1 Measurement of wireless transceivers

1.1 Introduction

This book is entitled *Microwave and Wireless Measurement Techniques*, since the objective is to identify and understand measurement theory and practice in wireless systems.

In this book, the concept of a wireless system is applied to the collection of sub-systems that are designed to behave in a particular way and to apply a certain procedure to the signal itself, in order to convert a low-frequency information signal, usually called the baseband signal, to a radio-frequency (RF) signal, and transmit it over the air, and vice versa.

Figure 1.1 presents a typical commercial wireless system architecture. The main blocks are amplifiers, filters, mixers, oscillators, passive components, and domain converters, namely digital to analog and vice versa.

In each of these sub-systems the measurement instruments will be measuring voltages and currents as in any other electrical circuit. In basic terms, what we are measuring are always voltages, like a voltmeter will do for low-frequency signals. The problem here is stated as how we are going to be able to capture a high-frequency signal and identify and quantify its amplitude or phase difference with a reference signal. This is actually the problem throughout the book, and we will start by identifying the main figures of merit that deserve to be measured in each of the identified sub-systems.

In order to do that, we will start by analyzing a general sub-system that can be described by a network. In RF systems it can be a single-port, two-port, or three-port network. The two-port network is the most common.

1.2 Linear two-port networks

1.2.1 Microwave description

A two-port network, Fig. 1.2, is a network in which the terminal voltages and currents relate to each other in a certain way.

The relationships between the voltages and currents of a two-port network can be given by matrix parameters such as Z-parameters, Y-parameters, or ABCD parameters. The reader can find more information in [1, 2].

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

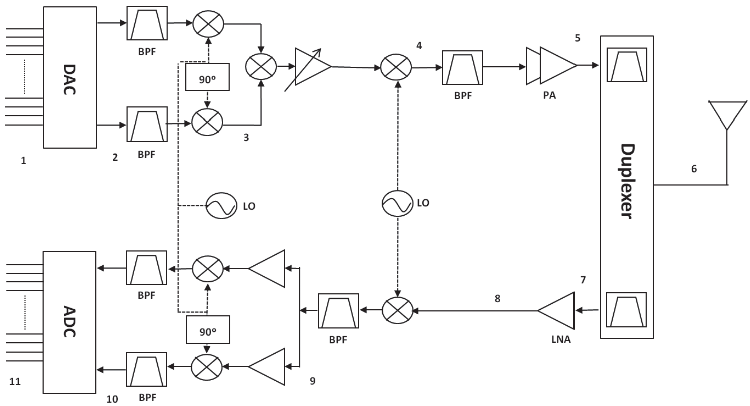

Figure 1.1 A typical wireless system architecture, with a full receiver and transmitter stage.

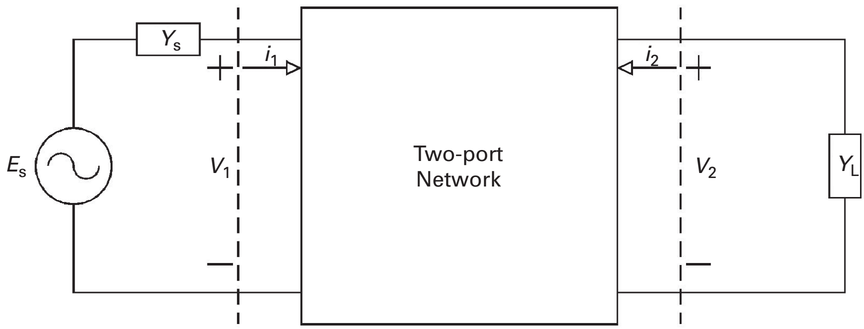

Figure 1.2 A two-port network, presenting the interactions of voltages and currents at its ports.

