2

Issues in RFIC Design, Noise, Linearity, and Filtering

2.1 Introduction

In this chapter we will have a brief look at some general issues in RF circuit design. Nonidealities we will consider include noise and nonlinearity. We will also consider the effect of filtering. An ideal circuit, such as an amplifier, produces a perfect copy of the input signal at the output. In a real circuit, the amplifier will introduce both noise and distortion to that waveform. Noise, which is present in all resistors and active devices, limits the minimum detectable signal in a radio. At the other amplitude extreme, nonlinearities in the circuit blocks will cause the output signal to become distorted, limiting the maximum signal amplitude.

At the system level, specifications for linearity and noise as well as many other parameters must be determined before the circuit can be designed. In this chapter, before we look at circuit details, we will look at some of these system issues in more detail. In order to design radio frequency integrated circuits with realistic specifications, we need to understand the impact of noise on minimum detectable signals and the effect of nonlinearity on distortion. Knowledge of noise floors and distortion will be used to understand the requirements for circuit parameters.

2.2 Noise

Signal detection is more difficult in the presence of noise. In addition to the desired signal, the receiver is also picking up noise from the rest of the universe.

Any matter above 0K contains thermal energy. This thermal energy moves atoms and electrons around in a random way, leading to random currents in circuits, which are also noise. Noise can also come from man-made sources such as microwave ovens, cell phones, pagers, and radio antennas. Circuit designers are mostly concerned with how much noise is being added by the circuits in the transceiver. At the input to the receiver, there will be some noise power present that defines the noise floor. The minimum detectable signal must be higher than the noise floor by some *signal-to-noise ratio* (SNR) to detect signals reliably and to compensate for additional noise added by circuitry. These concepts will be described in the following sections.

We note that to find the total noise due to a number of sources, the relationship of the sources with each other has to be considered. The most common assumption is that all noise sources are random and have no relationship with each other, so they are said to be uncorrelated. In such a case, noise power is added instead of noise voltage. Similarly, if noise at different frequencies is uncorrelated, noise power is added. We note that signals, like noise, can also be uncorrelated, such as signals at different unrelated frequencies. In such a case, one finds the total output signal by adding the powers. On the other hand, if two sources are correlated, the voltages can be added. As an example, correlated noise is seen at the outputs of two separate paths that have the same origin.

2.2.1 Thermal Noise

One of the most common noise sources in a circuit is a resistor. Noise in resistors is generated by thermal energy causing random electron motion [1–3]. The thermal noise spectral density in a resistor is given by

$$N_{\text{resistor}} = 4kTR \tag{2.1}$$

where T is the Kelvin temperature of the resistor, k is Boltzmann's constant (1.38×10^{-23} J/K), and R is the value of the resistor. Noise power spectral density is expressed using volts squared per hertz (power spectral density). In order to find out how much power a resistor produces in a finite bandwidth, simply multiply (2.1) by the bandwidth of interest Δf:

$$v_n^2 = 4kTR\Delta f \tag{2.2}$$

where v_n is the rms value of the noise voltage in the bandwidth Δf. This can also be written equivalently as a noise current rather than a noise voltage:

$$i_n^2 = \frac{4kT\Delta f}{R} \tag{2.3}$$

Thermal noise is white noise, meaning it has a constant power spectral density with respect to frequency (valid up to approximately 6,000 GHz) [4]. The model for noise in a resistor is shown in Figure 2.1.

2.2.2 Available Noise Power

Maximum power is transferred to the load when R_{LOAD} is equal to R. Then v_o is equal to $v_n/2$. The output power spectral density P_o is then given by

$$P_o = \frac{v_o^2}{R} = \frac{v_n^2}{4R} = kT \tag{2.4}$$

Thus, available power is kT, independent of resistor size. Note that kT is in watts per hertz, which is a power density. To get total power out P_{out} in watts, multiply by the bandwidth, with the result that

$$P_{\mathrm{out}} = kTB \tag{2.5}$$

2.2.3 Available Power from Antenna

The noise from an antenna can be modeled as a resistor [5]. Thus, as in the previous section, the available power from an antenna is given by

$$P_{\mathrm{available}} = kT = 4 \times 10^{-21} \text{ W/Hz} \tag{2.6}$$

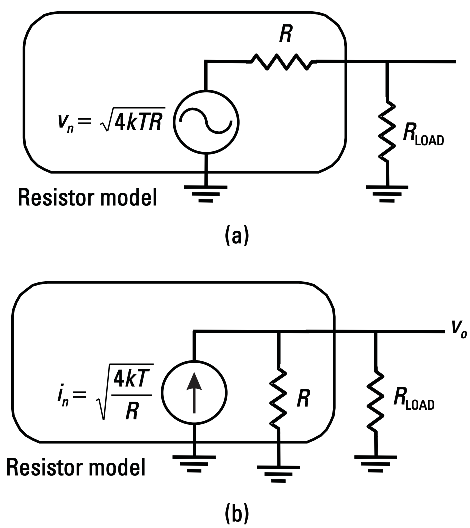

Figure 2.1 Resistor noise model: (a) with a voltage source, and (b) with a current source.

at $T = 290$K, or in dBm per hertz,

$$P_{\text{available}} = 10\log_{10}\left(\frac{4 \times 10^{-21}}{1 \times 10^{-3}}\right) = -174 \text{ dBm/Hz} \tag{2.7}$$

Note that using 290K as the temperature of the resistor modeling the antenna is appropriate for cell phone applications where the antenna is pointed at the horizon. However, if the antenna were pointed at the sky, the equivalent noise temperature would be much lower, more typically 50K [6].

For any receiver required to receive a given signal bandwidth, the minimum detectable signal can now be determined. As can be seen from (2.5), the noise floor depends on the bandwidth. For example, with a bandwidth of 200 kHz, the noise floor is

$$\text{Noise floor} = kTB = 4 \times 10^{-21} \times 200{,}000 = 8 \times 10^{-16} \tag{2.8}$$

More commonly, the noise floor would be expressed in dBm, as in the following for the example shown above:

$$\text{Noise floor} = -174 \text{ dBm/Hz} + 10\log_{10}(200{,}000) = -121 \text{ dBm} \tag{2.9}$$

Thus, we can now also formally define signal-to-noise ratio. If the signal has a power of *S,* then the SNR is

$$\text{SNR} = \frac{S}{\text{Noise floor}} \tag{2.10}$$

Thus, if the electronics added no noise and if the detector required a signal-to-noise ratio of 0 dB, then a signal at −121 dBm could just be detected. The minimum detectable signal in a receiver is also referred to as the receiver sensitivity. However, the SNR required to detect bits reliably (e.g., bit error rate (BER) = 10^{-3}) is typically not 0 dB. The actual required SNR depends on a variety of factors, such as bit rate, energy per bit, IF filter bandwidth, detection method (e.g., synchronous or not), and interference levels. Such calculations are the topics for a digital communications course [6, 7] and will not be discussed further here. But typical results for a bit error rate of 10^{-3} is about 7 dB for *quadrature phase shift keying* (QPSK), about 12 dB for 16 *quadrature amplitude modulation* (QAM), and about 17 dB for 64 QAM, though often higher numbers are quoted to leave a safety margin. It should be noted that for data transmission, lower BER is often required (e.g., 10^{-6}), resulting in an SNR requirement of 11 dB or more for QPSK. Thus, the input signal level must be above the noise floor level by at least this amount. Consequently, the minimum detectable signal level in a 200-kHz bandwidth is more like −114 dBm (assuming no noise is added by the electronics).

2.2.4 The Concept of Noise Figure

Noise added by electronics will be directly added to the noise from the input. Thus, for reliable detection, the previously calculated minimum detectable signal level must be modified to include the noise from the active circuitry. Noise from the electronics is described by noise factor F, which is a measure of how much the signal-to-noise ratio is degraded through the system. We note that

$$S_o = G \cdot S_i \tag{2.11}$$

where S_i is the input signal power, S_o is the output signal power, and G is the power gain S_o/S_i. We derive the following equation for noise factor:

$$F = \frac{\mathrm{SNR}_i}{\mathrm{SNR}_o} = \frac{S_i/N_{i\,(\mathrm{source})}}{S_o/N_{o\,(\mathrm{total})}} = \frac{S_i/N_{i\,(\mathrm{source})}}{(S_i \cdot G)/N_{o\,(\mathrm{total})}} = \frac{N_{o\,(\mathrm{total})}}{G \cdot N_{i\,(\mathrm{source})}} \tag{2.12}$$

where $N_{o\,(\mathrm{total})}$ is the total noise at the output. If $N_{o\,(\mathrm{source})}$ is the noise at the output originating at the source, and $N_{o\,(\mathrm{added})}$ is the noise at the output added by the electronic circuitry, then we can write:

$$N_{o\,(\mathrm{total})} = N_{o\,(\mathrm{source})} + N_{o\,(\mathrm{added})} \tag{2.13}$$

Noise factor can be written in several useful alternative forms:

$$F = \frac{N_{o\,(\mathrm{total})}}{G \cdot N_{i\,(\mathrm{source})}} = \frac{N_{o\,(\mathrm{total})}}{N_{o\,(\mathrm{source})}} = \frac{N_{o\,(\mathrm{source})} + N_{o\,(\mathrm{added})}}{N_{o\,(\mathrm{source})}} = 1 + \frac{N_{o\,(\mathrm{added})}}{N_{o\,(\mathrm{source})}} \tag{2.14}$$

This shows that the minimum possible noise factor, which occurs if the electronics adds no noise, is equal to 1. Noise figure NF is related to noise factor F by

$$\mathrm{NF} = 10 \log_{10} F \tag{2.15}$$

Thus, while noise factor is at least 1, noise figure is at least 0 dB. In other words, an electronic system that adds no noise has a noise figure of 0 dB.

