tag{2.104}$$

$$\approx \frac{4kT\gamma}{g_m}, \tag{2.105}$$

*(Continues)*

**Example 2.18** *(Continued)*

where it is assumed $g_m r_O \gg 1$. Leaving the input open as shown in Fig. 2.46(c), the reader can show that (Problem 2.12)

$$\overline{V_{n2}^2} = \overline{I_n^2} r_O^2. \tag{2.106}$$

Defined as the output voltage divided by the input current, the transimpedance gain of the stage is given by $g_m r_O R_1$ (why?). It follows that

$$\overline{I_{n,in}^2} = \frac{\overline{I_n^2} r_O^2}{g_m^2 r_O^2 R_1^2} \tag{2.107}$$

$$= \frac{4kT\gamma}{g_m R_1^2}. \tag{2.108}$$

From the above example, it may appear that the noise of $M_1$ is "counted" twice. It can be shown that [1] the two input-referred noise sources are necessary and sufficient, but often correlated.

**Example 2.19**

Explain why the output noise of a circuit depends on the output impedance of the *preceding* stage.

**Solution:**

Modeling the noise of the circuit by input-referred sources as shown in Fig. 2.47, we observe that some of $\overline{I_n^2}$ flows through $Z_1$, generating a noise voltage at the input that depends on $|Z_1|$. Thus, the output noise, $V_{n,out}$, also depends on $|Z_1|$.

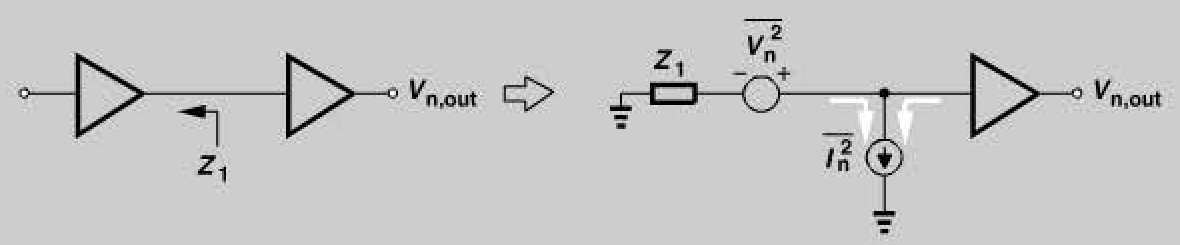



**Figure 2.47** *Noise in a cascade.*

The computation and use of input-referred noise sources prove difficult at high frequencies. For example, it is quite challenging to measure the transimpedance gain of an RF stage. For this reason, RF designers employ the concept of "noise figure" as another metric of noise performance that more easily lends itself to measurement.

**Noise Figure** In circuit and system design, we are interested in the signal-to-noise ratio (SNR), defined as the signal power divided by the noise power. It is therefore helpful to



ask, how does the SNR degrade as the signal travels through a given circuit? If the circuit contains no noise, then the output SNR is *equal* to the input SNR even if the circuit acts as an attenuator.[18] To quantify how noisy the circuit is, we define its noise figure (NF) as

$$\mathrm{NF} = \frac{SNR_{in}}{SNR_{out}} \tag{2.109}$$

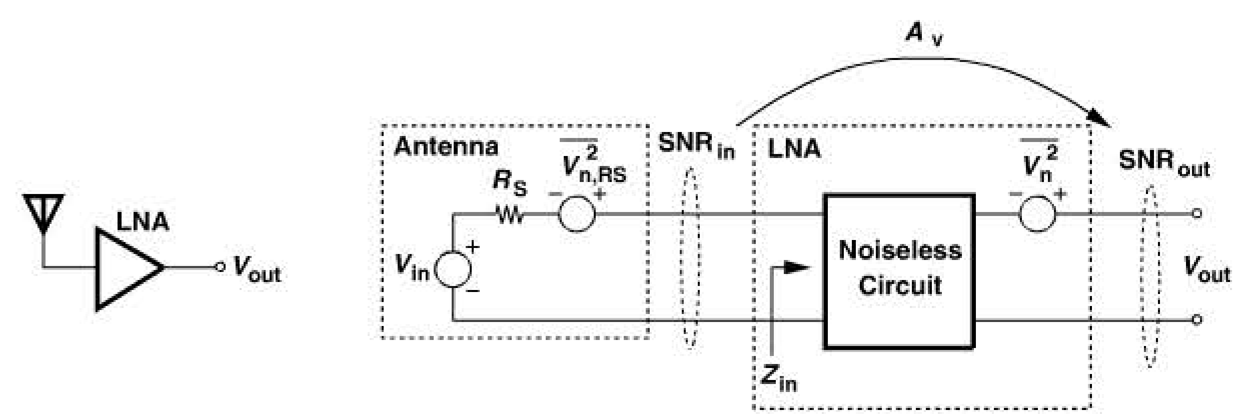
such that it is equal to 1 for a noiseless stage. Since each quantity in this ratio has a dimension of power (or voltage squared), we express NF in decibels as

$$\mathrm{NF}|_{\mathrm{dB}} = 10 \log \frac{SNR_{in}}{SNR_{out}}. \tag{2.110}$$

Note that most texts call (2.109) the "noise factor" and (2.110) the noise figure. We do not make this distinction in this book.

Compared to input-referred noise, the definition of NF in (2.109) may appear rather complicated: it depends on not only the noise of the circuit under consideration but the SNR provided by the *preceding stage*. In fact, if the input signal contains no noise, then $SNR_{in} = \infty$ and $\mathrm{NF} = \infty$, even though the circuit may have finite internal noise. For such a case, NF is not a meaningful parameter and only the input-referred noise can be specified.

Calculation of the noise figure is generally simpler than Eq. (2.109) may suggest. For example, suppose a low-noise amplifier senses the signal received by an antenna [Fig. 2.48(a)]. As predicted by Eq. (2.92), the antenna "radiation resistance," $R_S$, produces thermal noise, leading to the model shown in Fig. 2.48(b). Here, $\overline{V_{n,RS}^2}$ represents the thermal noise of the antenna, and $\overline{V_n^2}$ the output noise of the LNA. We must compute $SNR_{in}$ at the LNA input and $SNR_{out}$ at its output.




(a) (b)

**Figure 2.48** *(a) Antenna followed by LNA, (b) equivalent circuit.*

18. Because the input signal and the input noise are attenuated by the same factor.

If the LNA exhibits an input impedance of $Z_{in}$, then both $V_{in}$ and $V_{RS}$ experience an attenuation factor of $\alpha = Z_{in}/(Z_{in} + R_S)$ as they appear at the input of the LNA. That is,

$$SNR_{in} = \frac{|\alpha|^2 V_{in}^2}{|\alpha|^2 \overline{V_{RS}^2}}, \tag{2.111}$$

where $V_{in}$ denotes the rms value of the signal received by the antenna.

To determine $SNR_{out}$, we assume a voltage gain of $A_v$ from the LNA input to the output and recognize that the output signal power is equal to $V_{in}^2|\alpha|^2A_v^2$. The output noise consists of two components: (a) the noise of the antenna amplified by the LNA, $\overline{V_{RS}^2}|\alpha|^2A_v^2$, and (b) the output noise of the LNA, $\overline{V_n^2}$. Since these two components are uncorrelated, we simply add the PSDs and write

$$SNR_{out} = \frac{V_{in}^2|\alpha|^2A_v^2}{\overline{V_{RS}^2}|\alpha|^2A_v^2 + \overline{V_n^2}}. \tag{2.112}$$

It follows that

$$\mathrm{NF} = \frac{V_{in}^2}{4kTR_S} \cdot \frac{\overline{V_{RS}^2}|\alpha|^2A_v^2 + \overline{V_n^2}}{V_{in}^2|\alpha|^2A_v^2} \tag{2.113}$$

$$= \frac{1}{\overline{V_{RS}^2}} \cdot \frac{\overline{V_{RS}^2}|\alpha|^2A_v^2 + \overline{V_n^2}}{|\alpha|^2A_v^2} \tag{2.114}$$

$$= 1 + \frac{\overline{V_n^2}}{|\alpha|^2A_v^2} \cdot \frac{1}{\overline{V_{RS}^2}}. \tag{2.115}$$

This result leads to another definition of the NF: the total noise at the output divided by the noise at the output due to the source impedance. The NF is usually specified for a 1-Hz bandwidth at a given frequency, and hence sometimes called the "spot noise figure" to emphasize the small bandwidth.

Equation (2.115) suggests that the NF depends on the *source impedance*, not only through $\overline{V_{RS}^2}$ but also through $\overline{V_n^2}$ (Example 2.19). In fact, if we model the noise by *input-referred* sources, then the input noise current, $\overline{I_{n,in}^2}$, partially flows through $R_S$, generating a source-dependent noise voltage of $\overline{I_{n,in}^2}R_S^2$ at the input and hence a proportional noise at the output. Thus, the NF must be specified with respect to a source impedance—typically 50 Ω.

For hand analysis and simulations, it is possible to reduce the right-hand side of Eq. (2.114) to a simpler form by noting that the numerator is the *total* noise measured at the output:

$$\mathrm{NF} = \frac{1}{4kTR_S} \cdot \frac{\overline{V_{n,out}^2}}{A_0^2}, \tag{2.116}$$

where $\overline{V_{n,out}^2}$ includes both the source impedance noise and the LNA noise, and $A_0 = |\alpha|A_v$ is the voltage gain from $V_{in}$ to $V_{out}$ (rather than the gain from the LNA input to its output). We loosely say, "to calculate the NF, we simply divide the total output noise by the gain from $V_{in}$ to $V_{out}$ and normalize the result to the noise of $R_S$." Alternatively, we can say from (2.115) that "we calculate the output noise due to the amplifier ($\overline{V_n^2}$), divide it by the gain, normalize it to $4kTR_S$, and add 1 to the result."



It is important to note that the above derivations are valid even if no actual *power* is transferred from the antenna to the LNA or from the LNA to a load. For example, if $Z_{in}$ in Fig. 2.48(b) goes to infinity, no power is delivered to the LNA, but all of the derivations remain valid because they are based on *voltage* (squared) quantities rather than power quantities. In other words, so long as the derivations incorporate noise and signal voltages, no inconsistency arises in the presence of impedance mismatches or even infinite input impedances. This is a critical difference in thinking between modern RF design and traditional microwave design.

**Example 2.20**

Compute the noise figure of a shunt resistor $R_P$ with respect to a source impedance $R_S$ [Fig. 2.49(a)].

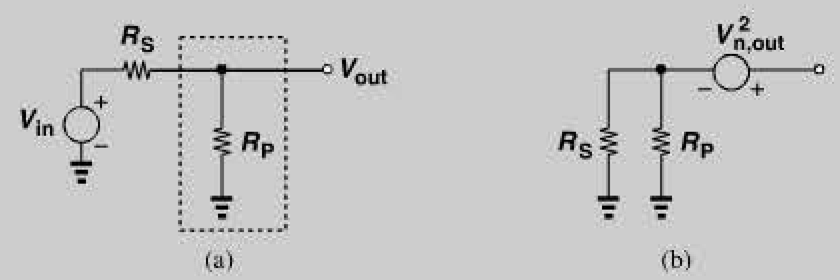



**Figure 2.49** *(a) Circuit consisting of a single parallel resistor, (b) model for NF calculation.*

**Solution:**

From Fig. 2.49(b), the total output noise voltage is obtained by setting $V_{in}$ to zero:

$$\overline{V_{n,out}^2} = 4kT(R_S||R_P). \tag{2.117}$$

The gain is equal to

$$A_0 = \frac{R_P}{R_P + R_S}. \tag{2.118}$$

Thus,

$$\mathrm{NF} = 4kT(R_S||R_P)\frac{(R_S + R_P)^2}{R_P^2}\frac{1}{4kTR_S} \tag{2.119}$$

$$= 1 + \frac{R_S}{R_P}. \tag{2.120}$$

The NF is therefore minimized by *maximizing* $R_P$. Note that if $R_P = R_S$ to provide impedance matching, then the NF cannot be less than 3 dB. We will return to this critical point in the context of LNA design in Chapter 5.

### Example 2.21

Determine the noise figure of the common-source stage shown in Fig. 2.50(a) with respect to a source impedance $R_S$. Neglect the capacitances and flicker noise of $M_1$ and assume $I_1$ is ideal.

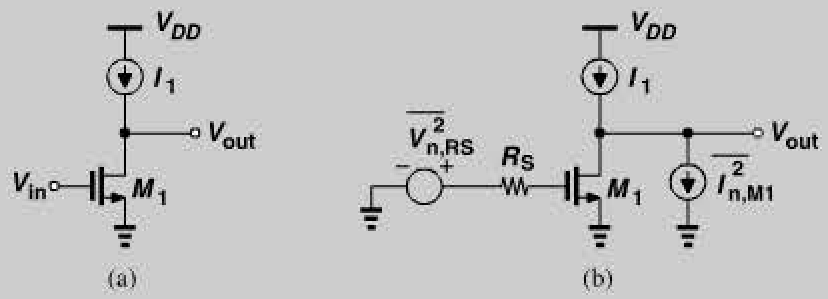



**Figure 2.50** *(a) CS stage, (b) inclusion of noise.*

**Solution:**

From Fig. 2.50(b), the output noise consists of two components: (a) that due to $M_1$, $\overline{I^2_{n,M1}}r_O^2$, and (b) the amplified noise of $R_S$, $\overline{V^2_{RS}}(g_m r_O)^2$. It follows that

$$\mathrm{NF} = \frac{4kT\gamma g_m r_O^2 + 4kTR_S(g_m r_O)^2}{(g_m r_O)^2}\cdot\frac{1}{4kTR_S} \tag{2.121}$$

$$= \frac{\gamma}{g_m R_S} + 1. \tag{2.122}$$

This result implies that the NF falls as $R_S$ rises. Does this mean that, even though the amplifier remains unchanged, the overall system noise performance *improves* as $R_S$ increases?! This interesting point is studied in Problems 2.18 and 2.19.

**Noise Figure of Cascaded Stages** Since many stages appear in a receiver chain, it is desirable to determine the NF of the overall cascade in terms of that of each stage. Consider the cascade depicted in Fig. 2.51(a), where $A_{v1}$ and $A_{v2}$ denote the *unloaded* voltage gain of the two stages. The input and output impedances and the output noise voltages of the two stages are also shown.[19]

We first obtain the NF of the cascade using a direct method; according to (2.115), we simply calculate the total noise at the output due to the two stages, divide by $(V_{out}/V_{in})^2$, normalize to $4kTR_S$, and add one to the result. Taking the loadings into account, we write the overall voltage gain as

$$A_0 = \frac{V_{out}}{V_{in}} = \frac{R_{in1}}{R_{in1}+R_S}A_{v1}\frac{R_{in2}}{R_{in2}+R_{out1}}A_{v2}. \tag{2.123}$$

19. We assume for simplicity that the reactive components of the input and output impedances are nulled but the final result is valid even if they are not.



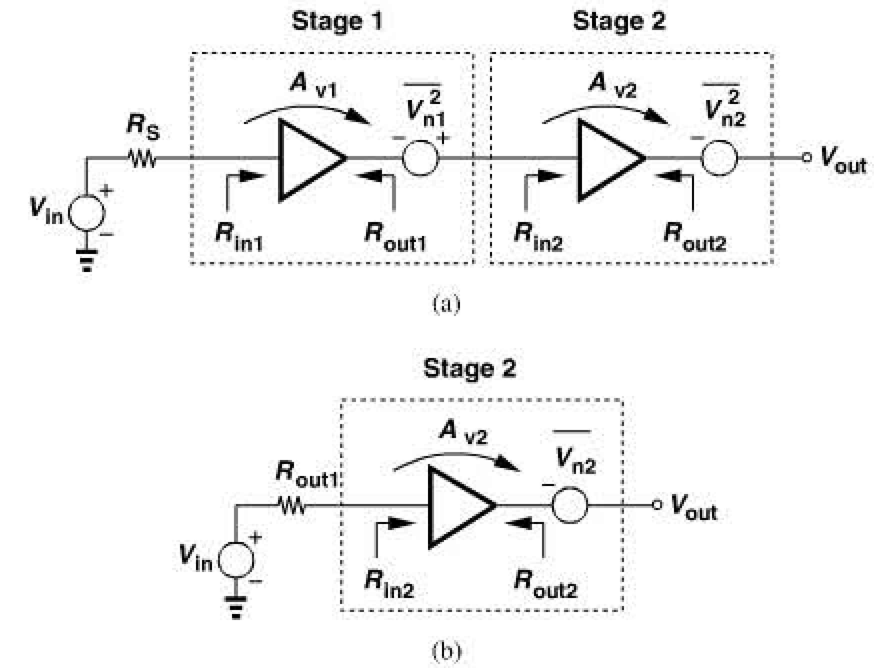



**Figure 2.51** *(a) Noise in a cascade of stages, (b) simplified diagram.*

The output noise due to the two stages, denoted by $\overline{V^2_{n,out}}$, consists of two components: (a) $\overline{V^2_{n2}}$, and (b) $\overline{V^2_{n1}}$ amplified by the second stage. Since $V_{n1}$ sees an impedance of $R_{out1}$ to its left and $R_{in2}$ to its right, it is scaled by a factor of $R_{in2}/(R_{in2}+R_{out1})$ as it appears at the input of the second stage. Thus,

$$\overline{V^2_{n,out}} = \overline{V^2_{n2}} + \overline{V^2_{n1}}\frac{R^2_{in2}}{(R_{in2}+R_{out1})^2}A^2_{v2}. \tag{2.124}$$

The overall NF is therefore expressed as

$$\mathrm{NF_{tot}} = 1 + \frac{\overline{V^2_{n,out}}}{A_0^2}\cdot\frac{1}{4kTR_S} \tag{2.125}$$

$$= 1 + \frac{\overline{V^2_{n1}}}{\left(\dfrac{R_{in1}}{R_{in1}+R_S}\right)^2 A^2_{v1}}\cdot\frac{1}{4kTR_S}$$

$$+ \frac{\overline{V^2_{n2}}}{\left(\dfrac{R_{in1}}{R_{in1}+R_S}\right)^2 A^2_{v1}\left(\dfrac{R_{in2}}{R_{in2}+R_{out1}}\right)^2 A^2_{v2}}\cdot\frac{1}{4kTR_S} \tag{2.126}$$

The first two terms constitute the *NF* of the first stage, $NF_1$, with respect to a source impedance of $R_S$. The third term represents the noise of the second stage, but how can it be expressed in terms of the *noise figure* of this stage?