The objective is always to relate the input and output voltages and currents by using certain relationships. One of these examples using Y-parameters is described by the following equation:

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{1.1}$$

where

$$y_{11} = \left.\frac{i_1}{v_1}\right|_{v_2=0}$$

$$y_{12} = \left.\frac{i_1}{v_2}\right|_{v_1=0}$$

$$y_{21} = \left.\frac{i_2}{v_1}\right|_{v_2=0}$$

$$y_{22} = \left.\frac{i_2}{v_2}\right|_{v_1=0}$$

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

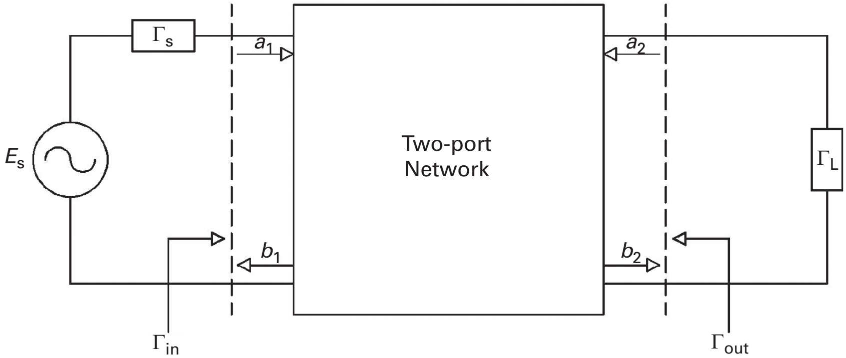

Figure 1.3 Two-port scattering parameters, where the incident and reflected waves can be seen in each port.

As can be seen, these Y-parameters can be easily calculated by considering the other port voltage equal to zero, which means that the other port should be short-circuited. For instance, y_{11} is the ratio of the measured current at port 1 and the applied voltage at port 1 by which port 2 is short-circuited.

Unfortunately, when we are dealing with high-frequency signals, a short circuit is not so simple to realize, and in that case more robust high-frequency parameters should be used.

In that sense some scientists started to think of alternative ways to describe a two-port network, and came up with the idea of using traveling voltage waves [1, 2]. In this case there is an incident traveling voltage wave and a scattered traveling voltage wave at each port, and the network parameters become a description of these traveling voltage waves, Fig. 1.3.

One of the most well-known matrices used to describe these relations consists of the scattering parameters, or S-parameters, by which the scattered traveling voltage waves are related to the incident traveling voltage waves in each port.

In this case each voltage and current in each port will be divided into an incident and a scattered traveling voltage wave, $V^+(x)$ and $V^-(x)$, where the $+$ sign refers to the incident traveling voltage wave and the $-$ sign refers to the reflected traveling voltage wave. The same can be said about the currents, where $I^+(x) = V^+(x)/Z_0$ and $I^-(x) = V^-(x)/Z_0$, Z_0 being the characteristic impedance of the port. The value x now appears since we are dealing with waves that travel across the space, being guided or not, so $V^+(x) = Ae^{-\gamma x}$ [1, 2].

These equations can be further simplified and normalized to be used efficiently:

$$v(x) = \frac{V(x)}{\sqrt{Z_0}}$$
$$i(x) = \sqrt{Z_0}I(x) \tag{1.2}$$

Then each normalized voltage and current can be decomposed into its incident and scattered wave. The incident wave is denoted $a(x)$ and the scattered one $b(x)$:

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

$$\begin{aligned} v(x) &= a(x) + b(x) \\ i(x) &= a(x) - b(x) \end{aligned} \tag{1.3}$$

where

$$\begin{aligned} a(x) &= \frac{V^+(x)}{\sqrt{Z_0}} \\ b(x) &= \frac{V^-(x)}{\sqrt{Z_0}} \end{aligned} \tag{1.4}$$

with

$$\begin{aligned} V &= \sqrt{Z_0}(a + b) \\ I &= \frac{1}{\sqrt{Z_0}}(a - b) \end{aligned}$$

Fortunately, we also know that in a load the reflected wave can be related to the incident wave using its reflection coefficient $\Gamma(x)$:

$$b(x) = \Gamma(x)a(x)$$

or

$$\Gamma(x) = \frac{b(x)}{a(x)} \tag{1.5}$$

In this way it is then possible to calculate and use a new form of matrix parameter to describe these wave relationships in a two-port network, namely the scattering parameters:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{1.6}$$

where

$$S_{ij} = \left.\frac{b_i(x)}{a_j(x)}\right|_{a_k=0 \text{ to } k \neq j} \tag{1.7}$$

As can be deduced from the equations, and in contrast to the Y-parameters, for the calculation of each parameter, the other port should have no reflected wave. This corresponds to matching the other port to the impedance of Z_0. This is easier to achieve at high frequencies than realizing a short circuit or an open circuit, as used for Y- and Z-parameters, respectively.