In the receiver chain, for components with loss (such as switches and filters), the noise figure is equal to the attenuation of the signal. For example, a filter with 3 dB of loss has a noise figure of 3 dB. This is explained by noting that output noise is approximately equal to input noise, but signal is attenuated by 3 dB. Thus, there has been a degradation of SNR by 3 dB.

2.2.5 The Noise Figure of an Amplifier Circuit

We can now make use of the definition of noise figure just developed and apply it to an amplifier circuit [8]. For the purposes of developing (2.14) into a more useful form, it is assumed that all practical amplifiers can be characterized by an input-referred noise model, such as the one shown in Figure 2.2, where the amplifier is characterized with current gain A_i. (It will be shown in later chapters how to take a practical amplifier and make it fit this model.) In this model, all noise sources in the circuit are lumped into a series noise voltage source v_n and a parallel current noise source i_n placed in front of a noiseless transfer function.

If the amplifier has finite input impedance, then the input current will be split by some ratio α between the amplifier and the source admittance Y_s:

$$\mathrm{SNR_{in}} = \frac{\alpha^2 i_{\mathrm{in}}^2}{\alpha^2 i_{\mathrm{ns}}^2} \tag{2.16}$$

Assuming that the input-referred noise sources are correlated, the output signal-to-noise ratio is

$$\mathrm{SNR_{out}} = \frac{\alpha^2 A_i^2\, i_{\mathrm{in}}^2}{\alpha^2 A_i^2 \left(i_{\mathrm{ns}}^2 + |i_n + v_n Y_s|^2\right)} \tag{2.17}$$

Thus, the noise factor can now be written in terms of the preceding two equations:

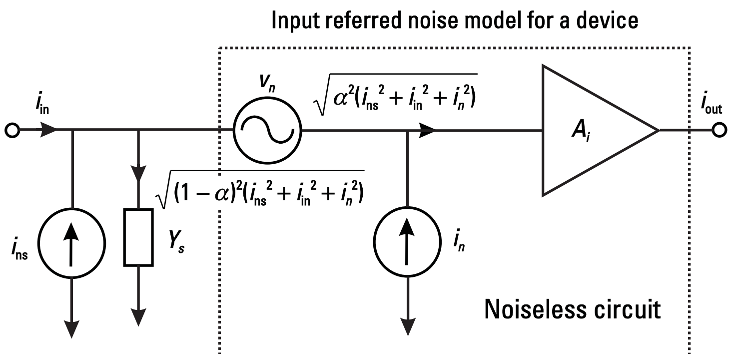

Figure 2.2 Input-referred noise model for a device.

$$F = \frac{i_{ns}^2 + |i_n + v_n Y_s|^2}{i_{ns}^2} = \frac{N_{o\,(\mathrm{total})}}{N_{o\,(\mathrm{source})}} \tag{2.18}$$

This can also be interpreted as the ratio of the total output noise to the total output noise due to the source admittance.

In (2.17), it was assumed that the two input noise sources were correlated with each other. In general, they will not be correlated with each other, but rather the current i_n will be partially correlated with v_n and partially uncorrelated. We can expand both current and voltage into these two explicit parts:

$$i_n = i_c + i_u \tag{2.19}$$

$$v_n = v_c + v_u \tag{2.20}$$

In addition, the correlated components will be related by the ratio

$$i_c = Y_c v_c \tag{2.21}$$

where Y_c is the correlation admittance.

The noise figure can now be written as

$$\mathrm{NF} = 1 + \frac{i_u^2 + |Y_c + Y_s|^2 v_c^2 + v_u^2 |Y_s|^2}{i_{ns}^2} \tag{2.22}$$

The noise currents and voltages can also be written in terms of equivalent resistance and admittance (these resistors would have the same noise behavior):

$$R_c = \frac{v_c^2}{4kT\Delta f} \tag{2.23}$$

$$R_u = \frac{v_u^2}{4kT\Delta f} \tag{2.24}$$

$$G_u = \frac{i_u^2}{4kT\Delta f} \tag{2.25}$$

$$G_s = \frac{i_{ns}^2}{4kT\Delta f} \tag{2.26}$$

Thus, the noise figure is now written in terms of these parameters:

$$\mathrm{NF} = 1 + \frac{G_u + |Y_c + Y_s|^2 R_c + |Y_s|^2 R_u}{G_s} \tag{2.27}$$

$$\mathrm{NF} = 1 + \frac{G_u + [(G_c + G_s)^2 + (B_c + B_s)^2] R_c + (G_s^2 + B_s^2) R_u}{G_s} \tag{2.28}$$

It can be seen from this equation that NF is dependent on the equivalent source impedance.

Equation (2.28) can be used not only to determine the noise figure, but also to determine the source loading conditions that will minimize the noise figure. Differentiating with respect to G_s and B_s and setting the derivative to zero yields the following two conditions for minimum noise (G_{opt} and B_{opt}) after several pages of math:

$$G_{\mathrm{opt}} = \sqrt{\frac{G_u + R_u\left(\dfrac{R_c B_c}{R_c + R_u}\right)^2 + G_c^2 R_c + \left(B_c - \dfrac{R_c B_c}{R_c + R_u}\right)^2 R_c}{R_c + R_u}} \tag{2.29}$$

$$B_{\mathrm{opt}} = \frac{-R_c B_c}{R_c + R_u} \tag{2.30}$$

2.2.6 The Noise Figure of Components in Series

For components in series, as shown in Figure 2.3, one can calculate the total output noise ($N_{o\,(\mathrm{total})}$) and output noise due to the source ($N_{o\,(\mathrm{source})}$) to determine the noise figure.

The output signal S_o is given by

$$S_o = S_i \cdot G_i \cdot G_2 \cdot G_3 \tag{2.31}$$

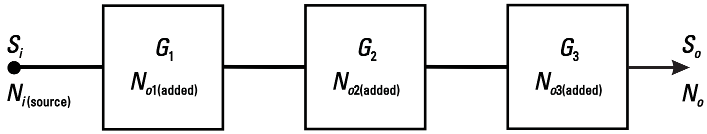

Figure 2.3 Noise figure in cascaded circuits with gain and noise added shown in each.

The input noise is

$$N_{i\,(\text{source})} = kT \tag{2.32}$$

The total output noise is

$$\begin{aligned} N_{o\,(\text{total})} &= N_{i\,(\text{source})} G_1 G_2 G_3 + N_{o1(\text{added})} G_2 G_3 \\ &\quad + N_{o2(\text{added})} G_3 + N_{o3(\text{added})} \end{aligned} \tag{2.33}$$

The output noise due to the source is

$$N_{o\,(\text{source})} = N_{i\,(\text{source})} G_1 G_2 G_3 \tag{2.34}$$

Finally, the noise factor can be determined as

$$\begin{aligned} F &= \frac{N_{o\,(\text{total})}}{N_{o\,(\text{source})}} = 1 + \frac{N_{o1(\text{added})}}{N_{i\,(\text{source})} G_1} + \frac{N_{o2(\text{added})}}{N_{i\,(\text{source})} G_1 G_2} + \frac{N_{o3(\text{added})}}{N_{i\,(\text{source})} G_1 G_2 G_3} \\ &= F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} \end{aligned} \tag{2.35}$$

The above formula shows how the presence of gain preceding a stage causes the effective noise figure to be reduced compared to the measured noise figure of a stage by itself. For this reason, we typically design systems with a low-noise amplifier at the front of the system. We note that the noise figure of each block is typically determined for the case in which a standard input source (e.g., 50Ω) is connected. The above formula can also be used to derive an equivalent model of each block as shown in Figure 2.4. If the input noise when measuring noise figure is

$$N_{i\,(\text{source})} = kT \tag{2.36}$$

and noting from manipulation of (2.14) that

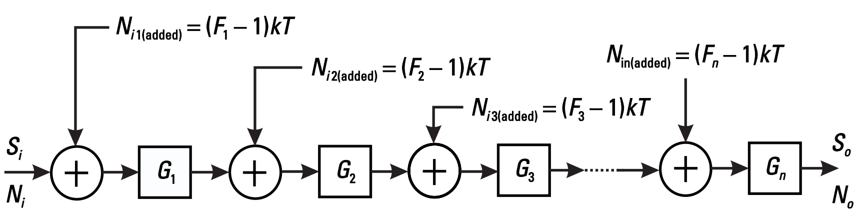

Figure 2.4 Equivalent noise model of a circuit.

$$N_{o1(\text{added})} = (F - 1)\, N_{o\,(\text{source})} \tag{2.37}$$

Now dividing both sides of (2.37) by G_1,

$$N_{i\,(\text{added})} = (F - 1)\,\frac{N_{o\,(\text{source})}}{G_1} = (F - 1)\, N_{i\,(\text{source})} = (F - 1)\, kT \tag{2.38}$$

Then the total input-referred noise to the first stage is

$$N_{i1} = N_{i\,(\text{source})} + (F_1 - 1)\, kT = kT + (F_1 - 1)\, kT = kTF_1 \tag{2.39}$$

Thus, the input-referred noise model for cascaded stages as shown in Figure 2.4 can be derived.