Let us now consider the second stage by itself and determine its noise figure with respect to a source impedance of $R_{out1}$ [Fig. 2.51(b)]. Using (2.115) again, we have

$$NF_2 = 1 + \frac{\overline{V_{n2}^2}}{\dfrac{R_{in2}^2}{(R_{in2}+R_{out1})^2}A_{v2}^2}\frac{1}{4kTR_{out1}}. \tag{2.127}$$

It follows from (2.126) and (2.127) that

$$NF_{tot} = NF_1 + \frac{NF_2 - 1}{\dfrac{R_{in1}^2}{(R_{in1}+R_S)^2}A_{v1}^2\dfrac{R_S}{R_{out1}}}. \tag{2.128}$$

What does the denominator represent? This quantity is in fact the "available power gain" of the first stage, defined as the "available power" at its output, $P_{out,av}$ (the power that it would deliver to a matched load) divided by the available source power, $P_{S,av}$ (the power that the source would deliver to a matched load). This can be readily verified by finding the power that the first stage in Fig. 2.51(a) would deliver to a load equal to $R_{out1}$:

$$P_{out,av} = V_{in}^2\frac{R_{in1}^2}{(R_S+R_{in1})^2}A_{v1}^2 \cdot \frac{1}{4R_{out1}}. \tag{2.129}$$

Similarly, the power that $V_{in}$ would deliver to a load of $R_S$ is given by

$$P_{S,av} = \frac{V_{in}^2}{4R_S}. \tag{2.130}$$

The ratio of (2.129) and (2.130) is indeed equal to the denominator in (2.128).

With these observations, we write

$$NF_{tot} = NF_1 + \frac{NF_2 - 1}{A_{P1}}, \tag{2.131}$$



where $A_{P1}$ denotes the "available power gain" of the first stage. It is important to bear in mind that $NF_2$ is computed with respect to the output impedance of the first stage. For $m$ stages,

$$NF_{tot} = 1 + (NF_1 - 1) + \frac{NF_2 - 1}{A_{P1}} + \cdots + \frac{NF_m - 1}{A_{P1}\cdots A_{P(m-1)}}. \tag{2.132}$$

Called "Friis' equation" [7], this result suggests that the noise contributed by each stage decreases as the total gain preceding that stage increases, implying that the first few stages in a cascade are the most critical. Conversely, if a stage suffers from attenuation (loss), then the NF of the following circuits is "amplified" when referred to the input of that stage.



**Example 2.22**

Determine the $NF$ of the cascade of common-source stages shown in Fig. 2.52. Neglect the transistor capacitances and flicker noise.

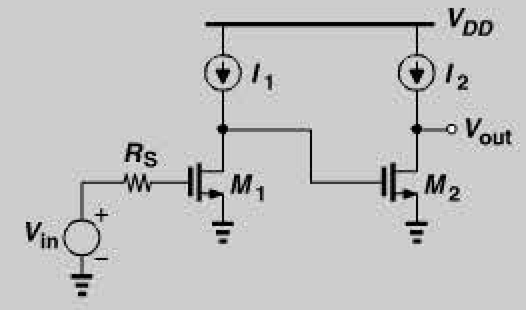



**Figure 2.52** *Cascade of CS stages for noise figure calculation.*

**Solution:**

Which approach is simpler to use here, the direct method or Friis' equation? Since $R_{in1} = R_{in2} = \infty$, Eq. (2.126) reduces to

$$NF = 1 + \frac{\overline{V_{n1}^2}}{A_{v1}^2}\frac{1}{4kTR_S} + \frac{\overline{V_{n2}^2}}{A_{v1}^2A_{v2}^2}\frac{1}{4kTR_S}, \tag{2.133}$$

where $\overline{V_{n1}^2} = 4kT\gamma g_{m1}r_{O1}^2$, $\overline{V_{n2}^2} = 4kT\gamma g_{m2}r_{O2}^2$, $A_{v1} = g_{m1}r_{O1}$, and $A_{v2} = g_{m2}r_{O2}$. With all of these quantities readily available, we simply substitute for their values in (2.133), obtaining

$$NF = 1 + \frac{\gamma}{g_{m1}R_S} + \frac{\gamma}{g_{m1}^2r_{O1}^2g_{m2}R_S}. \tag{2.134}$$

On the other hand, Friis' equation requires the calculation of the available power gain of the first stage and the $NF$ of the second stage with respect to a source impedance of $r_{O1}$, leading to lengthy algebra.

The foregoing example represents a typical situation in modern RF design: the interface between the two stages does not have a 50-Ω impedance *and* no attempt has been made to provide impedance matching between the two stages. In such cases, Friis' equation becomes cumbersome, making direct calculation of the $NF$ more attractive.

While the above example assumes an infinite input impedance for the second stage, the direct method can be extended to more realistic cases with the aid of Eq. (2.126). Even in the presence of complex input and output impedances, Eq. (2.126) indicates that (1) $\overline{V_{n1}^2}$ must be divided by the *unloaded* gain from $V_{in}$ to the output of the first stage; (2) the output noise of the second stage, $\overline{V_{n2}^2}$, must be calculated with this stage driven by the output impedance of the first stage;[20] and (3) $\overline{V_{n2}^2}$ must be divided by the total voltage gain from $V_{in}$ to $V_{out}$.

20. Recall from Example 2.19 that the output noise of a circuit may depend on the source impedance driving it, but the source impedance noise is excluded from $\overline{V_{n2}^2}$.

### Example 2.23

Determine the noise figure of the circuit shown in Fig. 2.53(a). Neglect transistor capacitances, flicker noise, channel-length modulation, and body effect.

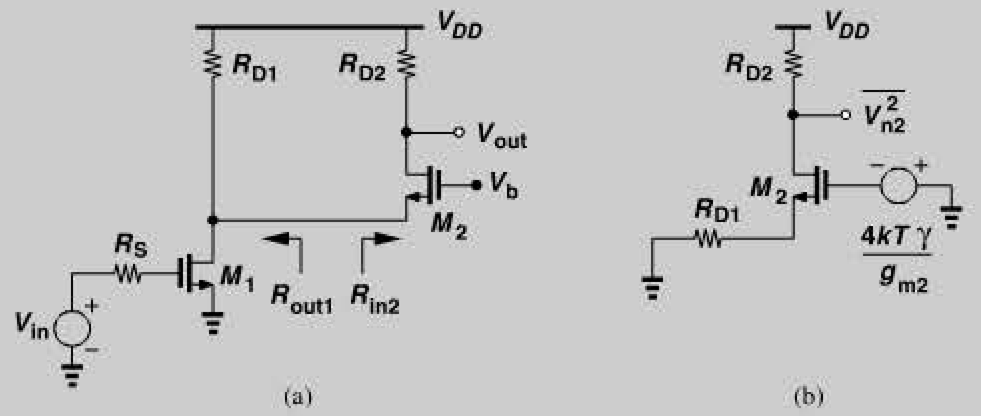



**Figure 2.53** *(a) Cascade of CS and CG stages, (b) simplified diagram.*

**Solution:**

For the first stage, $A_{v1} = -g_{m1}R_{D1}$ and the unloaded output noise is equal to

$$\overline{V_{n1}^2} = 4kT\gamma g_{m1}R_{D1}^2 + 4kTR_{D1}. \tag{2.135}$$

For the second stage, the reader can show from Fig. 2.53(b) that

$$\overline{V_{n2}^2} = \frac{4kT\gamma}{g_{m2}}\left(\frac{R_{D2}}{\frac{1}{g_{m2}}+R_{D1}}\right)^2 + 4kTR_{D2}. \tag{2.136}$$

Note that the output impedance of the first stage is included in the calculation of $\overline{V_{n2}^2}$ but the *noise* of $R_{D1}$ is not.

We now substitute these values in Eq. (2.126), bearing in mind that $R_{in2} = 1/g_{m2}$ and $A_{v2} = g_{m2}R_{D2}$.

$$NF_{tot} = 1 + \frac{4kT\gamma g_{m1}R_D^2 + 4kTR_{D1}}{g_{m1}^2R_{D1}^2}\cdot\frac{1}{4kTR_S} + \frac{\dfrac{4kT\gamma}{g_{m2}}\left(\dfrac{R_{D2}}{g_{m2}^{-1}+R_{D2}}\right)^2 + 4kTR_{D2}}{g_{m1}^2R_{D1}^2\left(\dfrac{g_{m2}^{-1}}{g_{m2}^{-1}+R_{D1}}\right)^2 g_{m2}^2R_{D2}^2}\cdot\frac{1}{4kTR_S}. \tag{2.137}$$

**Noise Figure of Lossy Circuits** Passive circuits such as filters appear at the front end of RF transceivers and their loss proves critical (Chapter 4). The loss arises from unwanted resistive components within the circuit that convert the input power to heat, thereby producing a smaller signal power at the output. Furthermore, recall from Fig. 2.37 that resistive components also *generate* thermal noise. That is, passive lossy circuits both attenuate the signal and introduce noise.



We wish to prove that the noise figure of a passive (reciprocal) circuit is equal to its "power loss," defined as $L = P_{in}/P_{out}$, where $P_{in}$ is the available source power and $P_{out}$ the available power at the output. As mentioned in the derivation of Friis' equation, the available power is the power that a given source or circuit would deliver to a conjugate-matched load. The proof is straightforward if the input and output are matched (Problem 2.20). We consider a more general case here.

Consider the arrangement shown in Fig. 2.54(a), where the lossy circuit is driven by a source impedance of $R_S$ while driving a load impedance of $R_L$.[21] From Eq. (2.130), the available source power is $P_{in} = V_{in}^2/(4R_S)$. To determine the available output power, we construct the Thevenin equivalent shown in Fig. 2.54(b), obtaining $P_{out} = V_{Thev}^2/(4R_{out})$. Thus, the loss is given by

$$L = \frac{V_{in}^2}{V_{Thev}^2}\frac{R_{out}}{R_S}. \tag{2.138}$$

To calculate the noise figure, we utilize the theorem illustrated in Fig. 2.37 and the equivalent circuit shown in Fig. 2.54(c) to write

$$\overline{V_{n,out}^2} = 4kTR_{out}\frac{R_L^2}{(R_L+R_{out})^2}. \tag{2.139}$$

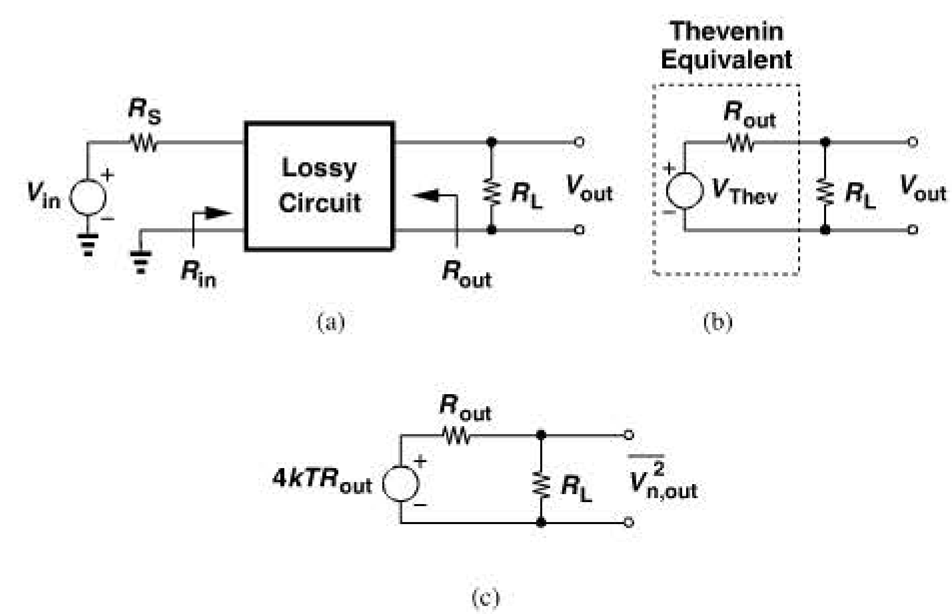



**Figure 2.54** *(a) Lossy passive network, (b) Thevenin equivalent, (c) simplified diagram.*

21. For simplicity, we assume the reactive parts of the impedances are cancelled but the final result is valid even if they are not.

Note that $R_L$ is assumed noiseless so that only the noise figure of the lossy circuit can be determined. The voltage gain from $V_{in}$ to $V_{out}$ is found by noting that, in response to $V_{in}$, the circuit produces an output voltage of $V_{out} = V_{Thev}R_L/(R_L + R_{out})$ [Fig. 2.54(b)]. That is,

$$A_0 = \frac{V_{Thev}}{V_{in}} \frac{R_L}{R_L + R_{out}}. \tag{2.140}$$

The $NF$ is equal to (2.139) divided by the square of (2.140) and normalized to $4kTR_S$:

$$NF = 4kTR_{out} \frac{V_{in}^2}{V_{Thev}^2} \frac{1}{4kTR_S} \tag{2.141}$$
$$= L. \tag{2.142}$$

**Example 2.24**

The receiver shown in Fig. 2.55 incorporates a front-end band-pass filter (BPF) to suppress some of the interferers that may desensitize the LNA. If the filter has a loss of $L$ and the LNA a noise figure of $NF_{LNA}$, calculate the overall noise figure.

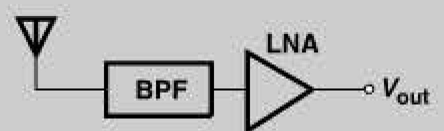



**Figure 2.55** *Cascade of BPF and LNA.*

**Solution:**

Denoting the noise figure of the filter by $NF_{filt}$, we write Friis' equation as

$$NF_{tot} = NF_{filt} + \frac{NF_{LNA} - 1}{L^{-1}} \tag{2.143}$$
$$= L + (NF_{LNA} - 1)L \tag{2.144}$$
$$= L \cdot NF_{LNA}, \tag{2.145}$$

where $NF_{LNA}$ is calculated with respect to the output resistance of the filter. For example, if $L = 1.5$ dB and $NF_{LNA} = 2$ dB, then $NF_{tot} = 3.5$ dB.

## 2.4 SENSITIVITY AND DYNAMIC RANGE

The performance of RF receivers is characterized by many parameters. We study two, namely, sensitivity and dynamic range, here and defer the others to Chapter 3.



### 2.4.1 Sensitivity

The sensitivity is defined as the minimum signal level that a receiver can detect with "acceptable quality." In the presence of excessive noise, the detected signal becomes unintelligible and carries little information. We define acceptable quality as sufficient signal-to-noise ratio, which itself depends on the type of modulation and the corruption (e.g., bit error rate) that the system can tolerate. Typical required SNR levels are in the range of 6 to 25 dB (Chapter 3).