Example 2.1 Noise Calculations

Figure 2.5 shows a 50-Ω source resistance loaded with 50Ω. Determine how much noise voltage per unit bandwidth is present at the output. Then, for any R_L, what is the maximum noise power that this source can deliver to any load? Also find the noise factor, assuming that R_L does not contribute to noise factor, and compare to the case where R_L does contribute to noise factor.

Solution

The noise from the 50Ω source is $\sqrt{4kTR} \approx 0.9\ \text{nV}/\sqrt{\text{Hz}}$ at a temperature of 290K, which, after the voltage divider, becomes one half of this value, or $v_o = 0.45\ \text{nV}/\sqrt{\text{Hz}}$.

Now, for maximum power transfer, the load must remain matched, so $R_L = R_S = 50\Omega$. Then the complete available power from the source is delivered to the load. In this case,

$$P_o = \frac{v_o^2}{4R_L} = P_{\text{in(available)}}$$

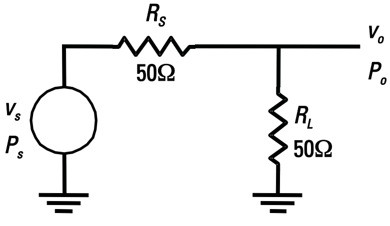

Figure 2.5 Simple circuit used for noise calculations.

$$P_{\text{in(available)}} = \frac{v_o^2}{4R_L} = \frac{4kTR_S}{4R_L} = kT = 4 \times 10^{-21}$$

At the output, the complete noise power (available) appears, and so if R_L is noiseless, the noise factor = 1. However, if R_L has noise of $\sqrt{4kTR_L}$ V/$\sqrt{\text{Hz}}$, then at the output, the total noise power is $2kT$, where kT is from R_S and kT is from R_L. Therefore, for a resistively matched circuit, the noise figure is 3 dB. Note that the output noise voltage is 0.45 nV/$\sqrt{\text{Hz}}$ from each resistor for a total of $\sqrt{2} \cdot 0.45$ nV/$\sqrt{\text{Hz}}$ = 0.636 nV/$\sqrt{\text{Hz}}$ (with noise the power adds because the noise voltage is uncorrelated).

Example 2.2 Noise Calculation with Gain Stages

In this example, Figure 2.6, a voltage gain of 20 has been added to the original circuit of Figure 2.5. All resistor values are still 50Ω. Determine the noise at the output of the circuit due to all resistors and then determine the circuit noise figure and signal-to-noise ratio assuming a 1-MHz bandwidth and the input is a 1-V sine wave.

Solution

In this example, at v_x the noise is still due to only R_S and R_2. As before, the noise at this point is 0.636 nV/$\sqrt{\text{Hz}}$. The signal at this point is 0.5V, thus at point v_y the signal is 10V and the noise due to the two input resistors R_S and R_2 is $0.636 \cdot 20 = 12.72$ nV/$\sqrt{\text{Hz}}$. At the output, the signal and noise from the input sources, as well as the noise from the two output resistors, all see a voltage divider. Thus, one can calculate the individual components. For the combination of R_S and R_2, one obtains

$$v_{R_S+R_2} = 0.5 \times 12.72 = 6.36 \text{ nV}/\sqrt{\text{Hz}}$$

The noise from the source can be determined from this equation:

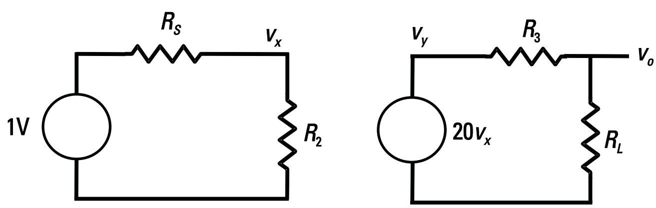

Figure 2.6 Noise calculation with a gain stage.

$$v_{R_S} = \frac{6.36\ \text{nV}/\sqrt{\text{Hz}}}{\sqrt{2}} = 4.5\ \text{nV}/\sqrt{\text{Hz}}$$

For the other resistors, the voltage is

$$v_{R_S} = 0.5 \cdot 0.9 = 0.45\ \text{nV}/\sqrt{\text{Hz}}$$

$$v_{R_L} = 0.5 \cdot 0.9 = 0.45\ \text{nV}/\sqrt{\text{Hz}}$$

Total output noise is given by

$$v_{\text{no(total)}} = \sqrt{v_{(R_S+R_L)}^2 + v_{R_S}^2 + v_{R_L}^2} = \sqrt{6.36^2 + 0.45^2 + 0.45^2}$$
$$= 6.392\ \text{nV}/\sqrt{\text{Hz}}$$

Therefore, the noise figure can now be determined:

$$\text{Noise factor} = F = \frac{N_{o\,\text{(total)}}}{N_{o\,\text{(source)}}} = \left(\frac{6.392}{4.5}\right)^2 = (1.417)^2 = 2.018$$

$$\text{NF} = 10 \log_{10} F = 10 \log_{10} 2.018 = 3.05\ \text{dB}$$

Since the output voltage also sees a voltage divider of 1/2, it has a value of 5V. Thus, the signal-to-noise ratio is

$$\frac{S}{N} = 20 \log\left(\frac{5}{\dfrac{6.392\ \text{nV}}{\sqrt{\text{Hz}}} \cdot \sqrt{1\ \text{MHz}}}\right) = 117.9\ \text{dB}$$

This example illustrates that noise from the source and amplifier input resistance are the dominant noise sources in the circuit. Each resistor at the input provides 4.5 nV/$\sqrt{\text{Hz}}$, while the two resistors behind the amplifier each only contribute 0.45 nV/$\sqrt{\text{Hz}}$. Thus, as explained earlier, after a gain stage, noise is less important.

Example 2.3 Effect of Impedance Mismatch on Noise Figure

Find the noise figure of Example 2.2 again, but now assume that $R_2 = 500\Omega$.

Solution

As before, the output noise due to the resistors is as follows:

$$v_{\text{no}(R_S)} = 0.9 \cdot \frac{500}{550} \cdot 20 \cdot 0.5 = 8.181\ \text{nV}/\sqrt{\text{Hz}}$$

where 500/550 accounts for the voltage division from the noise source to the node v_x.

$$v_{\text{no}(R_2)} = 0.9 \cdot \sqrt{10} \cdot \frac{50}{550} \cdot 20 \cdot 0.5 = 2.587 \text{ nV}/\sqrt{\text{Hz}}$$

where the $\sqrt{10}$ accounts for the higher noise in a 500-Ω resistor compared to a 50-Ω resistor.

$$v_{\text{no}(R_3)} = 0.9 \cdot 0.5 = 0.45 \text{ nV}/\sqrt{\text{Hz}}$$

$$v_{\text{no}(R_L)} = 0.9 \cdot 0.5 = 0.45 \text{ nV}/\sqrt{\text{Hz}}$$

The total output noise voltage is

$$\begin{aligned} v_{\text{no(total)}} &= \sqrt{v_{R_S}^2 + v_{R_2}^2 + v_{R_3}^2 + v_{R_L}^2} = \sqrt{8.181^2 + 2.587^2 + 0.45^2 + 0.45^2} \\ &= 8.604 \text{ nV}/\sqrt{\text{Hz}} \end{aligned}$$

$$\text{Noise factor} = F = \frac{N_{o\,(\text{total})}}{N_{o\,(\text{source})}} = \left(\frac{8.604}{8.181}\right)^2 = 1.106$$

$$\text{NF} = 10 \log_{10} F = 10 \log_{10} 1.106 = 0.438 \text{ dB}$$

Note: This circuit is unmatched at the input. This example illustrates that a mismatched circuit may have better noise performance than a matched one. However, this assumes that it is possible to build a voltage amplifier that requires little power at the input. This may be possible on an IC. However, if transmission lines are included, power transfer will suffer. A matching circuit may need to be added.

Example 2.4 Cascaded Noise Figure and Sensitivity Calculation

Find the effective noise figure and noise floor of the system shown in Figure 2.7. The system consists of a filter with 3-dB loss, followed by a switch with 1-dB loss, an LNA, and a mixer. Assume the system needs an SNR of 7 dB for a bit error rate of 10^{-3}. Also assume that the system bandwidth is 200 kHz.

Solution

Since the bandwidth of the system has been given as 200 kHz, the noise floor of the system can be determined:

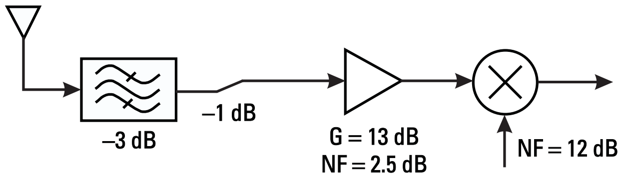

Figure 2.7 System for performance calculation.

$$\text{Noise floor} = -174 \text{ dBm} + 10 \log_{10}(200{,}000) = -121 \text{ dBm}$$

We make use of the cascaded noise figure equation and determine that the overall system noise figure is given by

$$NF_{TOTAL} = 3 \text{ dB} + 1 \text{ dB} + 10 \log_{10}\left[1.78 + \frac{15.84 - 1}{20}\right] \approx 8 \text{ dB}$$

Note that the LNA noise figure of 2.5 dB corresponds to a noise factor of 1.78 and the gain of 13 dB corresponds to a power gain of 20. Furthermore, the noise figure of 12 dB corresponds to a noise factor of 15.84.

Note that if the mixer also has gain, then possibly the noise due to the IF stage may be ignored. In a real system this would have to be checked, but here we will ignore noise in the IF stage.

Since it was stated that the system requires an SNR of 7 dB, the sensitivity of the system can now be determined:

$$\text{Sensitivity} = -121 \text{ dBm} + 7 \text{ dB} + 8 \text{ dB} = -106 \text{ dBm}$$

Thus, the smallest allowable input signal is −106 dBm. If this is not adequate for a given application, then a number of things can be done to improve this:

1. A smaller bandwidth could be used. This is usually fixed by IF requirements.
2. The loss in the preselect filter or switch could be reduced. For example, the LNA could be placed in front of one or both of these components.
3. The noise figure of the LNA could be improved.
4. The LNA gain could be increased reducing the effect of the mixer on the system NF.
5. A lower NF in the mixer would also improve the system NF.
6. If a lower SNR for the required BER could be tolerated, then this would also help.

2.3 Linearity and Distortion in RF Circuits

In an ideal system, the output is linearly related to the input. However, in any real device the transfer function is usually a lot more complicated. This can be due to active or passive devices in the circuit or the signal swing being limited by the power supply rails. Unavoidably, the gain curve for any component is never a perfectly straight line, as illustrated in Figure 2.8.

The resulting waveforms can appear as shown in Figure 2.9. For amplifier saturation, typically the top and bottom portions of the waveform are clipped equally, as shown in Figure 2.9(b). However, if the circuit is not biased between the two clipping levels, then clipping can be nonsymmetrical as shown in Figure 2.9(c).

2.3.1 Power Series Expansion

Mathematically, any nonlinear transfer function can be written as a series expansion of power terms unless the system contains memory, in which case a Volterra series is required [9, 10]:

$$v_{out} = k_0 + k_1 v_{in} + k_2 v_{in}^2 + k_3 v_{in}^3 + \ldots \tag{2.40}$$

To describe the nonlinearity perfectly, an infinite number of terms is required; however, in many practical circuits, the first three terms are sufficient to characterize the circuit with a fair degree of accuracy.

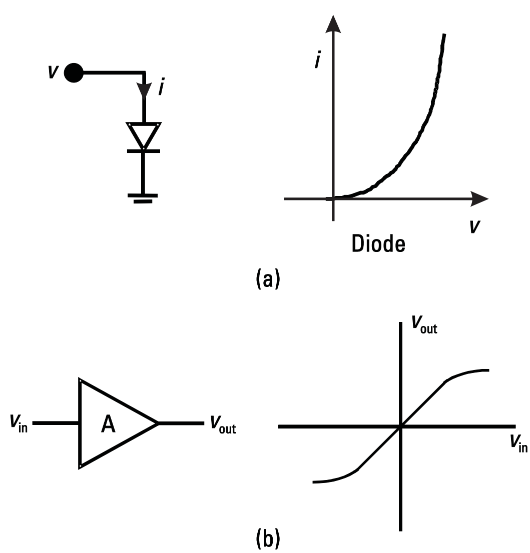

Figure 2.8 Illustration of the nonlinearity in (a) a diode, and (b) an amplifier.

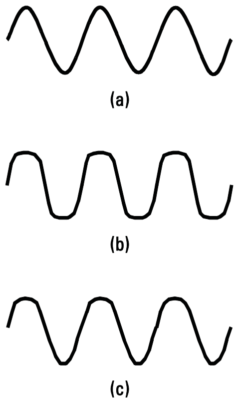

Figure 2.9 Distorted output waveforms: (a) input; (b) output, clipping; and (c) output, bias wrong.

Symmetrical saturation as shown in Figure 2.8(b) can be modeled with odd order terms; for example,

$$y = x - \frac{1}{10}x^3 \tag{2.41}$$

looks like Figure 2.10. In another example, an exponential nonlinearity as shown in Figure 2.8(a) has the form

$$x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots \tag{2.42}$$

which contains both even and odd power terms because it does not have symmetry about the y-axis. Real circuits will have more complex power series expansions.

One common way of characterizing the linearity of a circuit is called the two-tone test. In this test, an input consisting of two sine waves is applied to the circuit.

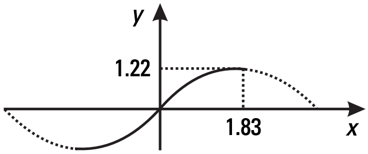

Figure 2.10 Example of output or input nonlinearity with first- and third-order terms.

$$v_{\text{in}} = v_1 \cos \omega_1 t + v_2 \cos \omega_2 t = X_1 + X_2 \tag{2.43}$$

When this tone is applied to the transfer function given in (2.40), the result is a number of terms:

$$v_0 = k_0 + \underbrace{k_1(X_1 + X_2)}_{\text{desired}} + \underbrace{k_2(X_1 + X_2)^2}_{\text{second order}} + \underbrace{k_3(X_1 + X_2)^3}_{\text{third order}} \tag{2.44}$$

$$\begin{aligned} v_0 = {} & k_0 + k_1(X_1 + X_2) + k_2(X_1^2 + 2X_1X_2 + X_2^2) \\ & + k_3(X_1^3 + 3X_1^2X_2 + 3X_1X_2^2 + X_1^3) \end{aligned} \tag{2.45}$$

These terms can be further broken down into various frequency components. For instance, the X_1^2 term has a zero frequency (dc) component and another at the second harmonic of the input:

$$X_1^2 = (v_1 \cos \omega_1 t)^2 = \frac{v_1^2}{2}(1 + \cos 2\omega_1 t) \tag{2.46}$$

The second-order terms can be expanded as follows:

$$(X_1 + X_2)^2 = \underbrace{X_1^2}_{\substack{\text{dc +} \\ \text{HD2}}} + \underbrace{2X_1X_2}_{\text{MIX}} + \underbrace{X_2^2}_{\substack{\text{dc +} \\ \text{HD2}}} \tag{2.47}$$

where second-order terms are composed of second harmonics HD2, and mixing components, here labeled MIX but sometimes labeled IM2 for second-order intermodulation. The mixing components will appear at the sum and difference frequencies of the two input signals. Note also that second-order terms cause an additional dc term to appear.

The third-order terms can be expanded as follows:

$$(X_1 + X_2)^3 = \underbrace{X_1^3}_{\substack{\text{FUND} \\ \text{+ HD3}}} + \underbrace{3X_1^2X_2}_{\substack{\text{IM3 +} \\ \text{FUND}}} + \underbrace{3X_1X_2^2}_{\substack{\text{IM3 +} \\ \text{FUND}}} + \underbrace{X_2^3}_{\substack{\text{FUND} \\ \text{+ HD3}}} \tag{2.48}$$

Third-order nonlinearity results in third harmonics HD3 and third-order intermodulation IM3. Expansion of both the HD3 and IM3 terms shows output signals appearing at the input frequencies. The effect is that third-order nonlinearity can change the gain, which is seen as gain compression. This is summarized in Table 2.1.