In order to calculate the sensitivity, we write

$$NF = \frac{SNR_{in}}{SNR_{out}} \tag{2.146}$$
$$= \frac{P_{sig}/P_{RS}}{SNR_{out}}, \tag{2.147}$$

where $P_{sig}$ denotes the input signal power and $P_{RS}$ the source resistance noise power, both per unit bandwidth. Do we express these quantities in $V^2$/Hz or W/Hz? Since the input impedance of the receiver is typically matched to that of the antenna (Chapter 4), the antenna indeed delivers signal power and noise power to the receiver. For this reason, it is common to express both quantities in W/Hz (or dBm/Hz). It follows that

$$P_{sig} = P_{RS} \cdot NF \cdot SNR_{out}. \tag{2.148}$$

Since the overall signal power is distributed across a certain bandwidth, $B$, the two sides of (2.148) must be integrated over the bandwidth so as to obtain the total mean squared power. Assuming a flat spectrum for the signal and the noise, we have

$$P_{sig,tot} = P_{RS} \cdot NF \cdot SNR_{out} \cdot B. \tag{2.149}$$

Equation (2.149) expresses the sensitivity as the minimum input signal that yields a given value for the output $SNR$. Changing the notation slightly and expressing the quantities in dB or dBm, we have[22]

$$P_{sen}|_{dBm} = P_{RS}|_{dBm/Hz} + NF|_{dB} + SNR_{min}|_{dB} + 10\log B, \tag{2.150}$$

where $P_{sen}$ is the sensitivity and $B$ is expressed in Hz. Note that (2.150) does not directly depend on the gain of the system. If the receiver is matched to the antenna, then from (2.91), $P_{RS} = kT = -174$ dBm/Hz and

$$P_{sen} = -174\,\text{dBm/Hz} + NF + 10\log B + SNR_{min}. \tag{2.151}$$

Note that the sum of the first three terms is the total integrated noise of the system (sometimes called the "noise floor").

22. Note that in conversion to dB or dBm, we take 10 log because these are power quantities.

**Example 2.25**

A GSM receiver requires a minimum *SNR* of 12 dB and has a channel bandwidth of 200 kHz. A wireless LAN receiver, on the other hand, specifies a minimum *SNR* of 23 dB and has a channel bandwidth of 20 MHz. Compare the sensitivities of these two systems if both have an *NF* of 7 dB.

**Solution:**

For the GSM receiver, $P_{sen} = -102$ dBm, whereas for the wireless LAN system, $P_{sen} = -71$ dBm. Does this mean that the latter is inferior? No, the latter employs a much wider bandwidth and a more efficient modulation to accommodate a data rate of 54 Mb/s. The GSM system handles a data rate of only 270 kb/s. In other words, specifying the sensitivity of a receiver without the data rate is not meaningful.

### 2.4.2 Dynamic Range

Dynamic range (DR) is loosely defined as the maximum input level that a receiver can "tolerate" divided by the minimum input level that it can detect (the sensitivity). This definition is quantified differently in different applications. For example, in analog circuits such as analog-to-digital converters, the DR is defined as the "full-scale" input level divided by the input level at which $SNR = 1$. The full scale is typically the input level beyond which a hard saturation occurs and can be easily determined by examining the circuit.

In RF design, on the other hand, the situation is more complicated. Consider a simple common-source stage. How do we define the input "full scale" for such a circuit? Is there a particular input level beyond which the circuit becomes excessively nonlinear? We may view the 1-dB compression point as such a level. But, what if the circuit senses two interferers and suffers from intermodulation?

In RF design, two definitions of DR have emerged. The first, simply called the dynamic range, refers to the maximum tolerable *desired* signal power divided by the minimum tolerable desired signal power (the sensitivity). Illustrated in Fig. 2.56(a), this DR is limited by compression at the upper end and noise at the lower end. For example, a cell phone coming close to a base station may receive a very large signal and must process it with

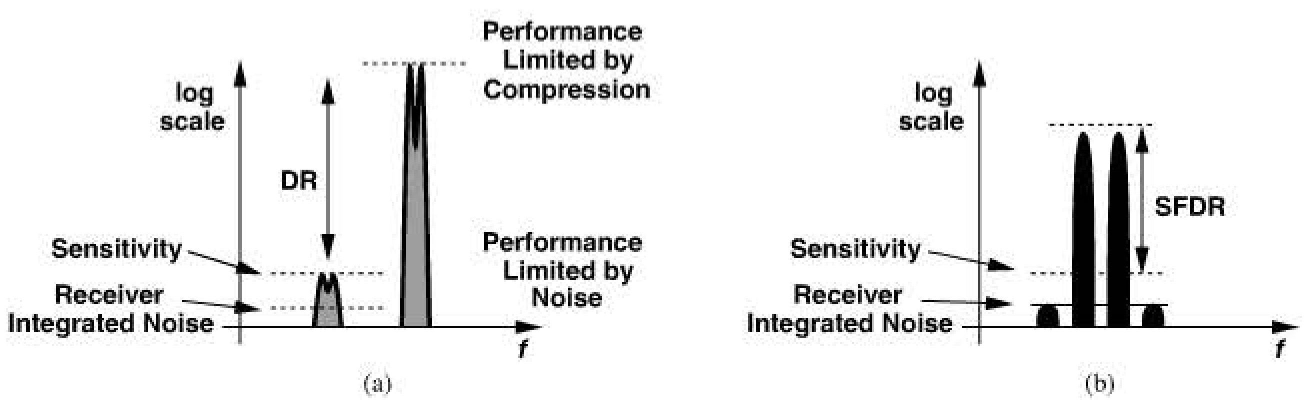



**Figure 2.56** *Definitions of (a) DR and (b) SFDR.*



acceptable distortion. In fact, the cell phone measures the signal strength and adjusts the receiver gain so as to avoid compression. Excluding interferers, this "compression-based" DR can exceed 100 dB because the upper end can be raised relatively easily.

The second type, called the "spurious-free dynamic range" (SFDR), represents limitations arising from both noise and interference. The lower end is still equal to the sensitivity, but the upper end is defined as the maximum input level in a *two-tone* test for which the third-order IM products do not exceed the integrated noise of the receiver. As shown in Fig. 2.56(b), two (modulated or unmodulated) tones having equal amplitudes are applied and their level is raised until the IM products reach the integrated noise.[23] The ratio of the power of each tone to the sensitivity yields the SFDR. The SFDR represents the maximum relative level of interferers that a receiver can tolerate while producing an acceptable signal quality from a small input level.

Where should the various levels depicted in Fig. 2.56(b) be measured, at the input of the circuit or at its output? Since the IM components appear only at the output, the output port serves as a more natural candidate for such a measurement. In this case, the sensitivity—usually an input-referred quantity—must be scaled by the gain of the circuit so that it is referred to the output. Alternatively, the output IM magnitudes can be divided by the gain so that they are referred to the input. We follow the latter approach in our SFDR calculations.

To determine the upper end of the SFDR, we rewrite Eq. (2.56) as

$$P_{IIP3} = P_{in} + \frac{P_{out} - P_{IM,out}}{2}, \tag{2.152}$$

where, for the sake of brevity, we have denoted $20 \log A_x$ as $P_x$ even though no actual power may be transferred at the input or output ports. Also, $P_{IM,out}$ represents the level of IM products at the output. If the circuit exhibits a gain of $G$ (in dB), then we can refer the IM level to the input by writing $P_{IM,in} = P_{IM,out} - G$. Similarly, the *input* level of each tone is given by $P_{in} = P_{out} - G$. Thus, (2.152) reduces to

$$P_{IIP3} = P_{in} + \frac{P_{in} - P_{IM,in}}{2} \tag{2.153}$$

$$= \frac{3P_{in} - P_{IM,in}}{2}, \tag{2.154}$$

and hence

$$P_{in} = \frac{2P_{IIP3} + P_{IM,in}}{3}. \tag{2.155}$$

The upper end of the SFDR is that value of $P_{in}$ which makes $P_{IM,in}$ equal to the integrated noise of the receiver:

$$P_{in,max} = \frac{2P_{IIP3} + (-174\,\text{dBm} + NF + 10 \log B)}{3}. \tag{2.156}$$

23. Note that the integrated noise is a single value (e.g., 100 $\mu V_{rms}$), not a *density*.

The SFDR is the difference (in dB) between $P_{in,max}$ and the sensitivity:

$$SFDR = P_{in,max} - (-174\,\text{dBm} + NF + 10\log B + SNR_{min}) \tag{2.157}$$

$$= \frac{2(P_{IIP3} + 174\,\text{dBm} - NF - 10\log B)}{3} - SNR_{min}. \tag{2.158}$$

For example, a GSM receiver with $NF = 7\,\text{dB}$, $P_{IIP3} = -15\,\text{dBm}$, and $SNR_{min} = 12\,\text{dB}$ achieves an SFDR of 54 dB, a substantially lower value than the dynamic range in the absence of interferers.

**Example 2.26**

The upper end of the dynamic range is limited by intermodulation in the presence of *two* interferers or desensitization in the presence of *one* interferer. Compare these two cases and determine which one is more restrictive.

**Solution:**

We must compare the upper end expressed by Eq. (2.156) with the 1-dB compression point:

$$P_{1-dB} \overset{?}{\gtrless} P_{in,max}. \tag{2.159}$$

Since $P_{1-dB} = P_{IIP3} - 9.6\,\text{dB}$,

$$P_{IIP3} - 9.6\,\text{dB} \overset{?}{\gtrless} \frac{2P_{IIP3} + (-174\,\text{dBm} + NF + 10\log B)}{3} \tag{2.160}$$

and hence

$$P_{IIP3} - 28.8\,\text{dB} \overset{?}{\gtrless} -174\,\text{dBm} + NF + 10\log B. \tag{2.161}$$

Since the right-hand side represents the receiver noise floor, we expect it to be much lower than the left-hand side. In fact, even for an extremely wideband channel of $B = 1$ GHz and $NF = 10\,\text{dB}$, the right-hand side is equal to $-74$ dBm, whereas, with a typical $P_{IIP3}$ of $-10$ to $-25$ dBm, the left-hand side still remains higher. It is therefore plausible to conclude that

$$P_{1-dB} > P_{in,max}. \tag{2.162}$$

It follows that the maximum tolerable level in a two-tone test is quite lower than that in a compression test, i.e., corruption by intermodulation between two interferers is much greater than compression due to one. The SFDR is therefore a more stringent characteristic of the system than the compression-based dynamic range.

## 2.5 PASSIVE IMPEDANCE TRANSFORMATION

At radio frequencies, we often employ passive networks to transform impedances—from high to low and vice versa, or from complex to real and vice versa. Called "matching networks," such circuits do not easily lend themselves to integration because their constituent devices, particularly inductors, suffer from loss if built on silicon chips. (We do use on-chip inductors in many RF building blocks.) Nonetheless, a basic understanding of impedance transformation is essential.



### 2.5.1 Quality Factor

In its simplest form, the quality factor, $Q$, indicates how close to ideal an energy-storing device is. An ideal capacitor dissipates no energy, exhibiting an infinite $Q$, but a series resistance, $R_S$ [Fig. 2.57(a)], reduces its $Q$ to

$$Q_S = \frac{\frac{1}{C\omega}}{R_S}, \tag{2.163}$$

where the numerator denotes the "desired" component and the denominator, the "undesired" component. If the resistive loss in the capacitor is modeled by a *parallel* resistance [Fig. 2.57(b)], then we must define the $Q$ as

$$Q_P = \frac{R_P}{\frac{1}{C\omega}}, \tag{2.164}$$

because an ideal (infinite $Q$) results only if $R_P = \infty$. As depicted in Figs. 2.57(c) and (d), similar concepts apply to inductors

$$Q_S = \frac{L\omega}{R_S} \tag{2.165}$$

$$Q_P = \frac{R_P}{L\omega}. \tag{2.166}$$

While a parallel resistance appears to have no physical meaning, modeling the loss by $R_P$ proves useful in many circuits such as amplifiers and oscillators (Chapters 5 and 8). We will also introduce other definitions of Q in Chapter 8.

### 2.5.2 Series-to-Parallel Conversion

Before studying transformation techniques, let us consider the series and parallel $RC$ sections shown in Fig. 2.58. What choice of values makes the two networks equivalent?

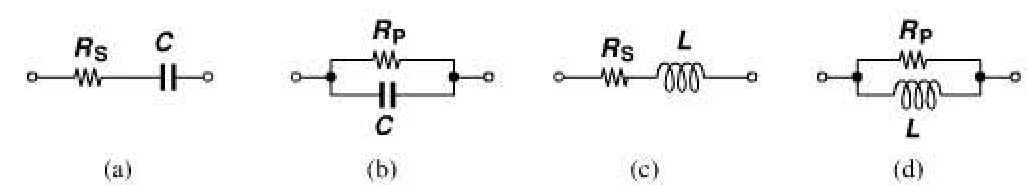



**Figure 2.57** *(a) Series RC circuit, (b) equivalent parallel circuit, (c) series RL circuit, (d) equivalent parallel circuit.*

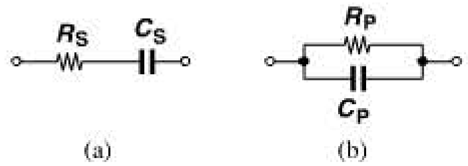



**Figure 2.58** *Series-to-parallel conversion.*

Equating the impedances,

$$\frac{R_S C_S s + 1}{C_S s} = \frac{R_P}{R_P C_P s + 1}, \tag{2.167}$$

and substituting $j\omega$ for $s$, we have

$$R_P C_S j\omega = 1 - R_P C_P R_S C_S \omega^2 + (R_P C_P + R_S C_S) j\omega, \tag{2.168}$$

and hence

$$R_P C_P R_S C_S \omega^2 = 1 \tag{2.169}$$

$$R_P C_P + R_S C_S - R_P C_S = 0. \tag{2.170}$$

Equation (2.169) implies that $Q_S = Q_P$.

Of course, the two impedances cannot remain equal at all frequencies. For example, the series section approaches an open circuit at low frequencies while the parallel section does not. Nevertheless, an approximation allows equivalence for a narrow frequency range. We first substitute for $R_P C_P$ in (2.169) from (2.170), obtaining

$$R_P = \frac{1}{R_S C_S^2 \omega^2} + R_S. \tag{2.171}$$

Utilizing the definition of $Q_S$ in (2.163), we have

$$R_P = (Q_S^2 + 1) R_S. \tag{2.172}$$

Substitution in (2.169) thus yields

$$C_P = \frac{Q_S^2}{Q_S^2 + 1} C_S. \tag{2.173}$$

So long as $Q_S^2 \gg 1$ (which is true for a finite frequency range),

$$R_P \approx Q_S^2 R_S \tag{2.174}$$

$$C_P \approx C_S. \tag{2.175}$$

That is, the series-to-parallel conversion retains the value of the capacitor but raises the resistance by a factor of $Q_S^2$. These approximations for $R_P$ and $C_P$ are relatively accurate because the quality factors encountered in practice typically exceed 4. Conversely,



parallel-to-series conversion reduces the resistance by a factor of $Q_P^2$. This statement applies to $RL$ sections as well.

### 2.5.3 Basic Matching Networks

A common situation in RF transmitter design is that a load resistance must be transformed to a lower value. The circuit shown in Fig. 2.59(a) accomplishes this task. As mentioned above, the capacitor in parallel with $R_L$ converts this resistance to a lower *series* component [Fig. 2.59(b)]. The inductance is inserted to cancel the equivalent series capacitance.

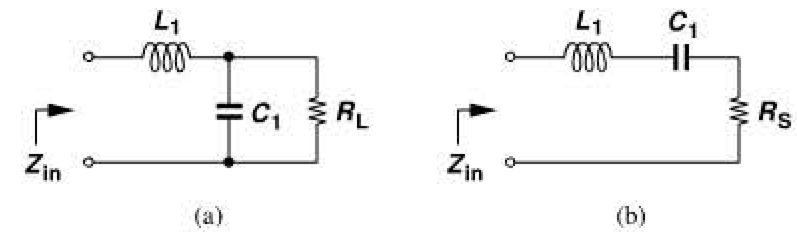



**Figure 2.59** *(a) Matching network, (b) equivalent circuit.*

Writing $Z_{in}$ from Fig. 2.59(a) and replacing $s$ with $j\omega$, we have

$$Z_{in}(j\omega) = \frac{R_L(1 - L_1 C_1 \omega^2) + jL_1\omega}{1 + jR_L C_1 \omega}. \tag{2.176}$$

Thus,

$$Re\{Z_{in}\} = \frac{R_L}{1 + R_L^2 C_1^2 \omega^2} \tag{2.177}$$

$$= \frac{R_L}{1 + Q_P^2}, \tag{2.178}$$

indicating that $R_L$ is transformed down by a factor of $1 + Q_P^2$. Also, setting the imaginary part to zero gives

$$L_1 = \frac{R_L^2 C_1}{1 + R_L^2 C_1^2 \omega^2} \tag{2.179}$$

$$= \frac{R_L^2 C_1}{1 + Q_P^2}. \tag{2.180}$$

If $Q_P^2 \gg 1$, then

$$Re\{Z_{in}\} \approx \frac{1}{R_L C_1^2 \omega^2} \tag{2.181}$$

$$L_1 = \frac{1}{C_1 \omega^2}. \tag{2.182}$$

The following example illustrates how the component values are chosen.

### Example 2.27

Design the matching network of Fig. 2.59(a) so as to transform $R_L = 50\ \Omega$ to $25\ \Omega$ at a center frequency of 5 GHz.