Table 2.1
Summary of Distortion Components

Frequency	Component Amplitude
dc	$k_o + \frac{k_2}{2}(v_1^2 + v_2^2)$
ω_1	$k_1 v_1 + k_3 v_1\left(\frac{3}{4}v_1^2 + \frac{3}{2}v_2^2\right)$
ω_2	$k_1 v_2 + k_3 v_2\left(\frac{3}{4}v_2^2 + \frac{3}{2}v_1^2\right)$
$2\omega_1$	$\frac{k_2 v_1^2}{2}$
$2\omega_2$	$\frac{k_2 v_2^2}{2}$
$\omega_1 \pm \omega_2$	$k_2 v_1 v_2$
$\omega_2 \pm \omega_1$	$k_2 v_1 v_2$
$3\omega_1$	$\frac{k_3 v_1^3}{4}$
$3\omega_2$	$\frac{k_3 v_2^3}{4}$
$2\omega_1 \pm \omega_2$	$\frac{3}{4}k_3 v_1^2 v_2$
$2\omega_2 \pm \omega_1$	$\frac{3}{4}k_3 v_1 v_2^2$

Note that in the case of an amplifier, only the terms at the input frequency are desired. Of all the unwanted terms, the last two at frequencies $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ are the most troublesome, since they can fall in the band of the desired outputs if ω_1 is close in frequency to ω_2 and therefore cannot be easily filtered out. These two tones are usually referred to as third-order intermodulation terms (IM3 products).

Example 2.5 Determination of Frequency Components Generated in a Nonlinear System

Consider a nonlinear circuit with 7- and 8-MHz tones applied at the input. Determine all output frequency components, assuming distortion components up to the third order.

Solution

Table 2.2 and Figure 2.11 show the outputs.

It is apparent that harmonics can be filtered out easily, while the third-order intermodulation terms, being close to the desired tones, may be difficult to filter.

Table 2.2
Outputs from Nonlinear Circuits with Inputs at $f_1 = 7$, $f_2 = 8$ MHz

	Symbolic Frequency	Example Frequency	Name	Comment
First order	f_1, f_2	7, 8	Fundamental	Desired output
Second order	$2f_1$, $2f_2$	14, 16	HD2 (harmonics)	Can filter
	$f_2 - f_1$, $f_2 + f_1$	2, 15	IM2 (mixing)	Can filter
Third order	$3f_1$, $3f_2$	21, 24	HD3 (harmonic)	Can filter harmonics
	$2f_1 - f_2$,	6	IM3 (intermod)	Close to fundamental,
	$2f_2 - f_1$	9	IM3 (intermod)	difficult to filter

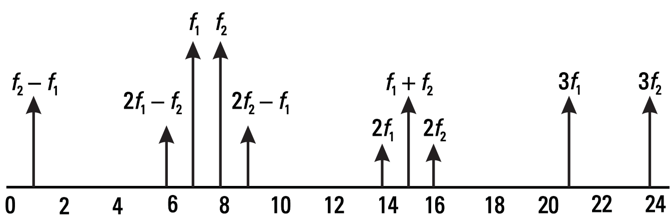

Figure 2.11 Output spectrum with inputs at 7 and 8 MHz.

2.3.2 Third-Order Intercept Point

One of the most common ways to test the linearity of a circuit is to apply two signals at the input, having equal amplitude and offset by some frequency, and plot fundamental output and intermodulation output power as a function of input power as shown in Figure 2.12. From the plot, the *third-order intercept point* (IP3) is determined. The third-order intercept point is a theoretical point where the amplitudes of the intermodulation tones at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ are equal to the amplitudes of the fundamental tones at ω_1 and ω_2.

From Table 2.1, if $v_1 = v_2 = v_i$, then the fundamental is given by

$$\text{fund} = k_1 v_i + \frac{9}{4} k_3 v_i^3 \tag{2.49}$$

The linear component of (2.49) given by

$$\text{fund} = k_1 v_i \tag{2.50}$$

can be compared to the third-order intermodulation term given by

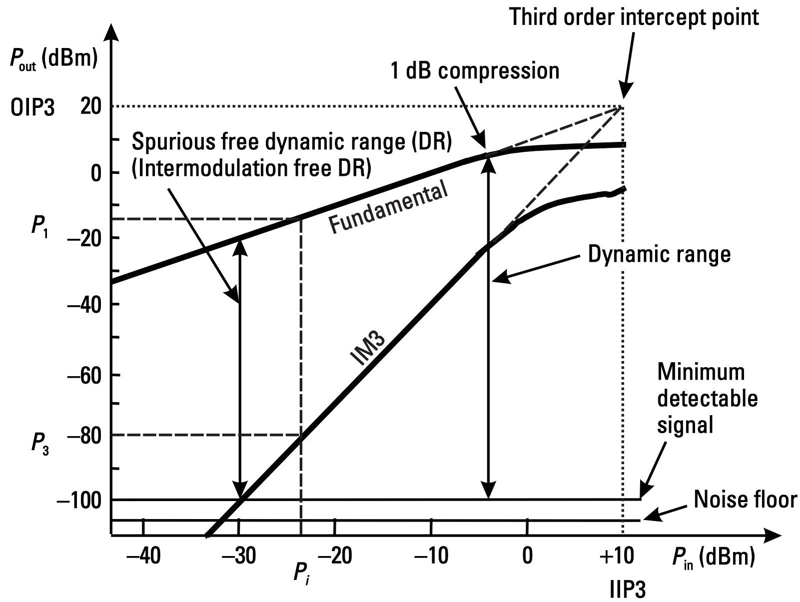

Figure 2.12 Plot of input output power of fundamental and IM3 versus input power.

$$\mathrm{IM3} = \frac{3}{4} k_3 v_i^3 \tag{2.51}$$

Note that for small v_i, the fundamental rises linearly (20 dB/decade) and that the IM3 terms rise as the cube of the input (60 dB/decade). A theoretical voltage at which these two tones will be equal can be defined:

$$\frac{\frac{3}{4} k_3 v_{\mathrm{IP3}}^3}{k_1 v_{\mathrm{IP3}}} = 1 \tag{2.52}$$

This can be solved for v_{IP3}:

$$v_{\mathrm{IP3}} = 2\sqrt{\frac{k_1}{3k_3}} \tag{2.53}$$

Note that (2.53) gives the input voltage at the third-order intercept point. The input power at this point is called the *input third-order intercept point* (IIP3). If IP3 is specified at the output, it is called the *output third-order intercept point* (OIP3).

Of course, the third-order intercept point cannot actually be measured directly, since by the time the amplifier reached this point, it would be heavily overloaded. Therefore, it is useful to describe a quick way to extrapolate it at a given power level. Assume that a device with power gain G has been measured to have an output power of P_1 at the fundamental frequency and a power of P_3 at the IM3 frequency for a given input power of P_i, as illustrated in Figure 2.12. Now, on a log plot (for example, when power is in dBm) of P_3 and P_1 versus P_i, the IM3 terms have a slope of 3 and the fundamental terms have a slope of 1. Therefore,

$$\frac{\text{OIP3} - P_1}{\text{IIP3} - P_i} = 1 \tag{2.54}$$

$$\frac{\text{OIP3} - P_3}{\text{IIP3} - P_i} = 3 \tag{2.55}$$

since subtraction on a log scale amounts to division of power.

Also note that

$$G = \text{OIP3} - \text{IIP3} = P_1 - P_i \tag{2.56}$$

These equations can be solved to give

$$\text{IIP3} = P_1 + \frac{1}{2}[P_1 - P_3] - G = P_i + \frac{1}{2}[P_1 - P_3] \tag{2.57}$$

2.3.3 Second-Order Intercept Point

A *second-order intercept point* (IP2) can be defined that is similar to the third-order intercept point. Which one is used depends largely on which is more important in the system of interest; for example, second-order distortion is particularly important in direct downconversion receivers.

If two tones are present at the input, then the second-order output is given by

$$v_{\text{IM2}} = k_2 v_i^2 \tag{2.58}$$

Note that in this case, the IM2 terms rise at 40 dB/decade rather than at 60 dB/decade, as in the case of the IM3 terms.

The theoretical voltage at which the IM2 term will be equal to the fundamental term given in (2.50) can be defined:

$$\frac{k_2 v_{\mathrm{IP2}}^2}{k_1 v_{\mathrm{IP2}}} = 1 \tag{2.59}$$

This can be solved for v_{IP2}:

$$v_{\mathrm{IP2}} = \frac{k_1}{k_2} \tag{2.60}$$

2.3.4 The 1-dB Compression Point

In addition to measuring the IP3 or IP2 of a circuit, the 1-dB compression point is another common way to measure linearity. This point is more directly measurable than IP3 and requires only one tone rather than two (although any number of tones can be used). The 1-dB compression point is simply the power level, specified at either the input or the output, where the output power is 1 dB less than it would have been in an ideally linear device. It is also marked in Figure 2.12.