**Solution:**

Assuming $Q_P^2 \gg 1$, we have from Eqs. (2.181) and (2.182), $C_1 = 0.90$ pF and $L_1 = 1.13$ nH, respectively. Unfortunately, however, $Q_P = 1.41$, indicating that Eqs. (2.178) and (2.180) must be used instead. We thus obtain $C_1 = 0.637$ pF and $L_1 = 0.796$ nH.

In order to transform a resistance to a higher value, the capacitive network shown in Fig. 2.60(a) can be used. The series-parallel conversion results derived previously provide insight here. If $Q^2 \gg 1$, the parallel combination of $C_1$ and $R_L$ can be converted to a series network [Fig. 2.60(b)], where $R_S \approx [R_L(C_1\omega)^2]^{-1}$ and $C_S \approx C_1$. Viewing $C_2$ and $C_1$ as one capacitor, $C_{eq}$, and converting the resulting series section to a parallel circuit [Fig. 2.60(c)], we have

$$R_{tot} = \frac{1}{R_S(C_{eq}\omega)^2} \tag{2.183}$$

$$= \left(1 + \frac{C_1}{C_2}\right)^2 R_L. \tag{2.184}$$

That is, the network "boosts" the value of $R_L$ by a factor of $(1 + C_1/C_2)^2$. Also,

$$C_{eq} = \frac{C_1 C_2}{C_1 + C_2}. \tag{2.185}$$

Note that the capacitive component must be cancelled by placing an inductor in *parallel* with the input.

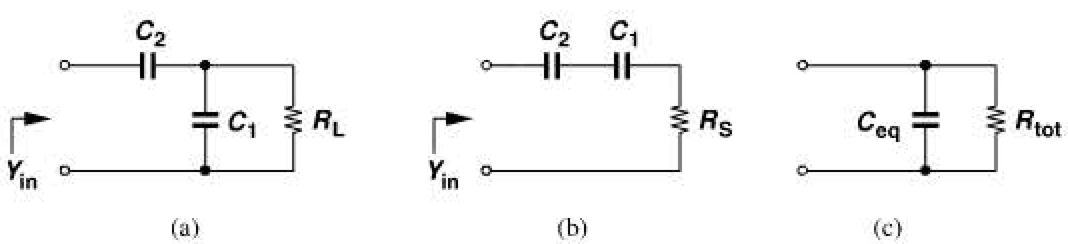



**Figure 2.60** *(a) Capacitive matching circuit, (b) simplified circuit with parallel-to-series conversion, (c) simplified circuit with series-to-parallel conversion.*

For low $Q$ values, the above derivations incur significant error. We thus compute the input admittance $(1/Y_{in})$ and replace $s$ with $j\omega$,

$$Y_{in} = \frac{j\omega C_2(1 + j\omega R_L C_1)}{1 + R_L(C_1 + C_2)j\omega}. \tag{2.186}$$



The real part of $Y_{in}$ yields the equivalent resistance seen to ground if we write

$$R_{tot} = \frac{1}{Re\{Y_{in}\}} \tag{2.187}$$

$$= \frac{1}{R_L C_2^2 \omega^2} + R_L\left(1 + \frac{C_1}{C_2}\right)^2. \tag{2.188}$$

In comparison with Eq. (2.184), this result contains an additional component, $(R_L C_2^2 \omega^2)^{-1}$.

### Example 2.28

Determine how the circuit shown in Fig. 2.61(a) transforms $R_L$.

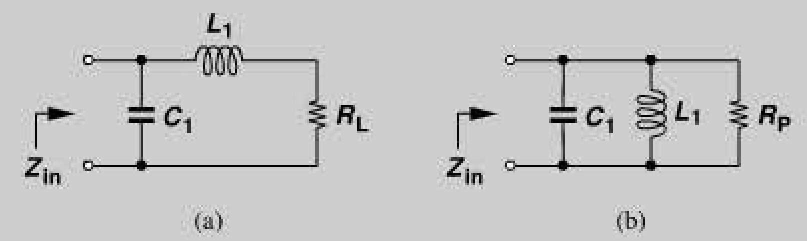



**Figure 2.61** *(a) Matching network, (b) simplified circuit.*

**Solution:**

We postulate that conversion of the $L_1$–$R_L$ branch to a parallel section produces a higher resistance. If $Q_S^2 = (L_1\omega/R_L)^2 \gg 1$, then the equivalent parallel resistance is obtained from Eq. (2.174) as

$$R_P = Q_S^2 R_L \tag{2.189}$$

$$= \frac{L_1^2\omega^2}{R_L}. \tag{2.190}$$

The parallel equivalent inductance is approximately equal to $L_1$ and is cancelled by $C_1$ [Fig. 2.61(b)].

The intuition gained from our analysis of matching networks leads to the four "$L$-section" topologies[24] shown in Fig. 2.62. In Fig. 2.62(a), $C_1$ transforms $R_L$ to a smaller series value and $L_1$ cancels $C_1$. Similarly, in Fig. 2.62(b), $L_1$ transforms $R_L$ to a smaller series value while $C_1$ resonates with $L_1$. In Fig. 2.62(c), $L_1$ transforms $R_L$ to a larger parallel value and $C_1$ cancels the resulting parallel inductance. A similar observation applies to Fig. 2.62(d).

How do these networks transform voltages and currents? As an example, consider the circuit in Fig. 2.62(a). For a sinusoidal input voltage with an rms value of $V_{in}$, the power

24. The term "$L$" is used because the capacitor and the inductor form the letter $L$ in the circuit diagram.

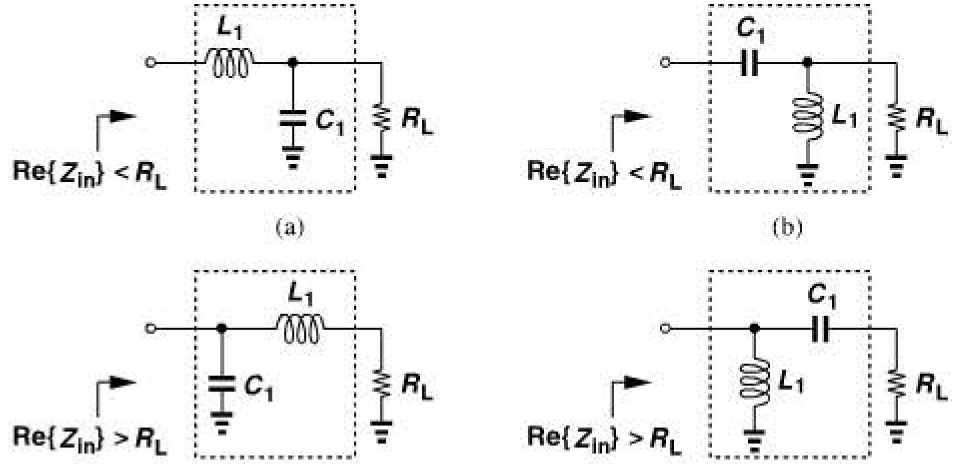



**Figure 2.62** *Four L sections used for matching.*

delivered to the input port is equal to $V_{in}^2/Re\{Z_{in}\}$, and that delivered to the load, $V_{out}^2/R_L$. If $L_1$ and $C_1$ are ideal, these two powers must be equal, yielding

$$\frac{V_{out}}{V_{in}} = \sqrt{\frac{R_L}{Re\{Z_{in}\}}}. \tag{2.191}$$

This result, of course, applies to any lossless matching network whose input impedance contains a zero imaginary part. Since $P_{in} = V_{in}I_{in}$ and $P_{out} = V_{out}I_{out}$, we also have

$$\frac{I_{out}}{I_{in}} = \sqrt{\frac{Re\{Z_{in}\}}{R_L}}. \tag{2.192}$$

For example, a network transforming $R_L$ to a *lower* value "amplifies" the voltage and attenuates the current by the above factor.

**Example 2.29**

A closer look at the $L$-sections in Figs. 2.62(a) and (c) suggests that one can be obtained from the other by swapping the input and output ports. Is it possible to generalize this observation?

**Solution:**

Yes, it is. Consider the arrangement shown in Fig. 2.63(a), where the passive network transforms $R_L$ by a factor of $\alpha$. Assuming the input port exhibits no imaginary component, we equate the power delivered to the network to the power delivered to the load:

$$\left(V_{in}\frac{\alpha R_L}{\alpha R_L + R_S}\right)^2 \cdot \frac{1}{\alpha R_L} = \frac{V_{out}^2}{R_L}. \tag{2.193}$$



**Example 2.29** (*Continued*)

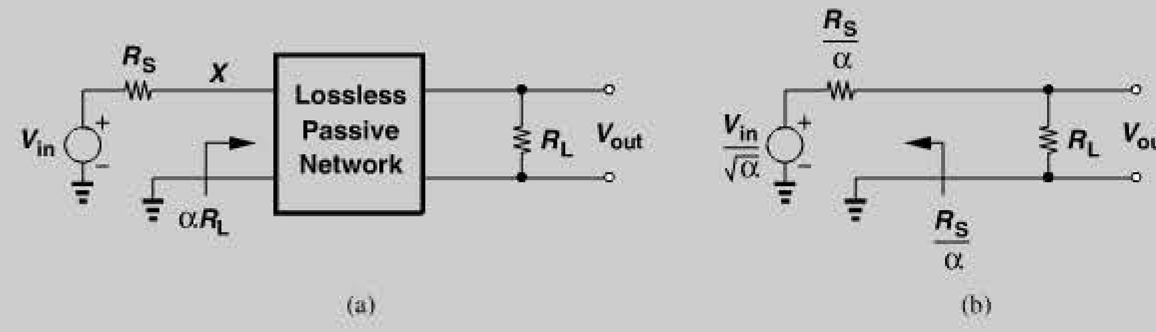



**Figure 2.63** *(a) Input and (b) output impedances of a lossless passive network.*

It follows that

$$V_{out} = \frac{V_{in}}{\sqrt{\alpha}} \cdot \frac{R_L}{R_L + \frac{R_S}{\alpha}}, \tag{2.194}$$

pointing to the Thevenin equivalent shown in Fig. 2.63(b). We observe that the network transforms $R_S$ by a factor of $1/\alpha$ and the input voltage by a factor of $1/\sqrt{\alpha}$, similar to that in Eq. (2.191). In other words, if the input and output ports of such a network are swapped, the resistance transformation ratio is simply inverted.

Transformers can also transform impedances. An ideal transformer having a turns ratio of $n$ "amplifies" the input voltage by a factor of $n$ (Fig. 2.64). Since no power is lost, $V_{in}^2/R_{in} = n^2V_{in}^2/R_L$ and hence $R_{in} = R_L/n^2$. The behavior of actual transformers, especially those fabricated monolithically, is studied in Chapter 7.

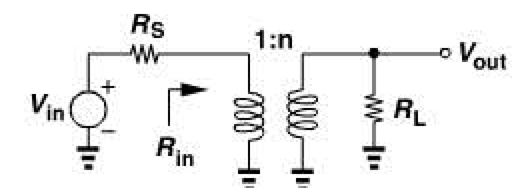



**Figure 2.64** *Impedance transformation by a physical transformer.*

The networks studied here operate across only a narrow bandwidth because the transformation ratio, e.g., $1 + Q^2$, varies with frequency, and the capacitance and inductance approximately resonate over a narrow frequency range. Broadband matching networks can be constructed, but they typically suffer from a high loss.

### 2.5.4 Loss in Matching Networks

Our study of matching networks has thus far neglected the loss of their constituent components, particularly, that of inductors. We analyze the effect of loss in a few cases here, but, in general, simulations are necessary to determine the behavior of complex lossy networks.

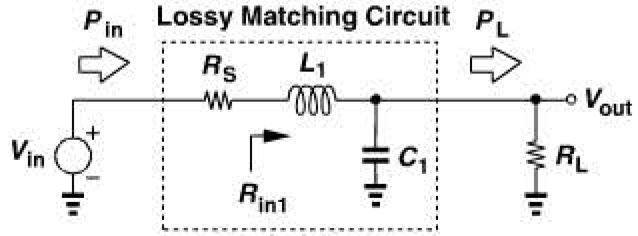



**Figure 2.65** *Lossy matching network with series resistence.*

Consider the matching network of Fig. 2.62(a), shown in Fig. 2.65 with the loss of $L_1$ modeled by a series resistance, $R_S$. We define the loss as the power provided by the input divided by that delivered to $R_L$. The former is equal to

$$P_{in} = \frac{V_{in}^2}{R_S + R_{in1}} \tag{2.195}$$

and the latter,

$$P_L = \left(V_{in}\frac{R_{in1}}{R_S + R_{in1}}\right)^2 \cdot \frac{1}{R_{in1}}, \tag{2.196}$$

because the power delivered to $R_{in1}$ is entirely absorbed by $R_L$. It follows that

$$\text{Loss} = \frac{P_{in}}{P_L} \tag{2.197}$$

$$= 1 + \frac{R_S}{R_{in1}}. \tag{2.198}$$

For example, if $R_S = 0.1R_{in1}$, then the (power) loss reaches 0.41 dB. Note that this network transforms $R_L$ to a *lower* value, $R_{in1} = R_L/(1 + Q_P^2)$, thereby suffering from loss even if $R_S$ appears small.

As another example, consider the network of Fig. 2.62(b), depicted in Fig. 2.66 with the loss of $L_1$ modeled by a parallel resistance, $R_P$. We note that the power delivered by $V_{in}$, $P_{in}$, is entirely absorbed by $R_P||R_L$:

$$P_{in} = \frac{V_{out}^2}{R_P||R_L} \tag{2.199}$$

$$= \frac{V_{out}^2}{R_L}\frac{R_P + R_L}{R_P}. \tag{2.200}$$

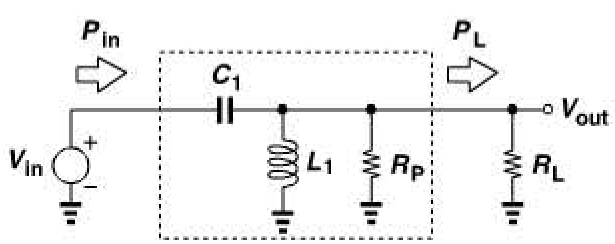



**Figure 2.66** *Lossy matching network with parallel resistence.*



Recognizing $V_{out}^2/R_L$ as the power delivered to the load, $P_L$, we have

$$\text{Loss} = 1 + \frac{R_L}{R_P}. \tag{2.201}$$

For example, if $R_P = 10R_L$, then the loss is equal to 0.41 dB.

## 2.6 SCATTERING PARAMETERS

Microwave theory deals mostly with power quantities rather than voltage or current quantities. Two reasons can explain this approach. First, traditional microwave design is based on transfer of *power* from one stage to the next. Second, the measurement of high-frequency voltages and currents in the laboratory proves very difficult, whereas that of average power is more straightforward. Microwave theory therefore models devices, circuits, and systems by parameters that can be obtained through the measurement of power quantities. They are called "scattering parameters" (S-parameters).

Before studying S-parameters, we introduce an example that provides a useful viewpoint. Consider the $L_1$–$C_1$ series combination depicted in Fig. 2.67. The circuit is driven by a sinusoidal source, $V_{in}$, having an output impedance of $R_S$. A load resistance of $R_L = R_S$ is tied to the output port. At an input frequency of $\omega = (\sqrt{L_1C_1})^{-1}$, $L_1$ and $C_1$ form a short circuit, providing a conjugate match between the source and the load. In analogy with transmission lines, we say the "incident wave" produced by the signal source is absorbed by $R_L$. At other frequencies, however, $L_1$ and $C_1$ attenuate the voltage delivered to $R_L$. Equivalently, we say the input port of the circuit generates a "reflected wave" that returns to the source. In other words, the difference between the incident power (the power that would be delivered to a matched load) and the reflected power represents the power delivered to the circuit.

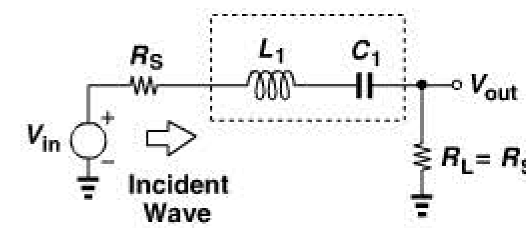



**Figure 2.67** *Incident wave in a network.*

The above viewpoint can be generalized for any two-port network. As illustrated in Fig. 2.68, we denote the incident and reflected waves at the input port by $V_1^+$ and $V_1^-$, respectively. Similar waves are denoted by $V_2^+$ and $V_2^-$, respectively, at the output. Note

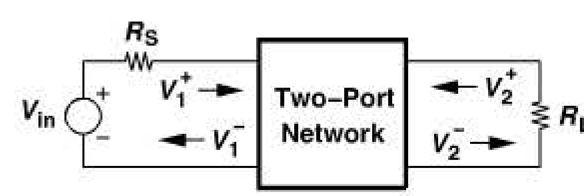



**Figure 2.68** *Illustration of incident and reflected waves at the input and output.*

that $V_1^+$ denotes a wave generated by $V_{in}$ as if the input impedance of the circuit were equal to $R_S$. Since that may not be the case, we include the reflected wave, $V_1^-$, so that the actual voltage measured at the input is equal to $V_1^+ + V_1^-$. Also, $V_2^+$ denotes the incident wave traveling *into* the output port or, equivalently, the wave *reflected* from $R_L$. These four quantities are uniquely related to one another through the S-parameters of the network:

$$V_1^- = S_{11}V_1^+ + S_{12}V_2^+ \tag{2.202}$$
$$V_2^- = S_{21}V_1^+ + S_{22}V_2^+. \tag{2.203}$$

With the aid of Fig. 2.69, we offer an intuitive interpretation for each parameter:

1. For $S_{11}$, we have from Fig. 2.69(a)



$$S_{11} = \frac{V_1^-}{V_1^+}|_{V_2^+=0}. \tag{2.204}$$

Thus, $S_{11}$ is the ratio of the reflected and incident waves at the input port when the reflection from $R_L$ (i.e., $V_2^+$) is zero. This parameter represents the accuracy of the input matching.