We first note that at 1-dB compression, the ratio of the actual output voltage v_o to the ideal output voltage v_{oi} is

$$20 \log_{10}\left(\frac{v_o}{v_{\mathrm{oi}}}\right) = -1 \text{ dB} \tag{2.61}$$

or

$$\frac{v_o}{v_{\mathrm{oi}}} = 0.89125 \tag{2.62}$$

Now referring again to Table 2.1, we note that the actual output voltage for a single tone is

$$v_o = k_1 v_i + \frac{3}{4} k_3 v_i^3 \tag{2.63}$$

for an input voltage v_i. The ideal output voltage is given by

$$v_{\mathrm{oi}} = k_1 v_i \tag{2.64}$$

Thus, the 1-dB compression point can be found by substituting (2.63) and (2.64) into (2.62):

$$\frac{k_1 v_{1dB} + \frac{3}{4} k_3 v_{1dB}^3}{k_1 v_{1dB}} = 0.89125 \tag{2.65}$$

Note that for a nonlinearity that causes compression, rather than one that causes expansion, k_3 has to be negative. Solving (2.65) for v_{1dB} gives

$$v_{1dB} = 0.38\sqrt{\frac{k_1}{k_3}} \tag{2.66}$$

If more than one tone is applied, the 1-dB compression point will occur for a lower input voltage. In the case of two equal amplitude tones applied to the system, the actual output power for one frequency is

$$v_o = k_1 v_i + \frac{9}{4} k_3 v_i^3 \tag{2.67}$$

The ideal output voltage is still given by (2.64). So now the ratio is

$$\frac{k_1 v_{1dB} + \frac{9}{4} k_3 v_{1dB}^3}{k_1 v_{1dB}} = 0.89125 \tag{2.68}$$

Therefore, the 1-dB compression voltage is now

$$v_{1dB} = 0.22\sqrt{\frac{k_1}{k_3}} \tag{2.69}$$

Thus, as more tones are added, this voltage will continue to get lower.

2.3.5 Relationships Between 1-dB Compression and IP3 Points

In the last two sections, formulas for the IP3 and the 1-dB compression point have been derived. Since we now have expressions for both these values, we can find a relationship between these two points. Taking the ratio of (2.53) and (2.66) gives

$$\frac{v_{IP3}}{v_{1dB}} = \frac{2\sqrt{\frac{k_1}{3k_3}}}{0.38\sqrt{\frac{k_1}{k_3}}} = 3.04 \tag{2.70}$$

Thus, these voltages are related by a factor of 3.04, or about 9.66 dB, independent of the particulars of the nonlinearity in question. In the case of the 1-dB compression point with two tones applied, the ratio is larger. In this case,

$$\frac{v_{\mathrm{IP3}}}{v_{1\mathrm{dB}}} = \frac{2\sqrt{\dfrac{k_1}{3k_3}}}{0.22\sqrt{\dfrac{k_1}{k_3}}} = 5.25 \tag{2.71}$$

Thus, these voltages are related by a factor of 5.25 or about 14.4 dB.

Thus, one can estimate that for a single tone, the compression point is about 10 dB below the intercept point, while for two tones, the 1-dB compression point is close to 15 dB below the intercept point. The difference between these two numbers is just the factor of three (4.77 dB) resulting from the second tone.

Note that this analysis is valid for third-order nonlinearity. For stronger nonlinearity (i.e., containing fifth-order terms), additional components are found at the fundamental as well as at the intermodulation frequencies. Nevertheless, the above is a good estimate of performance.

Example 2.6 Determining IIP3 and 1-dB Compression Point from Measurement Data

An amplifier designed to operate at 2 GHz with a gain of 10 dB has two signals of equal power applied at the input. One is at a frequency of 2.0 GHz and another at a frequency of 2.01 GHz. At the output, four tones are observed at 1.99, 2.0, 2.01, and 2.02 GHz. The power levels of the tones are −70, −20, −20, and −70 dBm, respectively. Determine the IIP3 and 1-dB compression point for this amplifier.

Solution

The tones at 1.99 and 2.02 GHz are the IP3 tones. We can use (2.57) directly to find the IIP3:

$$\mathrm{IIP3} = P_1 + \frac{1}{2}[P_1 - P_3] - G = -20 + \frac{1}{2}[-20 + 70] - 10 = -5 \text{ dBm}$$

The 1-dB compression point for a signal tone is 9.66 dB lower than this value, about −14.7 dBm at the input.

2.3.6 Broadband Measures of Linearity

Intercept and 1-dB compression points are two common measures of linearity, but they are by no means the only ones. Many others exist and, in fact, more could be defined. Two other measures of linearity that are common in wide-band systems handling many signals simultaneously are called *composite triple-order beat* (CTB) and *composite second-order beat* (CSO) [11, 12]. In these tests of linearity, N signals of voltage v_i are applied to the circuit equally spaced in frequency, as shown in Figure 2.13. Note here that, as an example, the tones are spaced 6 MHz apart (this is the spacing for a cable television system for which this is a popular way to characterize linearity). Note also that the tones are never placed at a frequency that is an exact multiple of the spacing (in this case, 6 MHz). This is done so that third-order terms and second-order terms fall at different frequencies. This will be clarified shortly.

If we take three of these signals, then the third-order nonlinearity gets a little more complicated than before:

$$\begin{aligned}(x_1 + x_2 + x_3)^3 = {} & \underbrace{x_1^3 + x_2^3 + x_3^3}_{\text{HD3}} \\ & + \underbrace{3x_1^2x_2 + 3x_1^2x_3 + 3x_2^2x_1 + 3x_3^2x_1 + 3x_2^2x_3 + 3x_3^2x_2}_{\text{IM3}} \\ & + \underbrace{6x_1x_2x_3}_{\text{TB}} \end{aligned} \tag{2.72}$$

The last term in the expression causes CTB in that it creates terms at frequencies $\omega_1 \pm \omega_2 \pm \omega_3$ of magnitude $1.5k_3v_i$ where $\omega_1 < \omega_2 < \omega_3$. This is twice as large as the IM3 products. Note that, except for the case where all three are added $(\omega_1 + \omega_2 + \omega_3)$, these tones can fall into any of the channels being used and many will fall into the same channel. For instance, in Figure

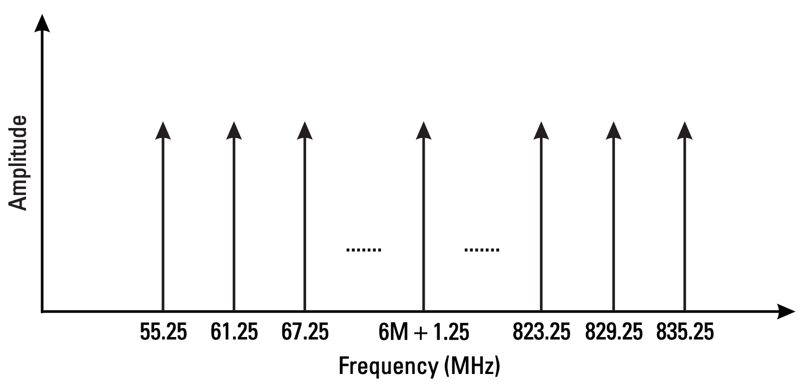

Figure 2.13 Equally spaced tones entering a broadband circuit.

2.13, 67.25 − 73.25 + 79.25 = 73.25 MHz and 49.25 − 55.25 + 79.25 = 73.25 MHz will both fall on the 73.25-MHz frequency. In fact, there will be many more *triple-beat* (TB) products than IM3 products. Thus, these terms become more important in a wide-band system. It can be shown that the maximum number of terms will fall on the tone at the middle of the band. With N tones, it can be shown that the number of tones falling there will be

$$\text{Tones} = \frac{3}{8}N^2 \tag{2.73}$$

We have already said that the voltage of these tones is twice that of the IP3 tones. We also note here that if the signal power is backed off from the IP3 power by some amount, the power in the IP3 tones will be backed off three times as much (calculated on a logarithmic scale). Therefore, if each fundamental tone is at a power level of P_s, then the power of the TB tones will be

$$\text{TB (dBm)} = P_{\text{IP3}} - 3(P_{\text{IP3}} - P_s) + 6 \tag{2.74}$$

where P_{IP3} is the IP3 power level for the given circuit.

Now, assuming that all tones add as power rather than voltage, and noting that CTB is usually specified as so many decibels down from the signal power,

$$\text{CTB (dB)} = P_s - \left[P_{\text{IP3}} - 3(P_{\text{IP3}} - P_s) + 6 + 10\log\left(\frac{3}{8}N^2\right)\right] \tag{2.75}$$

Note that CTB could be found using either input- or output-referred power levels.