2. For $S_{12}$, we have from Fig. 2.69(b)

$$S_{12} = \frac{V_1^-}{V_2^+}|_{V_1^+=0}. \tag{2.205}$$

Thus, $S_{12}$ is the ratio of the reflected wave at the input port to the incident wave into the output port when the input port is matched. In this case, the *output* port is driven by the signal source. This parameter characterizes the "reverse isolation" of the circuit, i.e., how much of the output signal couples to the input network.

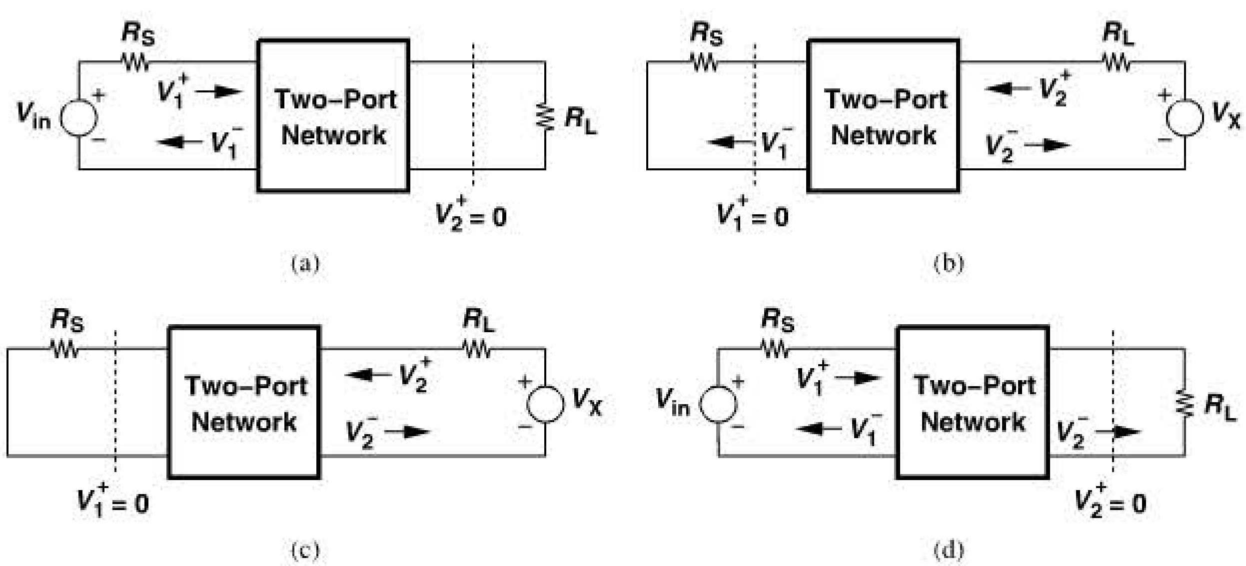



**Figure 2.69** *Illustration of four S-parameters.*



3. For $S_{22}$, we have from Fig. 2.69(c)

$$S_{22} = \frac{V_2^-}{V_2^+}|_{V_1^+=0}. \tag{2.206}$$

Thus, $S_{22}$ is the ratio of reflected and incident waves at the output when the reflection from $R_S$ (i.e., $V_1^+$) is zero. This parameter represents the accuracy of the output matching.

4. For $S_{21}$, we have from Fig. 2.69(d)

$$S_{21} = \frac{V_2^-}{V_1^+}|_{V_2^+=0}. \tag{2.207}$$

Thus, $S_{21}$ is the ratio of the wave incident on the load to that going to the input when the reflection from $R_L$ is zero. This parameter represents the gain of the circuit.

We should make a few remarks at this point. First, S-parameters generally have frequency-dependent complex values. Second, we often express S-parameters in units of dB as follows

$$S_{mn}|_{dB} = 20\log|S_{mn}|. \tag{2.208}$$

Third, the condition $V_2^+ = 0$ in Eqs. (2.204) and (2.207) requires that the reflection from $R_L$ be zero, but it does *not* mean that the output port of the circuit must be conjugate-matched to $R_L$. This condition simply means that if, hypothetically, a transmission line having a characteristic impedance equal to $R_S$ carries the output signal to $R_L$, then no wave is reflected from $R_L$. A similar note applies to the requirement $V_1^+ = 0$ in Eqs. (2.205) and (2.206). The conditions $V_1^+ = 0$ at the input or $V_2^+ = 0$ at the output facilitate high-frequency measurements while creating issues in modern RF design. As mentioned in Section 2.3.5 and exemplified by the cascade of stages in Fig. 2.53, modern RF design typically does not strive for matching between stages. Thus, if $S_{11}$ of the first stage must be measured with $R_L = R_S$ at its output, then its value may not represent the $S_{11}$ of the cascade.

In modern RF design, $S_{11}$ is the most commonly-used S parameter as it quantifies the accuracy of impedance matching at the input of receivers. Consider the arrangement shown in Fig. 2.70, where the receiver exhibits an input impedance of $Z_{in}$. The incident wave $V_1^+$ is given by $V_{in}/2$ (as if $Z_{in}$ were equal to $R_S$). Moreover, the total voltage at the receiver

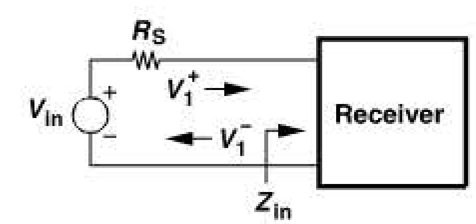



**Figure 2.70** *Receiver with incident and reflected waves.*

input is equal to $V_{in}Z_{in}/(Z_{in}+R_S)$, which is also equal to $V_1^+ + V_1^-$. Thus,

$$V_1^- = V_{in}\frac{Z_{in}}{Z_{in}+R_S} - \frac{V_{in}}{2} \tag{2.209}$$

$$= \frac{Z_{in}-R_S}{2(Z_{in}+R_S)}V_{in}. \tag{2.210}$$

It follows that

$$\frac{V_1^-}{V_1^+} = \frac{Z_{in}-R_S}{Z_{in}+R_S}. \tag{2.211}$$

Called the "input reflection coefficient" and denoted by $\Gamma_{in}$, this quantity can also be considered to be $S_{11}$ if we remove the condition $V_2^+ = 0$ in Eq. (2.204).

**Example 2.30**

Determine the S-parameters of the common-gate stage shown in Fig. 2.71(a). Neglect channel-length modulation and body effect.

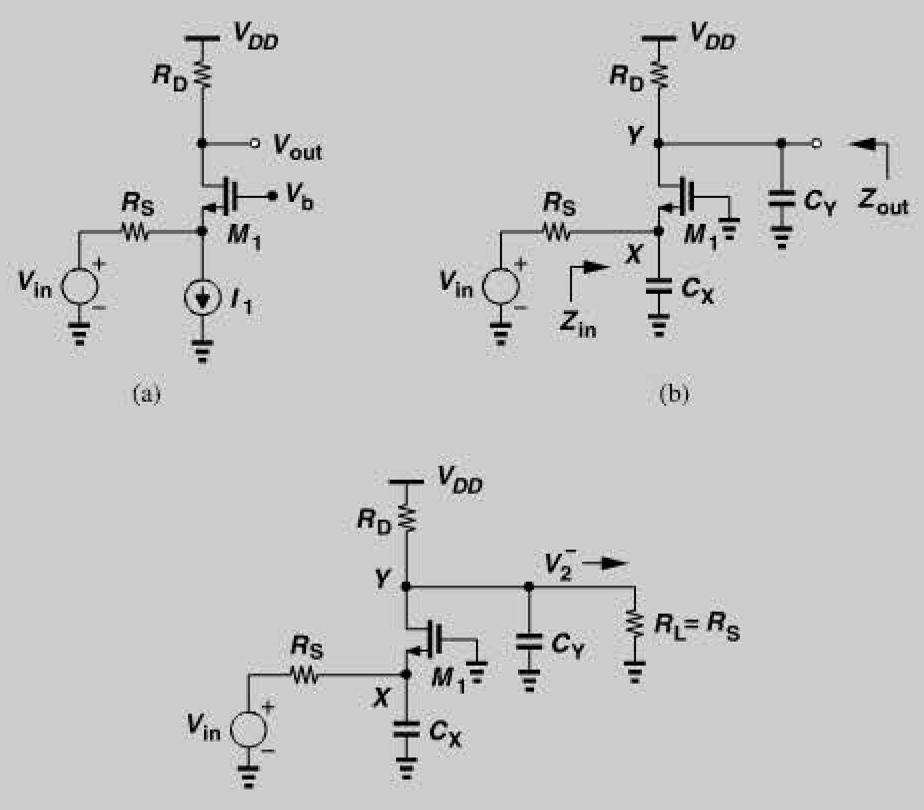



**Figure 2.71** *(a) CG stage for calculation of S-parameters, (b) inclusion of capacitors, (c) effect of reflected wave at output.*



**Example 2.30** *(Continued)*

**Solution:**

Drawing the circuit as shown in Fig. 2.71(b), where $C_X = C_{GS}+C_{SB}$ and $C_Y = C_{GD}+C_{DB}$, we write $Z_{in} = (1/g_m)||(C_Xs)^{-1}$ and

$$S_{11} = \frac{Z_{in}-R_S}{Z_{in}+R_S} \tag{2.212}$$

$$= \frac{1-g_mR_S-C_Xs}{1+g_mR_S+C_Xs}. \tag{2.213}$$

For $S_{12}$, we recognize that the arrangement of Fig. 2.71(b) yields no coupling from the output to the input if channel-length modulation is neglected. Thus, $S_{12}=0$. For $S_{22}$, we note that $Z_{out} = R_D||(C_Ys)^{-1}$ and hence

$$S_{22} = \frac{Z_{out}-R_S}{Z_{out}+R_S} \tag{2.214}$$

$$= -\frac{R_S-R_D+R_SR_DC_Ys}{R_S+R_D+R_SR_DC_Ys}. \tag{2.215}$$

Lastly, $S_{21}$ is obtained according to the configuration of Fig. 2.71(c). Since $V_2^-/V_{in} = (V_2^-/V_X)(V_X/V_{in})$, $V_2^-/V_X = g_m[R_D||R_S||(C_Ys)^{-1}]$, and $V_X/V_{in} = Z_{in}/(Z_{in}+R_S)$, we obtain

$$\frac{V_2^-}{V_{in}} = g_m\left(R_D||R_S||\frac{1}{C_Ys}\right)\frac{1}{1+g_mR_S+R_SC_Xs}. \tag{2.216}$$

It follows that

$$S_{21} = 2g_m\left(R_D||R_S||\frac{1}{C_Ys}\right)\frac{1}{1+g_mR_S+R_SC_Xs}. \tag{2.217}$$

## 2.7 ANALYSIS OF NONLINEAR DYNAMIC SYSTEMS[25]

In our treatment of systems in Section 2.2, we have assumed a static nonlinearity, e.g., in the form of $y(t) = \alpha_1x(t)+\alpha_2x^2(t)+\alpha_3x^3(t)$. In some cases, a circuit may exhibit dynamic nonlinearity, requiring a more complex analysis. In this section, we address this task.

### 2.7.1 Basic Considerations

Let us first consider a general nonlinear system with an input given by $x(t) = A_1\cos\omega_1t + A_2\cos\omega_2t$. We expect the output, $y(t)$, to contain harmonics at $n\omega_1$, $m\omega_2$, and IM products

25. This section can be skipped in a first reading.

at $k\omega_1 \pm q\omega_2$, where, $n$, $m$, $k$, and $q$ are integers. In other words,

$$y(t) = \sum_{n=1}^{\infty} a_n \cos(n\omega_1 t + \theta_n) + \sum_{n=1}^{\infty} b_n \cos(n\omega_2 t + \phi_n)$$
$$+ \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} c_{m,n} \cos(n\omega_1 t + m\omega_2 t + \phi_{n,m}). \qquad (2.218)$$

In the above equation, $a_n$, $b_n$, $c_{m,n}$, and the phase shifts are frequency-dependent quantities. If the differential equation governing the system is known, we can simply substitute for $y(t)$ from this expression, equate the like terms, and compute $a_n$, $b_n$, $c_{m,n}$, and the phase shifts. For example, consider the simple RC section shown in Fig. 2.72, where the capacitor is nonlinear and expressed as $C_1 = C_0(1 + \alpha V_{out})$. Adding the voltages across $R_1$ and $C_1$ and equating the result to $V_{in}$, we have

$$R_1 C_0(1 + \alpha V_{out})\frac{dV_{out}}{dt} + V_{out} = V_{in}. \qquad (2.219)$$

Now suppose $V_{in}(t) = V_0 \cos \omega_1 t + V_0 \cos \omega_2 t$ (as in a two-tone test) and assume the system is only "weakly" nonlinear, i.e., only the output terms at $\omega_1$, $\omega_2$, $\omega_1 \pm \omega_2$, $2\omega_1 \pm \omega_2$, and $2\omega_2 \pm \omega_1$ are significant. Thus, the output assumes the form

$$\begin{aligned} V_{out}(t) = {} & a_1 \cos(\omega_1 t + \phi_1) + b_1 \cos(\omega_2 t + \phi_2) + c_1 \cos[(\omega_1 + \omega_2)t + \phi_3] \\ & + c_2 \cos[(\omega_1 - \omega_2)t + \phi_4] + c_3 \cos[(2\omega_1 + \omega_2)t + \phi_5] \\ & + c_4 \cos[(\omega_1 + 2\omega_2)t + \phi_6] + c_5 \cos[(2\omega_1 - \omega_2)t + \phi_7] \\ & + c_6 \cos[(\omega_1 - 2\omega_2)t + \phi_8], \end{aligned} \qquad (2.220)$$

where, for simplicity, we have used $c_m$ and $\phi_m$. We must now substitute for $V_{out}(t)$ and $V_{in}(t)$ in (2.219), convert products of sinusoids to sums, bring all of the terms to one side of the equation, group them according to their frequencies, and equate the coefficient of each sinusoid to zero. We thus obtain a system of 16 nonlinear equations and 16 knowns ($a_1$, $b_1$, $c_1, \ldots, c_6, \phi_1, \ldots, \phi_8$).

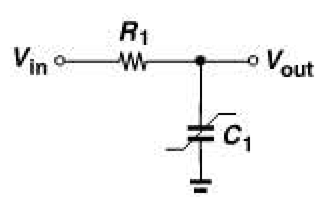



**Figure 2.72** *RC circuit with nonlinear capacitor.*

This type of analysis is called "harmonic balance" because it predicts the output frequencies and attempts to "balance" the two sides of the circuit's differential equation by including these components in $V_{out}(t)$. The mathematical labor in harmonic balance makes hand analysis difficult or impossible. The "Volterra series" approach, on the other hand, prescribes a *recursive* method that computes the response more accurately in successive



steps *without* the need for solving nonlinear equations. A detailed treatment of the concepts described below can be found in [10–14].

## 2.8 VOLTERRA SERIES

In order to understand how the Volterra series represents the time response of a system, we begin with a simple input form, $V_{in}(t) = V_0 \exp(j\omega_1 t)$. Of course, if we wish to obtain the response to a sinusoid of the form $V_0 \cos \omega_1 t = Re\{V_0 \exp(j\omega_1 t)\}$, we simply calculate the real part of the output.[26] (The use of the exponential form greatly simplifies the manipulation of the product terms.) For a linear, time-invariant system, the output is given by

$$V_{out}(t) = H(\omega_1)V_0 \exp(j\omega_1 t), \qquad (2.221)$$

where $H(\omega_1)$ is the Fourier transform of the impulse response. For example, if the capacitor in Fig. 2.72 is linear, i.e., $C_1 = C_0$, then we can substitute for $V_{out}$ and $V_{in}$ in Eq. (2.219):

$$R_1 C_0 H(\omega_1)(j\omega_1)V_0 \exp(j\omega_1 t) + H(\omega_1)V_0 \exp(j\omega_1 t) = V_0 \exp(j\omega_1 t). \qquad (2.222)$$

It follows that

$$H(\omega_1) = \frac{1}{R_1 C_0 j\omega_1 + 1}. \qquad (2.223)$$

Note that the phase shift introduced by the circuit is included in $H(\omega_1)$ here.