Similar to the CTB is the CSO, which can also be used to measure the linearity of a broadband system. Again, if we have N signals all at the same power level, we now consider the second-order distortion products of each pair of signals falling at frequencies $\omega_1 \pm \omega_2$. In this case, the signals fall at frequencies either above or below the carriers rather than right on top of them, as in the case of the triple-beat terms, provided that the carriers are not some even multiple of the channel spacing. For example, in Figure 2.13, 49.25 + 55.25 = 104.5 MHz. This is 1.25 MHz above the closest carrier at 103.25 MHz. All the sum terms will fall 1.25 MHz above the closest carrier, while the difference terms such as 763.25 − 841.25 = 78, will fall 1.25 MHz below the closest carrier at 79.25 MHz. Thus, the second-order and third-order terms can be measured separately. The number of terms that fall next to any given carrier will vary. Some of the $\omega_1 + \omega_2$ terms will fall out of band and the maximum number in band will fall next to the highest frequency carrier. The number of second-order beats above any given carrier is given by

$$N_B = (N-1)\frac{f - 2f_L + d}{2(f_H - f_L)} \tag{2.76}$$

where N is the number of carriers, f is the frequency of the measurement channel, f_L is the frequency of the lowest channel, f_H is the frequency of the highest channel, and d is the frequency offset from a multiple of the channel spacing (1.25 MHz in Figure 2.13).

For the case of the difference frequency second-order beats, there are more of these at lower frequencies, and the maximum number will be next to the lowest frequency carrier. In this case, the number of second-order products next to any carrier can be approximated by

$$N_B = (N-1)\left(1 - \frac{f - d}{f_H - f_L}\right) \tag{2.77}$$

Each of the second-order beats is an IP2 tone. Therefore, if each fundamental tone is at a power level of P_s, then the power of the *second-order beat* (SO) tones will be

$$\text{SO (dBm)} = P_{\text{IP2}} - 2(P_{\text{IP2}} - P_s) \tag{2.78}$$

Thus, the composite second-order beat product will be given by

$$\text{CSO (dB)} = P_s - [P_{\text{IP2}} - 2(P_{\text{IP2}} - P_s) + 10\log(N_B)] \tag{2.79}$$

2.4 Dynamic Range

So far, we have discussed noise and linearity in circuits. Noise determines how small a signal a receiver can handle, while linearity determines how large a signal a receiver can handle. If operation up to the 1-dB compression point is allowed (for about 10% distortion, or IM3 is about −20 dB with respect to the desired output), then the dynamic range is from the minimum detectable signal to this point. This is illustrated in Figure 2.12. In this figure, intermodulation components are above the minimum detectable signal for $P_{\text{in}} > -30$ dBm, for which $P_{\text{out}} = -20$ dBm. Thus, for any P_{out} between the minimum detectable signal of −100 dBm and −20 dBm, no intermodulation components can be seen, so the spurious free dynamic range is 80 dB.

Example 2.7 Determining Dynamic Range

In Example 2.4 we determined the sensitivity of a receiver system. Figure 2.14 shows this receiver again with the linearity of the mixer and LNA specified. Determine the dynamic range of this receiver.

Solution

The overall receiver has a gain of 19 dB. The minimum detectable signal from Example 2.4 is −106 dBm or −87 dBm at the output. The IIP3 of the LNA referred to the input is −5 dBm + 4 = −1 dBm. The IIP3 of the mixer referred to the input is 0 − 13 + 4 = −9 dBm. Therefore, the mixer dominates the IIP3 for the receiver. The 1-dB compression point will be 9.6 dB lower than this, or −18.6 dBm. Thus, the dynamic range of the system will be −18.6 + 106 = 87.4 dB.

Example 2.8 Effect of Bandwidth on Dynamic Range

The data transfer rate of the previous receiver can be greatly improved if we use a bandwidth of 80 MHz rather than 200 kHz. What does this do to the dynamic range of the receiver?

Solution

This system is the same as the last one except that now the bandwidth is 80 MHz. Thus, the noise floor is now

$$\text{Noise floor} = -174 \text{ dBm} + 10 \log_{10}(80 \times 10^6) = -95 \text{ dBm}$$

Assuming that the same signal-to-noise ratio is required:

$$\text{Sensitivity} = -95 \text{ dBm} + 7 \text{ dB} + 8 \text{ dB} = -80 \text{ dBm}$$

Thus, the dynamic range is now −15.6 + 80 = 64.4 dB. In order to get this back to the value in the previous system, we would need to increase the linearity of the receiver by 25.3 dB. As we will see in future chapters, this would be no easy task.

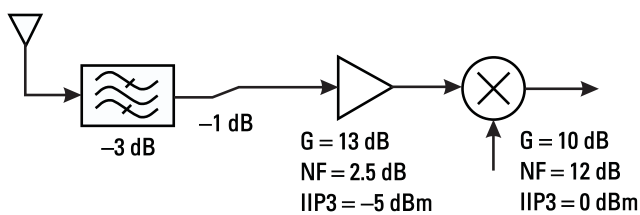

Figure 2.14 Circuit for system example.

2.5 Filtering Issues

To determine noise floor, the system bandwidth has to be known. The system bandwidth is set by filters, so it becomes necessary to discuss some of the filtering issues. There are additional reasons for needing filtering. The receiver must be able to maintain operation and to detect the desired signal in the presence of other signals often referred to as blocking signals. These other signals could be of large amplitude and could be close by in frequency. Such signals must be removed by filters, so a very general discussion of filters is in order. Actual monolithic filter circuits will be discussed in a later chapter.

2.5.1 Image Signals and Image Reject Filtering

The task of the receiver front end is to take the RF input and mix it either to baseband or to some IF where it can be more easily processed. A receiver in which the signal is taken directly to base band is called a *homodyne* or *direct-conversion receiver.* Although simpler than a receiver that takes the signal to some IF first (called a *superheterodyne receiver*), direct-conversion receivers suffer from numerous problems, including dc offsets, because much of the information is close to dc and also because of LO self-mixing [13]. A typical superheterodyne receiver front end consists of an LNA, an image filter, a mixer, and a VCO, as shown in Figure 2.15. An alternative to the image filter is to use an image reject mixer, which will be discussed in detail in Chapter 7. The image filter is required to suppress the unwanted image frequency, which is located a distance of two IFs away from the desired radio frequency [14]. Also, the image filter must prevent noise at the image frequency from mixing down to the IF and increasing the noise figure.

A superheterodyne receiver takes the desired RF input signal and mixes it with some reference signal to extract the difference frequency, as shown in Figure 2.16. The LO reference is mixed with the input to produce a signal at the difference frequency of the LO and RF. The problem is that a signal on the other side of the LO at the same distance from the LO will also mix down

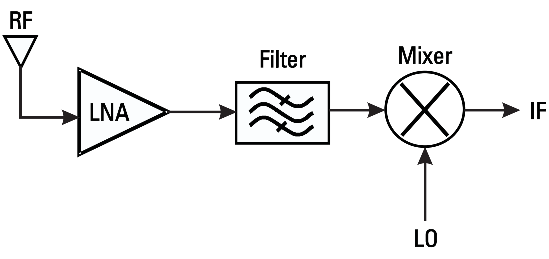

Figure 2.15 A block-level diagram of a superheterodyne receiver front end.

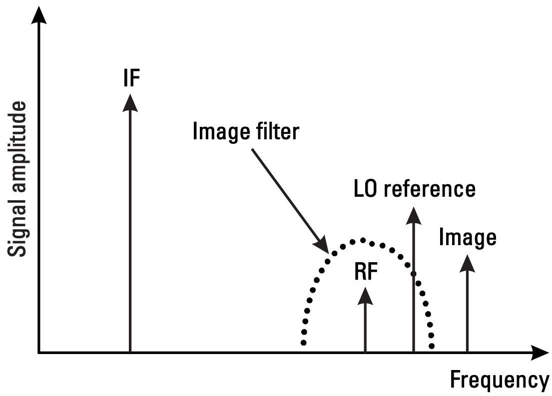

Figure 2.16 Translation of the RF signal to an IF in a superheterodyne receiver.

"on top" of the desired frequency. Thus, before mixing can take place, this unwanted image frequency must be removed. Typically, this is done with a filter that attenuates the image.

Thus, another important specification in a receiver is how much image rejection it has. Image rejection is defined as the ratio of the gain of the desired signal through the receiver G_{sig} to the gain of the image signal through the receiver G_{im}.

$$\text{IR} = 10 \log\left(\frac{G_{\text{sig}}}{G_{\text{im}}}\right) \tag{2.80}$$

The amount of filtering provided can be calculated by knowing the undesired frequency with respect to the filter center frequency, the filter bandwidth, and filter order. The following equation can be used for this calculation:

$$A_{\text{dB}} = \frac{n}{2} \cdot 20 \log\left(\frac{f_{\text{ud}} - f_c}{f_{\text{be}} - f_c}\right) = \frac{n}{2} \cdot 20 \log\left(2\frac{\Delta f}{f_{\text{BW}}}\right) \tag{2.81}$$

where A_{dB} is the attenuation in decibels, n is the filter order (and thus $n/2$ is the effective order on each edge), f_{ud} is the frequency of the undesired signal, f_c is the filter center frequency, f_{be} is the filter band edge, Δf is $f_{\text{ud}} - f_c$, and f_{BW} is $2(f_{\text{be}} - f_c)$.