As our next step, let us ask, how should the output response of a dynamic nonlinear system be expressed? To this end, we apply two tones to the input, $V_{in}(t) = V_0 \exp(j\omega_1 t) + V_0 \exp(j\omega_2 t)$, recognizing that the output consists of both linear and nonlinear responses. The former are of the form

$$V_{out1}(t) = H(\omega_1)V_0 \exp(j\omega_1 t) + H(\omega_2)V_0 \exp(j\omega_2 t), \qquad (2.224)$$

and the latter include exponentials such as $\exp[j(\omega_1 + \omega_2)t]$, etc. We expect that the coefficient of such an exponential is a function of both $\omega_1$ and $\omega_2$. We thus make a slight change in our notation: we denote $H(\omega_j)$ in Eq. (2.224) by $H_1(\omega_j)$ [to indicate first-order (linear) terms] and the coefficient of $\exp[j(\omega_1 + \omega_2)t]$ by $H_2(\omega_1, \omega_2)$. In other words, the overall output can be written as

$$\begin{aligned} V_{out}(t) = {} & H_1(\omega_1)V_0 \exp(j\omega_1 t) + H_1(\omega_2)V_0 \exp(j\omega_2 t) \\ & + H_2(\omega_1, \omega_2)V_0^2 \exp[j(\omega_1 + \omega_2)t] + \cdots. \end{aligned} \qquad (2.225)$$

How do we determine the terms at $2\omega_1$, $2\omega_2$, and $\omega_1 - \omega_2$? If $H_2(\omega_1, \omega_2)\exp[j(\omega_1 + \omega_2)t]$ represents the component at $\omega_1 + \omega_2$, then $H_2(\omega_1, \omega_1)\exp[j(2\omega_1)t]$ must model

26. From another point of view, in $V_0 \exp(j\omega_1 t) = V_0 \cos \omega_1 t + jV_0 \sin \omega_1 t$, the first term generates its own response, as does the second term; the two responses remain distinguishable by virtue of the factor $j$.

that at $2\omega_1$. Similarly, $H_2(\omega_2, \omega_2)$ and $H_2(\omega_1, -\omega_2)$ serve as coefficients for $\exp[j(2\omega_2)t]$ and $\exp[j(\omega_1 - \omega_2)t]$, respectively. In other words, a more complete form of Eq. (2.225) reads

$$\begin{aligned} V_{out}(t) = &H_1(\omega_1)V_0\exp(j\omega_1 t) + H_1(\omega_2)V_0\exp(j\omega_2 t) + H_2(\omega_1, \omega_1)V_0^2\exp(2j\omega_1 t) \\ &+ H_2(\omega_2, \omega_2)V_0^2\exp(2j\omega_2 t) + H_2(\omega_1, \omega_2)V_0^2\exp[j(\omega_1 + \omega_2)t] \\ &+ H_2(\omega_1, -\omega_2)V_0^2\exp[j(\omega_1 - \omega_2)t] + \cdots . \end{aligned} \tag{2.226}$$

Thus, our task is simply to compute $H_2(\omega_1, \omega_2)$.

**Example 2.31**

Determine $H_2(\omega_1, \omega_2)$ for the circuit of Fig. 2.72.

**Solution:**

We apply the input $V_{in}(t) = V_0\exp(j\omega_1 t) + V_0\exp(j\omega_2 t)$ and assume the output is of the form $V_{out}(t) = H_1(\omega_1)V_0\exp(j\omega_1 t) + H_1(\omega_2)V_0\exp(j\omega_2 t) + H_2(\omega_1, \omega_2)V_0^2\exp[j(\omega_1 + \omega_2)t]$. We substitute for $V_{out}$ and $V_{in}$ in Eq. (2.219):

$$\begin{aligned} &R_1C_0[1 + \alpha H_1(\omega_1)V_0e^{j\omega_1 t} + \alpha H_1(\omega_2)V_0e^{j\omega_2 t} + \alpha H_2(\omega_1, \omega_2)V_0^2e^{j(\omega_1+\omega_2)t}] \\ &\quad\times [H_1(\omega_1)j\omega_1V_0e^{j\omega_1 t} + H_1(\omega_2)j\omega_2V_0e^{j\omega_2 t} + H_2(\omega_1, \omega_2)j(\omega_1 + \omega_2) \\ &\quad\times V_0^2e^{j(\omega_1+\omega_2)t}] + H_1(\omega_1)e^{j\omega_1 t} + H_1(\omega_2)e^{j\omega_2 t} + H_2(\omega_1, \omega_2)V_0^2e^{j(\omega_1+\omega_2)t} \\ &\quad= V_0e^{j\omega_1 t} + V_0e^{j\omega_2 t}. \end{aligned} \tag{2.227}$$

To obtain $H_2$, we only consider the terms containing $\omega_1 + \omega_2$:

$$\begin{aligned} &R_1C_0[\alpha H_1(\omega_1)H_1(\omega_2)j\omega_1V_0^2e^{j(\omega_1+\omega_2)t} + \alpha H_1(\omega_2)H_1(\omega_1)j\omega_2V_0^2e^{j(\omega_1+\omega_2)t} \\ &\quad + H_2(\omega_1, \omega_2)j(\omega_1 + \omega_2)V_0^2e^{j(\omega_1+\omega_2)t}] + H_2(\omega_1, \omega_2) \\ &\quad\times V_0^2e^{j(\omega_1+\omega_2)t} = 0 \end{aligned} \tag{2.228}$$

That is,

$$H_2(\omega_1, \omega_2) = -\frac{\alpha R_1C_0 j(\omega_1 + \omega_2)H_1(\omega_1)H_1(\omega_2)}{R_1C_0 j(\omega_1 + \omega_2) + 1}. \tag{2.229}$$

Noting that the denominator resembles that of (2.223) but with $\omega_1$ replaced by $\omega_1 + \omega_2$, we simplify $H_2(\omega_1, \omega_2)$ to

$$H_2(\omega_1, \omega_2) = -\alpha R_1C_0 j(\omega_1 + \omega_2)H_1(\omega_1)H_1(\omega_2)H_1(\omega_1 + \omega_2). \tag{2.230}$$

Why did we assume $V_{out}(t) = H_1(\omega_1)V_0\exp(j\omega_1 t) + H_1(\omega_2)V_0\exp(j\omega_2 t) + H_2V_0^2(\omega_1, \omega_2)\exp[j(\omega_1 + \omega_2)t]$ while we know that $V_{out}(t)$ also contains terms at $2\omega_1$, $2\omega_2$, and $\omega_1 - \omega_2$? This is because these other exponentials do not yield terms of the form $\exp[j(\omega_1 + \omega_2)t]$.



**Example 2.32**

If an input $V_0\exp(j\omega_1 t)$ is applied to the circuit of Fig. 2.72, determine the amplitude of the second harmonic at the output.

**Solution:**

As mentioned earlier, the component at $2\omega_1$ is obtained as $H_2(\omega_1, \omega_1)V_0^2\exp[j(\omega_1 + \omega_1)t]$. Thus, the amplitude is equal to

$$|A_{2\omega 1}| = |\alpha R_1C_0(2\omega_1)H_1^2(\omega_1)H_1(2\omega_1)|V_0^2 \tag{2.231}$$

$$= \frac{2|\alpha|R_1C_0\omega_1V_0^2}{(R_1^2C_0^2\omega_1^2 + 1)\sqrt{4R_1^2C_0^2\omega_1^2 + 1}}. \tag{2.232}$$

We observe that $A_{2\omega 1}$ falls to zero as $\omega_1$ approaches zero because $C_1$ draws little current, and also as $\omega_1$ goes to infinity because the second harmonic is suppressed by the low-pass nature of the circuit.

**Example 2.33**

If two tones of equal amplitude are applied to the circuit of Fig. 2.72, determine the ratio of the amplitudes of the components at $\omega_1 + \omega_2$ and $\omega_1 - \omega_2$. Recall that $H_1(\omega) = (R_1C_0 j\omega + 1)^{-1}$.

**Solution:**

From Eq. (2.230), the ratio is given by

$$\left|\frac{A_{\omega 1+\omega 2}}{A_{\omega 1-\omega 2}}\right| = \left|\frac{H_2(\omega_1, \omega_2)}{H_2(\omega_1, -\omega_2)}\right| \tag{2.233}$$

$$= \left|\frac{(\omega_1 + \omega_2)H_1(\omega_2)H_1(\omega_1 + \omega_2)}{(\omega_1 - \omega_2)H_1(-\omega_2)H_1(\omega_1 - \omega_2)}\right|. \tag{2.234}$$

Since $|H_1(\omega_2)| = |H_1(-\omega_2)|$, we have

$$\left|\frac{A_{\omega 1+\omega 2}}{A_{\omega 1-\omega 2}}\right| = \frac{(\omega_1 + \omega_2)\sqrt{R_1^2C_0^2(\omega_1 - \omega_2)^2 + 1}}{|\omega_1 - \omega_2|\sqrt{R_1^2C_0^2(\omega_1 + \omega_2)^2 + 1}}. \tag{2.235}$$

The foregoing examples point to a methodical approach that allows us to compute the second harmonic or second-order IM components with a moderate amount of algebra. But how about higher-order harmonics or IM products? We surmise that for $N$th-order terms, we must apply the input $V_{in}(t) = V_0\exp(j\omega_1 t) + \cdots + V_0\exp(j\omega_N t)$ and compute $H_n(\omega_1, \dots, \omega_n)$ as the coefficient of the $\exp[j(\omega_1 + \cdots + \omega_n)t]$ terms in the output. The output can therefore be expressed as

$$V_{out}(t) = \sum_{k=1}^{N} H_1(\omega_k)V_0 \exp(j\omega_k t) + \sum_{m=1}^{N}\sum_{k=1}^{N} H_2(\omega_m, \pm\omega_k)V_0^2 \exp[j(\omega_m \pm \omega_k)t]$$

$$+ \sum_{n=1}^{N}\sum_{m=1}^{N}\sum_{k=1}^{N} H_3(\omega_n, \pm\omega_m, \pm\omega_k)V_0^3 \exp[j(\omega_n \pm \omega_m \pm \omega_k)t] + \cdots. \quad (2.236)$$

The above representation of the output is called the Volterra series. As exemplified by (2.230), $H_m(\omega_1, \ldots, \omega_m)$ can be computed in terms of $H_1, \ldots, H_{m-1}$ with no need to solve nonlinear equations. We call $H_m$ the $m$-th "Volterra kernel."

**Example 2.34**

Determine the third Volterra kernel for the circuit of Fig. 2.72.

**Solution:**

We assume $V_{in}(t) = V_0 \exp(j\omega_1 t) + V_0 \exp(j\omega_2 t) + V_0 \exp(j\omega_3 t)$. Since the output contains many components, we introduce the short hands $H_{1(1)} = H_1(\omega_1)V_0 \exp(j\omega_1 t)$, $H_{1(2)} = H_1(\omega_2)V_0 \exp(j\omega_2 t)$, etc., $H_{2(1,2)} = H_2(\omega_1, \omega_2)V_0^2 \exp[j(\omega_1 + \omega_2)t]$, etc., and $H_{3(1,2,3)} = H_3(\omega_1, \omega_2, \omega_3)V_0^3 \exp[j(\omega_1 + \omega_2 + \omega_3)t]$. We express the output as

$$V_{out}(t) = H_{1(1)} + H_{1(2)} + H_{1(3)} + H_{2(1,2)} + H_{2(1,3)} + H_{2(2,3)} + H_{2(1,1)}$$

$$+ H_{2(2,2)} + H_{2(3,3)} + H_{3(1,2,3)} + \cdots. \quad (2.237)$$

We must substitute for $V_{out}$ and $V_{in}$ in Eq. (2.219) and group all of the terms that contain $\omega_1 + \omega_2 + \omega_3$. To obtain such terms in the product of $\alpha V_{out}$ and $dV_{out}/dt$, we note that $\alpha H_{2(1,2)} j\omega_3 H_{1(3)}$ and $\alpha H_{1(3)} j(\omega_1 + \omega_2)H_{2(1,2)}$ produce an exponential of the form $\exp[j(\omega_1 + \omega_2)t]\exp(j\omega_3)$. Similarly, $\alpha H_{2(2,3)} j\omega_1 H_{1(1)}$, $\alpha H_{1(1)} j(\omega_2 + \omega_3)H_{2(2,3)}$, $\alpha H_{2(1,3)} j\omega_2 H_{1(2)}$, and $\alpha H_{1(2)} j(\omega_1 + \omega_3)H_{2(1,3)}$ result in $\omega_1 + \omega_2 + \omega_3$. Finally, the product of $\alpha V_{out}$ and $dV_{out}/dt$ also contains $1 \times j(\omega_1 + \omega_2 + \omega_3)H_{3(1,2,3)}$. Grouping all of the terms, we have

$$H_3(\omega_1, \omega_2, \omega_3)$$

$$= -j\alpha R_1 C_0 \frac{H_2(\omega_1, \omega_2)\omega_3 H_1(\omega_3) + H_2(\omega_2, \omega_3)\omega_1 H_1(\omega_1) + H_2(\omega_1, \omega_3)\omega_2 H_1(\omega_2)}{R_1 C_0 j(\omega_1 + \omega_2 + \omega_3) + 1}$$

$$- j\alpha R_1 C_0 \frac{H_1(\omega_1)(\omega_2 + \omega_3)H_2(\omega_2, \omega_3) + H_1(\omega_2)(\omega_1 + \omega_3)H_2(\omega_1, \omega_3)}{R_1 C_0 j(\omega_1 + \omega_2 + \omega_3) + 1}$$

$$- j\alpha R_1 C_0 \frac{H_1(\omega_3)(\omega_1 + \omega_2)H_2(\omega_1, \omega_2)}{R_1 C_0 j(\omega_1 + \omega_2 + \omega_3) + 1}. \quad (2.238)$$

Note that $H_{2(1,1)}$, etc., do not appear here and could have been omitted from Eq. (2.237). With the third Volterra kernel available, we can compute the amplitude of critical terms. For example, the third-order IM components in a two-tone test are obtained by substituting $\omega_1$ for $\omega_3$ and $-\omega_2$ for $\omega_2$.



The reader may wonder if the Volterra series can be used with inputs other than exponentials. This is indeed possible [14] but beyond the scope of this book.

The approach described in this section is called the "harmonic" method of kernel calculation. In summary, this method proceeds as follows:

1. Assume $V_{in}(t) = V_0 \exp(j\omega_1 t)$ and $V_{out}(t) = H_1(\omega_1)V_0 \exp(j\omega_1 t)$. Substitute for $V_{out}$ and $V_{in}$ in the system's differential equation, group the terms that contain $\exp(j\omega_1 t)$, and compute the first (linear) kernel, $H_1(\omega_1)$.
2. Assume $V_{in}(t) = V_0 \exp(j\omega_1 t) + V_0 \exp(j\omega_2 t)$ and $V_{out}(t) = H_1(\omega_1)V_0 \exp(j\omega_1 t) + H_1(\omega_2)V_0 \exp(j\omega_2 t) + H_2(\omega_1, \omega_2)V_0^2 \exp[j(\omega_1 + \omega_2)t]$. Make substitutions in the differential equation, group the terms that contain $\exp[j(\omega_1 + \omega_2)t]$, and determine the second kernel, $H_2(\omega_1, \omega_2)$.
3. Assume $V_{in}(t) = V_0 \exp(j\omega_1 t) + V_0 \exp(j\omega_2 t) + V_0 \exp(j\omega_3 t)$ and $V_{out}(t)$ is given by Eq. (2.237). Make substitutions, group the terms that contain $\exp[j(\omega_1 + \omega_2 + \omega_3)t]$, and calculate the third kernel, $H_3(\omega_1, \omega_2, \omega_3)$.
4. To compute the amplitude of harmonics and IM components, choose $\omega_1, \omega_2, \ldots$ properly. For example, $H_2(\omega_1, \omega_1)$ yields the transfer function for $2\omega_1$ and $H_3(\omega_1, -\omega_2, \omega_1)$ the transfer function for $2\omega_1 - \omega_2$.

### 2.8.1 Method of Nonlinear Currents

As seen in Example 2.34, the harmonic method becomes rapidly more complex as $n$ increases. An alternative approach called the method of "nonlinear currents" is sometimes preferred as it reduces the algebra to some extent. We describe the method itself here and refer the reader to [13] for a formal proof of its validity.

The method of nonlinear currents proceeds as follows for a circuit that contains a two-terminal nonlinear device [13]:

1. Assume $V_{in}(t) = V_0 \exp(j\omega_1 t)$ and determine the linear response of the circuit by ignoring the nonlinearity. The "response" includes both the output of interest *and* the voltage across the nonlinear device.
2. Assume $V_{in}(t) = V_0 \exp(j\omega_1 t) + V_0 \exp(j\omega_2 t)$ and calculate the voltage across the nonlinear device, assuming it is linear. Now, compute the *nonlinear* component of the current flowing through the device, assuming the device is nonlinear.
3. Set the main input to *zero* and place a current source equal to the nonlinear component found in Step 2 in parallel with the nonlinear device.
4. Ignoring the nonlinearity of the device again, determine the circuit's response to the current source applied in Step 3. Again, the response includes the output of interest and the voltage across the nonlinear device.
5. Repeat Steps 2, 3, and 4 for higher-order responses. The overall response is equal to the output components found in Steps 1, 4, etc.

The following example illustrates the procedure.

### Example 2.35

Determine $H_3(\omega_1, \omega_2, \omega_3)$ for the circuit of Fig. 2.72.

**Solution:**

In this case, the output voltage also appears across the nonlinear device. We know that $H_1(\omega_1) = (R_1 C_0 j\omega_1 + 1)^{-1}$. Thus, with $V_{in}(t) = V_0 \exp(j\omega_1 t)$, the voltage across the capacitor is equal to

$$V_{C1}(t) = \frac{V_0}{R_1 C_0 j\omega_1 + 1} e^{j\omega_1 t}. \tag{2.239}$$

In the second step, we apply $V_{in}(t) = V_0 \exp(j\omega_1 t) + V_0 \exp(j\omega_2 t)$, obtaining the linear voltage across $C_1$ as

$$V_{C1}(t) = \frac{V_0 e^{j\omega_1 t}}{R_1 C_0 j\omega_1 + 1} + \frac{V_0 e^{j\omega_2 t}}{R_1 C_0 j\omega_2 + 1}. \tag{2.240}$$

With this voltage, we compute the nonlinear current flowing through $C_1$:

$$I_{C1,non}(t) = \alpha C_0 V_{C1} \frac{dV_{C1}}{dt} \tag{2.241}$$

$$= \alpha C_0 \left( \frac{V_0 e^{j\omega_1 t}}{R_1 C_0 j\omega_1 + 1} + \frac{V_0 e^{j\omega_2 t}}{R_1 C_0 j\omega_2 + 1} \right) \times \left( \frac{j\omega_1 V_0 e^{j\omega_1 t}}{R_1 C_0 j\omega_1 + 1} + \frac{j\omega_2 V_0 e^{j\omega_2 t}}{R_1 C_0 j\omega_2 + 1} \right). \tag{2.242}$$

Since only the component at $\omega_1 + \omega_2$ is of interest at this point, we rewrite the above expression as

$$I_{C1,non}(t) = \alpha C_0 \left[ \frac{j(\omega_1 + \omega_2) V_0^2 e^{j(\omega_1+\omega_2)t}}{(R_1 C_0 j\omega_1 + 1)(R_1 C_0 j\omega_2 + 1)} + \cdots \right] \tag{2.243}$$

$$= \alpha C_0 [ j(\omega_1 + \omega_2) V_0^2 e^{j(\omega_1+\omega_2)t} H_1(\omega_1) H_1(\omega_2) + \cdots ]. \tag{2.244}$$

In the third step, we set the input to zero, assume a linear capacitor, and apply $I_{C1,non}(t)$ in parallel with $C_1$ (Fig. 2.73). The current component at $\omega_1 + \omega_2$ flows through the parallel combination of $R_1$ and $C_0$, producing $V_{C1,non}(t)$:

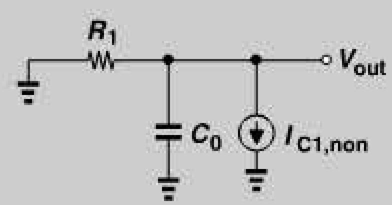



**Figure 2.73** *Inclusion of nonlinear current in RC section.*



### Example 2.35 (*Continued*)

$$V_{C1,non}(t) = -\alpha C_0 j(\omega_1 + \omega_2) V_0^2 e^{j(\omega_1+\omega_2)t} H_1(\omega_1) \times H_1(\omega_2) \frac{R_1}{R_1 C_0 j(\omega_1 + \omega_2) + 1} \tag{2.245}$$

$$= -\alpha R_1 C_0 j(\omega_1 + \omega_2) H_1(\omega_1) H_1(\omega_2) H_1(\omega_1 + \omega_2) V_0^2 e^{j(\omega_1+\omega_2)t}. \tag{2.246}$$

We note that the coefficient of $V_0^2 \exp[j(\omega_1 + \omega_2)t]$ in these two equations is the same as $H_2(\omega_1, \omega_2)$ in (2.229).

To determine $H_3(\omega_1, \omega_2, \omega_3)$, we must assume an input of the form $V_{in}(t) = V_0 \exp(j\omega_1 t) + V_0 \exp(j\omega_2 t) + V_0 \exp(j\omega_3 t)$ and write the voltage across $C_1$ as

$$V_{C1}(t) = H_1(\omega_1) V_0 e^{j\omega_1 t} + H_1(\omega_2) V_0 e^{j\omega_2 t} + H_1(\omega_3) V_0 e^{j\omega_3 t} + H_2(\omega_1, \omega_2) V_0^2 e^{j(\omega_1+\omega_2)t} + H_2(\omega_1, \omega_3) V_0^2 e^{j(\omega_1+\omega_3)t} + H_2(\omega_2, \omega_3) V_0^2 e^{j(\omega_2+\omega_3)t}. \tag{2.247}$$

Note that, in contrast to Eq. (2.240), we have included the second-order nonlinear terms in the voltage so as to calculate the third-order terms.[27] The nonlinear current through $C_1$ is thus equal to

$$I_{C1,non}(t) = \alpha C_0 V_{C1} \frac{dV_{C1}}{dt}. \tag{2.248}$$

We substitute for $V_{C1}$ and group the terms containing $\omega_1 + \omega_2 + \omega_3$:

$$\begin{aligned} I_{C1,non}(t) = \alpha C_0 [ & H_1(\omega_1) H_2(\omega_2, \omega_3) j(\omega_2 + \omega_3) + H_2(\omega_2, \omega_3) j\omega_1 H_1(\omega_1) \\ & + H_1(\omega_2) H_2(\omega_1, \omega_3) j(\omega_1 + \omega_3) + H_2(\omega_1, \omega_3) j\omega_2 H_1(\omega_2) \\ & + H_1(\omega_3) H_2(\omega_1, \omega_2) j(\omega_1 + \omega_2) + H_2(\omega_1, \omega_2) j\omega_3 H_1(\omega_3) ] V_0^3 e^{j(\omega_1+\omega_2+\omega_3)t} \\ & + \cdots. \end{aligned} \tag{2.249}$$

This current flows through the parallel combination of $R_1$ and $C_0$, yielding $V_{C1,non}(t)$. The reader can readily show that the coefficient of $\exp[j(\omega_1 + \omega_2 + \omega_3)t]$ in $V_{C1,non}(t)$ is the same as the third kernel expressed by Eq. (2.238).

The procedure described above applies to two-terminal nonlinear devices. For transistors, a similar approach can be taken. We illustrate this point with the aid of an example.

### Example 2.36

Figure 2.74(a) shows the input network of a commonly-used LNA (Chapter 5). Assuming that $g_m L_1 / C_{GS} = R_S$ (Chapter 5) and $I_D = \alpha(V_{GS} - V_{TH})^2$, determine the nonlinear terms in $I_{out}$. Neglect other capacitances, channel-length modulation, and body effect.

(*Continues*)

27. Other terms are excluded because they do not lead to a component at $\omega_1 + \omega_2 + \omega_3$.

**Example 2.36** (*Continued*)

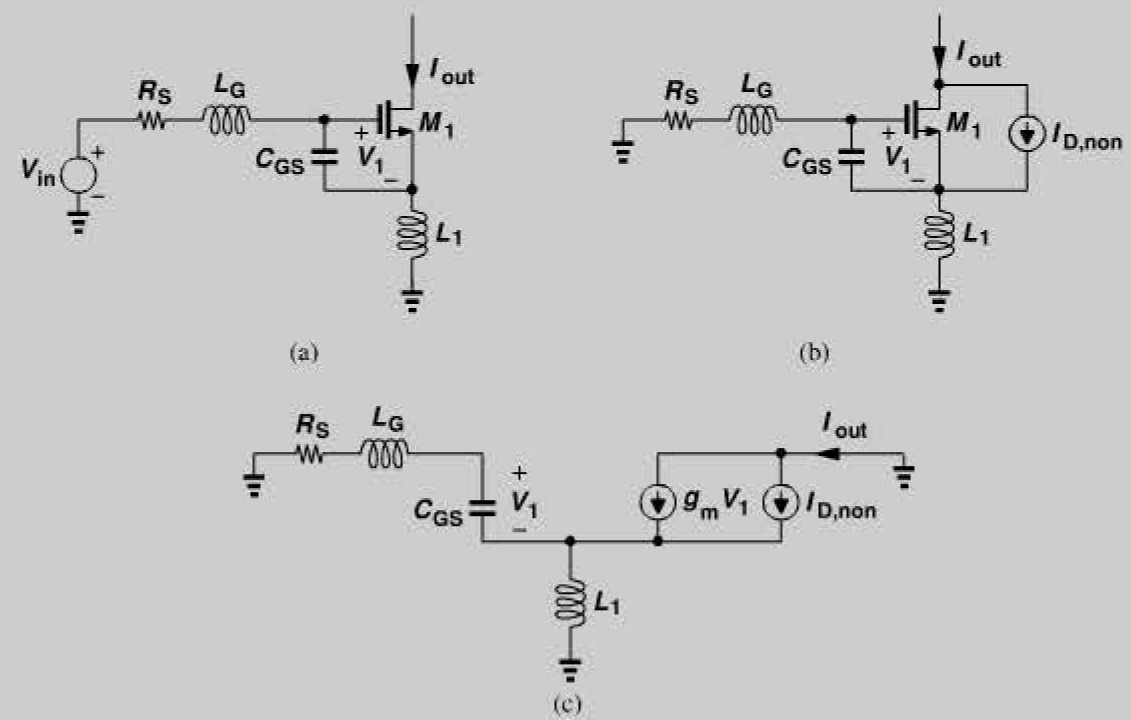



**Figure 2.74** *(a) CS stage with inductors in series with source and gate, (b) inclusion of nonlinear current, (c) computation of output current.*

### Solution:

In this circuit, two quantities are of interest, namely, the output current, $I_{out}$ ($= I_D$), and the gate-source voltage, $V_1$; the latter must be computed each time as it determines the nonlinear component in $I_D$.

Let us begin with the linear response. Since the current flowing through $L_1$ is equal to $V_1C_{GS}s + g_mV_1$ and that flowing through $R_S$ and $L_G$ equal to $V_1C_{GS}s$, we can write a KVL around the input loop as

$$V_{in} = (R_S + L_Gs)V_1C_{GS}s + V_1 + (V_1C_{GS}s + g_mV_1)L_1s. \tag{2.250}$$

It follows that

$$\frac{V_1}{V_{in}} = \frac{1}{(L_1 + L_G)C_{GS}s^2 + (R_SC_{GS} + g_mL_1)s + 1}. \tag{2.251}$$

Since we have assumed $g_mL_1/C_{GS} = R_S$, for $s = j\omega$ we obtain

$$\frac{V_1}{V_{in}}(j\omega) = \frac{1}{2g_mL_1j\omega + 1 - \dfrac{\omega^2}{\omega_0^2}} = H_1(\omega), \tag{2.252}$$

where $\omega_0^2 = [(L_1 + L_G)C_{GS}]^{-1}$. Note that $I_{out} = g_mV_1 = g_mH_1(\omega)V_{in}$.



**Example 2.36** (*Continued*)

Now, we assume $V_{in}(t) = V_0\exp(j\omega_1t) + V_0\exp(j\omega_2t)$ and write

$$V_1(t) = H_1(\omega_1)V_0e^{j\omega_1t} + H_1(\omega_2)V_0e^{j\omega_2t}. \tag{2.253}$$

Upon experiencing the characteristic $I_D = \alpha V_1^2$, this voltage results in a nonlinear current given by

$$I_{D,non} = 2\alpha H_1(\omega_1)H_1(\omega_2)V_0^2e^{j(\omega_1+\omega_2)t}. \tag{2.254}$$

In the next step, we set $V_{in}$ to zero and insert a current source having the above value in parallel with the drain current source [Fig. 2.74(b)]. We must compute $V_1$ in response to $I_{D,non}$, assuming the circuit is *linear*. From the equivalent circuit shown in Fig. 2.74(c), we have the following KVL:

$$(R_S + L_Gs)V_1C_{GS}s + V_1 + (g_mV_1 + I_{D,non} + V_1C_{GS}s)L_1s = 0. \tag{2.255}$$

Thus, for $s = j\omega$

$$\frac{V_1}{I_{D,non}}(j\omega) = \frac{-jL_1\omega}{2g_mL_1j\omega + 1 - \dfrac{\omega^2}{\omega_0^2}}. \tag{2.256}$$

Since $I_{D,non}$ contains a frequency component at $\omega_1 + \omega_2$, the above transfer function must be calculated at $\omega_1 + \omega_2$ and multiplied by $I_{D,non}$ to yield $V_1$. We therefore have

$$H_2(\omega_1, \omega_2) = \frac{-jL_1(\omega_1 + \omega_2)}{2g_mL_1j(\omega_1 + \omega_2) + 1 - \dfrac{(\omega_1 + \omega_2)^2}{\omega_0^2}}2\alpha H_1(\omega_1)H_1(\omega_2). \tag{2.257}$$

In our last step, we assume $V_{in}(t) = V_0\exp(j\omega_1t) + V_0\exp(j\omega_2t) + V_0\exp(j\omega_3t)$ and write

$$\begin{aligned}V_1(t) = {} & H_1(\omega_1)V_0e^{j\omega_1t} + H_1(\omega_2)V_0e^{j\omega_2t} + H_1(\omega_3)V_0e^{j\omega_3t} + H_2(\omega_1, \omega_2)V_0^2e^{j(\omega_1+\omega_2)t} \\ & + H_2(\omega_1, \omega_3)V_0^2e^{j(\omega_1+\omega_3)t} + H_2(\omega_2, \omega_3)V_0^2e^{j(\omega_2+\omega_3)t}.\end{aligned} \tag{2.258}$$

Since $I_D = \alpha V_1^2$, the nonlinear current at $\omega_1 + \omega_2 + \omega_3$ is expressed as

$$\begin{aligned}I_{D,non} = {} & 2\alpha[H_1(\omega_1)H_2(\omega_2, \omega_3) + H_1(\omega_2)H_2(\omega_1, \omega_3) \\ & + H_1(\omega_3)H_2(\omega_1, \omega_2)]V_0^3e^{j(\omega_1+\omega_2+\omega_3)t}.\end{aligned} \tag{2.259}$$

The third-order nonlinear component in the output of interest, $I_{out}$, is equal to the above expression. We note that, even though the transistor exhibits only second-order nonlinearity, the degeneration (feedback) caused by $L_1$ results in higher-order terms.

The reader is encouraged to repeat this analysis using the harmonic method and see that it is much more complex.

## REFERENCES


[1] B. Razavi, *Design of Analog CMOS Integrated Circuits,* Boston: McGraw-Hill, 2001.
[2] L. W. Couch, *Digital and Analog Communication Systems,* Fourth Edition, New York: Macmillan Co., 1993.
[3] A. van der Ziel, "Thermal Noise in Field Effect Transistors," *Proc. IRE*, vol. 50, pp. 1808–1812, Aug. 1962.
[4] A. A. Abidi, "High-Frequency Noise Measurements on FETs with Small Dimensions," *IEEE Trans. Electron Devices,* vol. 33, pp. 1801–1805, Nov. 1986.
[5] A. J. Sholten et al., "Accurate Thermal Noise Model of Deep-Submicron CMOS," *IEDM Dig. Tech. Papers,* pp. 155–158, Dec. 1999.
[6] B. Razavi, "Impact of Distributed Gate Resistance on the Performance of MOS Devices," *IEEE Trans. Circuits and Systems- Part I,* vol. 41, pp. 750–754, Nov. 1994.
[7] H. T. Friis, "Noise Figure of Radio Receivers," *Proc. IRE*, vol. 32, pp. 419–422, July 1944.
[8] A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* Third Edition, New York: McGraw-Hill, 1991.
[9] R. W. Bennet, "Methods of Solving Noise Problems," *Proc. IRE*, vol. 44, pp. 609–638, May 1956.
[10] S. Narayanan, "Application of Volterra Series to Intermodulation Distortion Analysis of Transistor Feedback Amplifiers," *IEEE Tran. Circuit Theory,* vol. 17, pp. 518–527, Nov. 1970.
[11] P. Wambacq et al., "High-Frequency Distortion Analysis of Analog Integrated Circuits," *IEEE Tran. Circuits and Systems, II,* vol. 46, pp. 335–334, March 1999.
[12] P. Wambaq and W. Sansen, *Distortion Analysis of Analog Integrated Circuits,* Norwell, MA: Kluwer, 1998.
[13] J. Bussganag, L. Ehrman, and J. W. Graham, "Analysis of Nonlinear Systems with Multiple Inputs," *Proc. IEEE*, vol. 62, pp. 1088–1119, Aug. 1974.
[14] E. Bedrosian and S. O. Rice, "The Output Properties of Volterra Systems (Nonlinear Systems with Memory) Driven by Harmonic and Gaussian Inputs," *Proc. IEEE,* vol. 59, pp. 1688–1707, Dec. 1971.