Example 2.9 Image Reject Filtering

A system has an RF band from 902 to 928 MHz and a 200-kHz channel bandwidth and channel spacing. The first IF is at 70 MHz. With a 26-MHz image-reject filter, determine the order of filter required to get a worst-case image rejection of better than 50 dB.

Solution

The frequency spectrum is shown in Figure 2.17. At RF, the local oscillator frequency f_{LO} is tuned to be 70 MHz above the desired RF signal so that the desired signal will be mixed down to IF at 70 MHz. Thus, f_{LO} is adjustable between 972 and 998 MHz to allow signals between 902 and 928 MHz to be received. Any signal or noise 70 MHz above f_{LO} will also mix into the IF stage. This is known as the *image frequency.* An image reject filter is required to prevent any image signals from entering the mixer. The worst case will be when the image frequency is closest to the filter frequency. This occurs when the input is at 902 MHz, the LO is at 972 MHz, and the image is 1,042 MHz. The required filter order n can be calculated by solving (2.81) using $f_{BW} = 26$ MHz and $\Delta f = 70 + 44 + 13 = 127$ MHz as follows:

$$n = \frac{2 \cdot A_{dB}}{20 \cdot \log(2\Delta f / f_{BW})} = 5.05$$

Since the order is an even number, a sixth-order filter is used and total attenuation is calculated to be 59.4 dB.

2.5.2 Blockers and Blocker Filtering

Large unwanted signals can block the desired signal. This can happen when the desired signal is small and the undesired signal is large, for example, when the desired signal is far away and the undesired signal is close. If the result is that the receiver is overloaded, the desired signal cannot be received. This situation is known as *blocking.* If the blockers are in the desired frequency band, then filters do not help until the IF stage is reached.

Example 2.10 How Blockers Are Used To Determine Linearity

Consider the typical blocker specifications for a *Global System Mobile* (GSM) receiver shown in Figure 2.18. In the presence of the blockers, the input signal

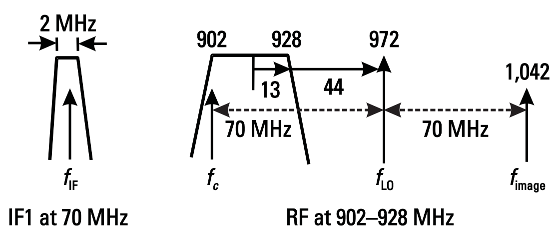

Figure 2.17 Signal spectrum for filter example.

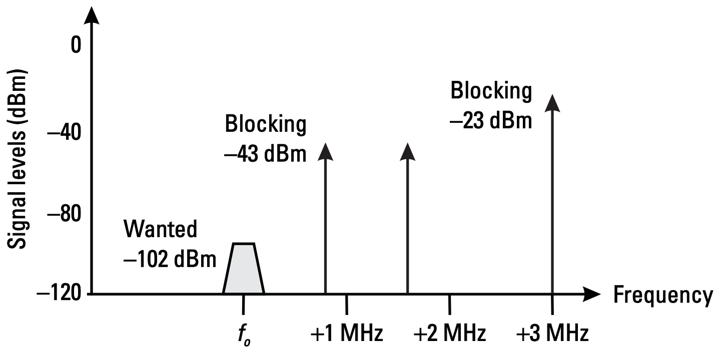

Figure 2.18 GSM minimum detectable signal and blocker levels.

is at −102 dBm and the required signal-to-noise ratio, with some safety margin, is 11 dB. Calculate the required input linearity of the GSM receiver.

Solution

This is an example of the so-called near-far problem that occurs when the desired signal is far away and one or more interfering signals are close by and hence much larger than the wanted signal. So what will be the effect of the blockers? With nonlinearity, third-order intermodulation between the pair of blockers will cause interference directly on top of the signal. The level of this disturbance must be low enough so that the signal can still be detected. The other potential problem is that the large blocker at –23 dBm can cause the amplifier to saturate, rendering the amplifier helpless to respond to the desired signal, which is much smaller. In other words, the receiver has been blocked.

As an estimate, the blocker inputs at –43 dBm will result in third-order intermodulation components (referred to the input) which must be less than −113 dBm, so there is still 11 dB of SNR at the input. Thus, the third-order components (at −113 dBm) are 70 dB below the fundamental components (at –43 dBm). Using (2.57) with P_i at –43 dBm and $[P_1 - P_3] = 70$ dB results in IIP3 of about −8 dBm. Going by this number, the 1-dB compression point is at about −18 dBm at the input. Thus, the single input blocker at –23 dBm is still 5 dB away from the 1-dB compression point. This sounds safe, although there will now be gain through the LNA and the mixer. The blocker will not be filtered until after the mixer, so one must be careful not to saturate any of the components along this path.

The blocking signals can cause problems in a receiver through another mechanism known as *reciprocal mixing*. For a blocker at an offset of Δf from the desired signal, if the oscillator also has a component at the same offset Δf from the carrier, then the blocking signal will be mixed directly to the IF.

Example 2.11 Calculating Maximum Level of Synthesizer Spurs

For the previous GSM specifications, calculate the allowable noise in a synthesizer in the presence of the blocking signals.

Solution Any tone in the synthesizer at 600-kHz offset will mix with the blocker which is at −43 dBm and mix it to the IF stage, where it will interfere with the wanted signal. The blocker can be mixed with noise anywhere in the 200-kHz bandwidth, so a further 53 dB is added to the noise. We note that to be able to detect the wanted signal reliably, as in the previous example, we need the signal to be about 11 dB or so above the mixed-down blocker. Therefore, the mixed-down blocker must be less than −113 dBm. Therefore, the maximum synthesizer noise power at 600-kHz offset is calculated as $-113 + 43 - 53 = -123$ dB lower than the desired oscillating amplitude measured in a 1-Hz bandwidth. This is an illustration of what is known as *phase noise* and will be discussed in more detail in Chapter 8.

References

[1] Papoulis, A., *Probability, Random Variables, and Stochastic Processes,* New York: McGraw-Hill, 1984.

[2] Sze, S. M., *Physics of Semiconductor Devices,* 2nd ed., New York: John Wiley & Sons, 1981.

[3] Gray, P. R., et al., *Analysis and Design of Analog Integrated Circuits,* 4th ed., New York: John Wiley & Sons, 2001.

[4] Stremler, F. G., *Introduction to Communication Systems,* Reading, MA: Addison-Wesley, 1977.

[5] Jordan, E. C., and K. G. Balmain, *Electromagnetic Waves and Radiating Systems,* 2nd ed., Englewood Cliffs, NJ: Prentice Hall, 1968.

[6] Rappaport, T. S., *Wireless Communications,* Upper Saddle River, NJ: Prentice Hall, 1996.

[7] Proakis, J. G., *Digital Communications,* 3rd ed., New York: McGraw-Hill, 1995.

[8] Gonzalez, G., *Microwave Transistor Amplifiers,* Upper Saddle River, NJ: Prentice Hall, 1997.

[9] Wambacq, P., and W. Sansen, *Distortion Analysis of Analog Integrated Circuits,* Norwell, MA: Kluwer, 1998.

[10] Wambacq, P., et al., "High-Frequency Distortion Analysis of Analog Integrated Circuits," *IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing,* Vol. 46, No. 3, March 1999, pp. 335–345.

[11] "Some Notes on Composite Second and Third Order Intermodulation Distortions," *Matrix Technical Notes MTN-108,* Middlesex, NJ: Matrix Test Equipment, http://www.matrixtest.com/Literat/mtn108.htm, accessed Dec. 15, 1998.

[12] "The Relationship of Intercept Points and Composite Distortions," *Matrix Technical Notes MTN-109,* Middlesex, NJ: Matrix Test Equipment, http://www.matrixtest.com/Literat/mtn109.htm, Feb. 18, 1998.

[13] Razavi, B., *RF Microelectronics,* Upper Saddle River, NJ: Prentice Hall, 1998.

[14] Carson, R. S., *Radio Communications Concepts: Analog,* New York: John Wiley & Sons, 1990, Chapter 8.

Selected Bibliography

Fukui, H., *Low Noise Microwave Transistors and Amplifiers,* New York: John Wiley & Sons, 1981.

Larson, L. E., (ed.), *RF and Microwave Circuit Design for Wireless Communications,* Norwood, MA: Artech House, 1997.

Rohde, U. L., J. Whitaker, and A. Bateman, *Communications Receivers: DPS, Software Radios, and Design,* 3rd ed., New York: McGraw-Hill, 2000.

Sklar, B., *Digital Communications: Fundamentals and Applications,* 2nd ed., Englewood Cliffs, NJ: Prentice Hall, 2001.