## PROBLEMS

2.1. Two nonlinear stages are cascaded. If the input/output characteristic of each stage is approximated by a third-order polynomial, determine the $P_{1dB}$ of the cascade in terms of the $P_{1dB}$ of each stage.

2.2. Repeat Example 2.11 if one interferer has a level of $-3$ dBm and the other, $-35$ dBm.

2.3. If cascaded, stages having only *second-order* nonlinearity can yield a finite $IP_3$. For example, consider the cascade identical common-source stages shown in Fig. 2.75.

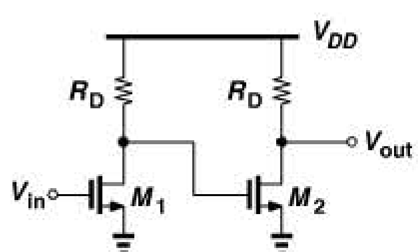



**Figure 2.75** *Cascade of CS stages.*



If each transistor operates in saturation and follows the ideal square-law behavior, determine the $IP_3$ of the cascade.

2.4. Determine the $IP_3$ and $P_{1dB}$ for a system whose characteristic is approximated by a fifth-order polynomial.

2.5. Consider the scenario shown in Fig. 2.76, where $\omega_3 - \omega_2 = \omega_2 - \omega_3$ and the band-pass filter provides an attenuation of 17 dB at $\omega_2$ and 37 dB at $\omega_3$.

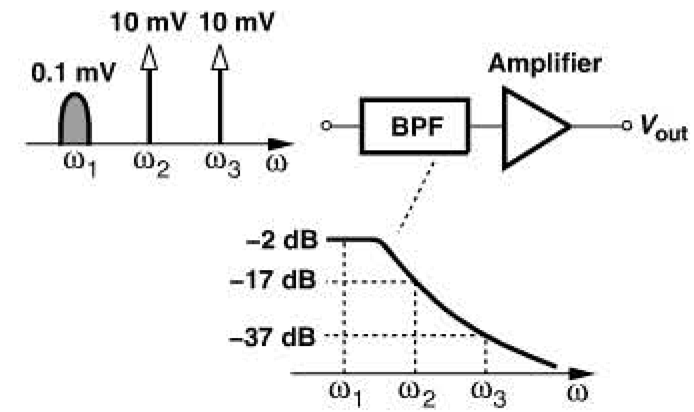



**Figure 2.76** *Cascade of BPF and amplifier.*

(a) Compute the $IIP_3$ of the amplifier such that the intermodulation product falling at $\omega_1$ is 20 dB below the desired signal.

(b) Suppose an amplifier with a voltage gain of 10 dB and $IIP_3 = 500\,\text{mV}_p$ precedes the band-pass filter. Calculate the $IIP_3$ of the overall chain. (Neglect second-order nonlinearities.)

2.6. Prove that the Fourier transform of the autocorrelation of a random signal yields the spectrum, i.e., the power measured in a 1-Hz bandwidth at each frequency.

2.7. A broadband circuit sensing an input $V_0 \cos \omega_0 t$ produces a third harmonic $V_3 \cos(3\omega_0 t)$. Determine the 1-dB compression point in terms of $V_0$ and $V_3$.

2.8. Prove that in Fig. 2.36, the noise power delivered by $R_1$ to $R_2$ is equal to that delivered by $R_2$ to $R_1$ if the resistors reside at the same temperature. What happens if they do not?

2.9. Explain why the channel thermal noise of a MOSFET is modeled by a current source tied between the source and drain terminals (rather than, say, between the gate and source terminals).

2.10. Prove that the channel thermal noise of a MOSFET can be referred to the gate as a voltage given by $4kT\gamma/g_m$. As shown in Fig. 2.77, the two circuits must generate the same current with the same terminal voltages.

2.11. Determine the NF of the circuit shown in Fig. 2.52 using Friis' equation.

2.12. Prove that the output noise voltage of the circuit shown in Fig. 2.46(c) is given by $\overline{V_{n2}^2} = \overline{I_{n1}^2} r_O^2$.

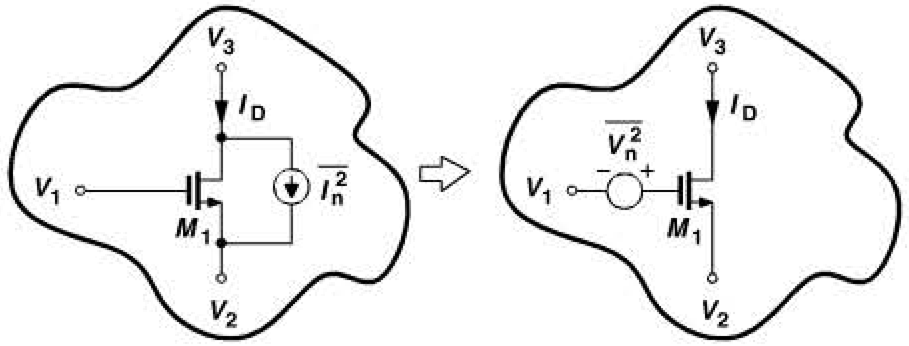



**Figure 2.77** *Equivalent circuits for noise of a MOSFET.*

2.13. Repeat Example 2.23 if the CS and CG stages are swapped. Does the NF change? Why?

2.14. Repeat Example 2.23 if $R_{D1}$ and $R_{D2}$ are replaced with ideal current sources and channel-length modulation is not neglected.

2.15. The input/output characteristic of a bipolar differential pair is given by $V_{out} = -2R_C I_{EE} \tanh[V_{in}/(2V_T)]$, where $R_C$ denotes the load resistance, $I_{EE}$ is the tail current, and $V_T = kT/q$. Determine the $IP_3$ of the circuit.

2.16. What happens to the noise figure of a circuit if the circuit is loaded by a noiseless impedance $Z_L$ at its output?

2.17. The noise figure of a circuit is known for a source impedance of $R_{S1}$. Is it possible to compute the noise figure for another source impedance $R_{S2}$? Explain in detail.

2.18. Equation (2.122) implies that the noise figure falls as $R_S$ rises. Assuming that the antenna voltage swing remains constant, explain what happens to the output SNR as $R_S$ increases.

2.19. Repeat Example 2.21 for the arrangement shown in Fig. 2.78, where the transformer amplifies its primary voltage by a factor of $n$ and transforms $R_S$ to a value of $n^2 R_S$.

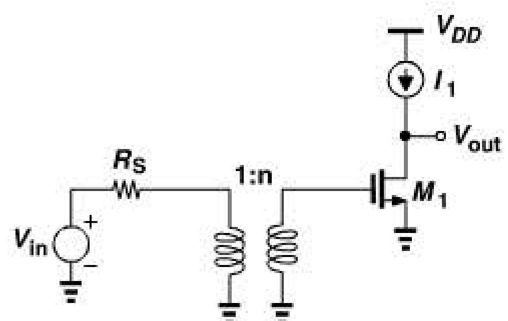



**Figure 2.78** *CS stage driven by a transformer.*

2.20. For matched inputs and outputs, prove that the NF of a passive (reciprocal) circuit is equal to its power loss.

2.21. Determine the noise figure of each circuit in Fig. 2.79 with respect to a source impedance $R_S$. Neglect channel-length modulation and body effect.



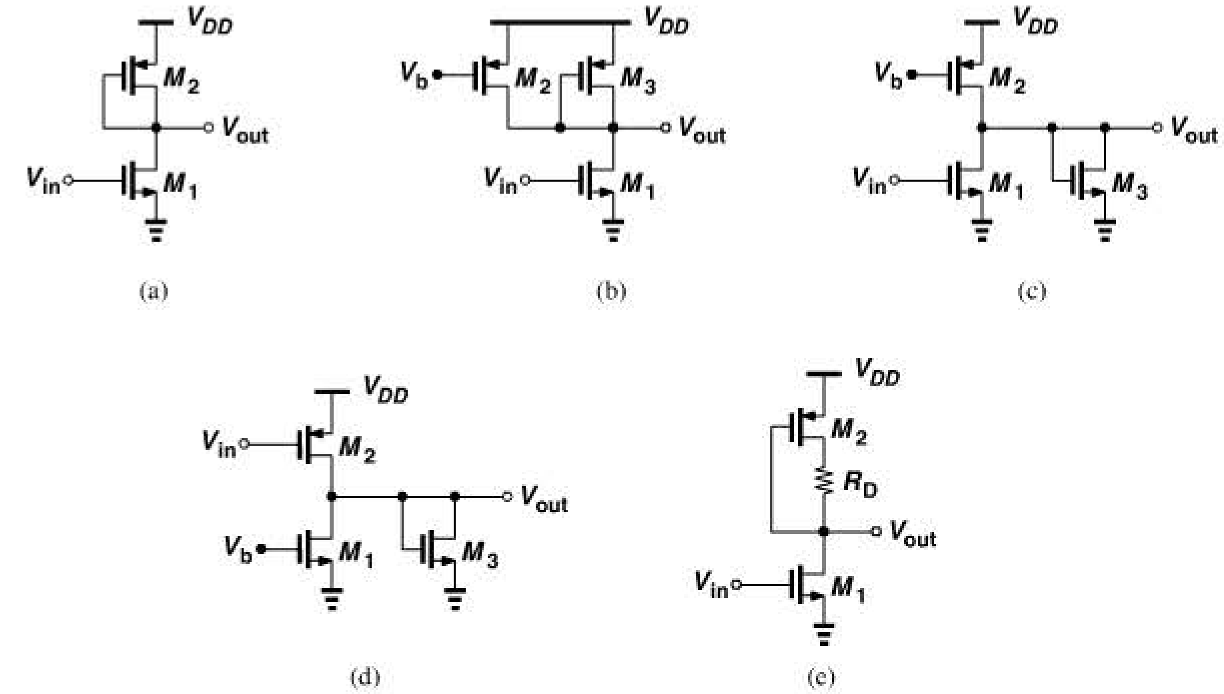



**Figure 2.79** *CS stages for NF calculation.*

2.22. Determine the noise figure of each circuit in Fig. 2.80 with respect to a source impedance $R_S$. Neglect channel-length modulation and body effect.

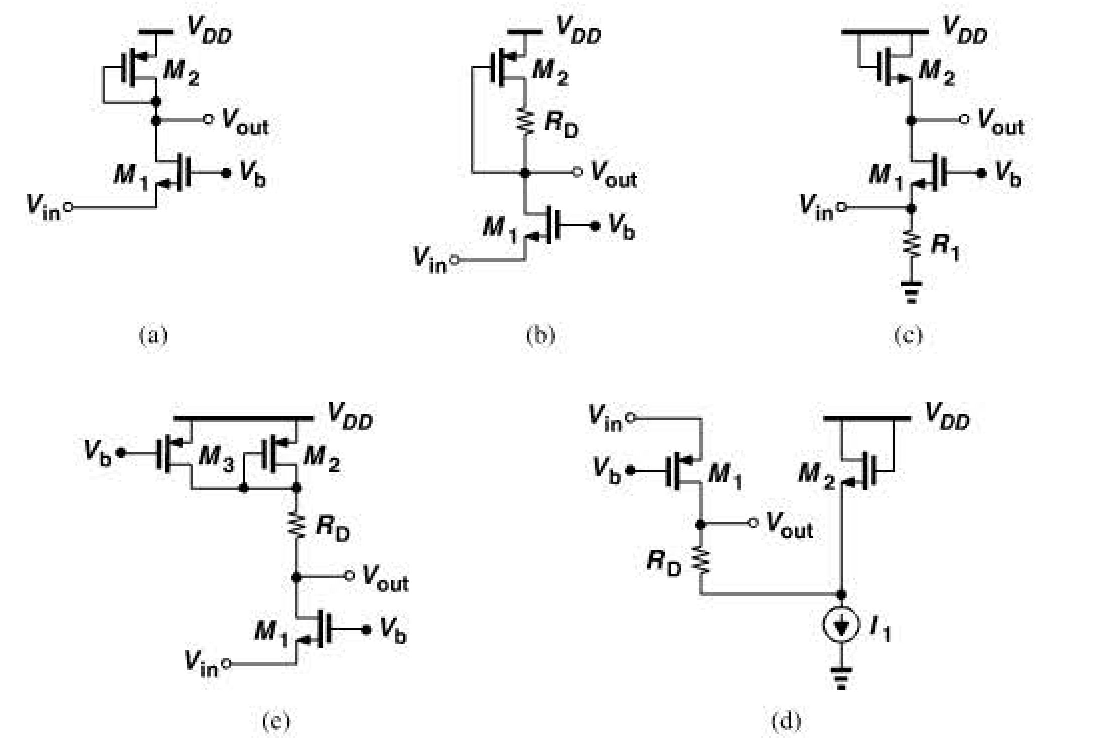



**Figure 2.80** *CG stages for NF calculation.*

2.23. Determine the noise figure of each circuit in Fig. 2.81 with respect to a source impedance $R_S$. Neglect channel-length modulation and body effect.

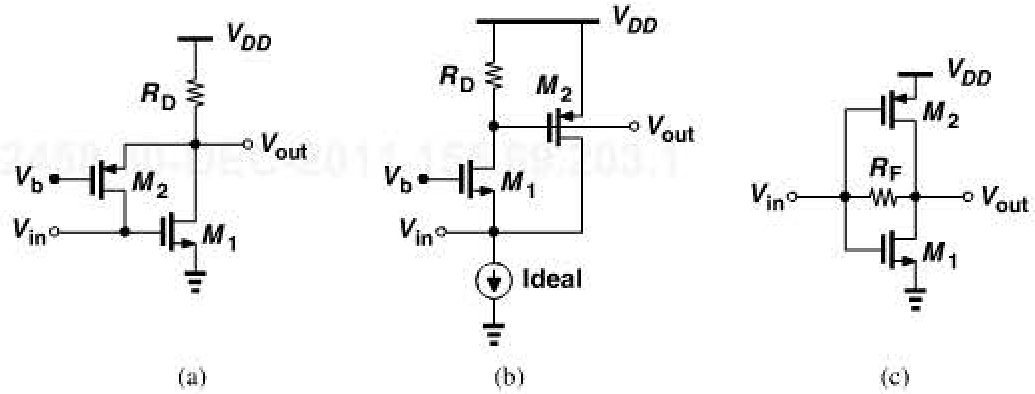



**Figure 2.81** *Stages for NF calculation.*

CHAPTER

3

# COMMUNICATION CONCEPTS

The design of highly-integrated RF transceivers requires a solid understanding of communication theory. For example, as mentioned in Chapter 2, the receiver sensitivity depends on the minimum acceptable signal-to-noise ratio, which itself depends on the type of modulation. In fact, today we rarely design a low-noise amplifier, an oscillator, etc., with no attention to the type of transceiver in which they are used. Furthermore, modern RF designers must regularly interact with digital signal processing engineers to trade functions and specifications and must therefore speak the same language.

This chapter provides a basic, yet necessary, understanding of modulation theory and wireless standards. Tailored to a reader who is ultimately interested in RF IC design rather than communication theory, the concepts are described in an intuitive language so that they can be incorporated in the reader's daily work. The outline of the chapter is shown below.

| Modulation | Mobile Systems | Multiple Access Technqiues | Wireless Standards |
|---|---|---|---|
| ▪ AM, PM, FM | ▪ Cellular System | ▪ Duplexing | ▪ GSM |
| ▪ Intersymbol Interference | ▪ Hand-off | ▪ FDMA | ▪ IS-95 CDMA |
| ▪ Signal Constellations | ▪ Multipath Fading | ▪ TDMA | ▪ Wideband CDMA |
| ▪ ASK, PSK, FSK | ▪ Diversity | ▪ CDMA | ▪ Bluetooth |
| ▪ QPSK, GMSK, QAM | | | ▪ IEEE802.11a/b/g |
| ▪ OFDM | | | |
| ▪ Spectral Regrowth | | | |

## 3.1 GENERAL CONSIDERATIONS

How does your voice enter a cell phone here and come out of another cell phone miles away? We wish to understand the incredible journey that your voice signal takes.

The transmitter in a cell phone must convert the voice, which is called a "baseband signal" because its spectrum (20 Hz to 20 kHz) is centered around zero frequency, to a "passband signal," i.e., one residing around a nonzero center frequency, $\omega_c$ [Fig. 3.1(b)]. We call $\omega_c$ the "carrier frequency."