Moreover, using this type of parameter allows us to immediately calculate a number of important parameters for the wireless sub-system. On looking at the next set of equations, it is possible to identify the input reflection coefficient immediately from S_{11}, or, similarly, the output reflection coefficient from S_{22}:

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

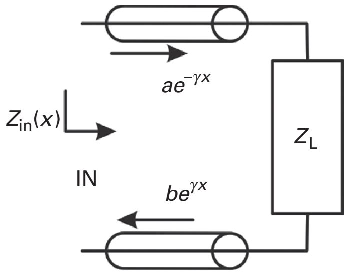

Figure 1.4 Power waves traversing a guided structure.

$$S_{11} = \left.\frac{b_1(x)}{a_1(x)}\right|_{a_2=0} = \frac{Z_1 - Z_0}{Z_1 + Z_0} \tag{1.8}$$

$$S_{22} = \left.\frac{b_2(x)}{a_2(x)}\right|_{a_1=0} = \frac{Z_2 - Z_0}{Z_2 + Z_0} \tag{1.9}$$

The same applies to the other two parameters, S_{21} and S_{12}, which correspond to the transmission coefficient and the reverse transmission coefficient, respectively. The square of their amplitude corresponds to the forward and reverse power gain when the other port is matched.

Note that in the derivation of these parameters it is assumed that the other port is matched. If that is not the case, the values can be somewhat erroneous. For instance, $\Gamma_{in}(x) = S_{11}$ only if the other port is matched or either S_{12} or S_{21} is equal to zero. If this is not the case, the input reflection should be calculated from

$$\Gamma_{in} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 + S_{22}\Gamma_L} \tag{1.10}$$

More information can be found in [1, 2].

With the parameters based on the wave representation that have now been defined, several quantities can be calculated. See Fig. 1.4.

For example, if the objective is to calculate the power at terminal IN, then

$$P = VI^* = aa^* - bb^* = |a|^2 - |b|^2 \tag{1.11}$$

Here $|a|^2$ actually corresponds to the incident power, while $|b|^2$ corresponds to the reflected power.

Important linear figures of merit that are common to most wireless sub-systems can now be defined using the S-parameters.

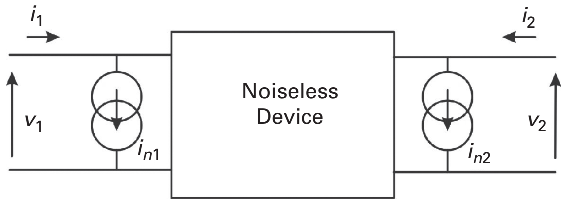

Figure 1.5 A noisy device, Y-parameter representation, including noise sources.

1.2.2 Noise

Another very important aspect to consider when dealing with RF and wireless systems is the amount of introduced noise. Since for RF systems the main goal is actually to achieve a good compromise between power and noise, in order to achieve a good noise-to-power ratio, the study of noise is fundamental. For that reason, let us briefly describe the noise behavior [3] in a two-port network.

A noisy two-port network can be represented by a noiseless two-port network and a noise current source at each port. An admittance representation can be developed.

The voltages and currents in each port can be related to the admittance matrix:

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = [Y] \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} i_{n1} \\ i_{n2} \end{bmatrix} \tag{1.12}$$

(Fig. 1.5). A correlation matrix C_Y can also be defined, as

$$[C_Y] = \begin{bmatrix} \langle i_{n1} i_{n1}^* \rangle & \langle i_{n1} i_{n2}^* \rangle \\ \langle i_{n2} i_{n1}^* \rangle & \langle i_{n2} i_{n2}^* \rangle \end{bmatrix} \tag{1.13}$$

The correlation matrix relates the properties of the noise in each port. For a passive two-port network, one has

$$[C_Y] = 4k_{\mathrm{B}} T \ \Delta f \ \mathrm{Re}(Y) \tag{1.14}$$

where k_{B} is the Boltzmann constant (1.381×10^{-23} J/K), T the temperature (typically 290 K), Δf the bandwidth, and Y the admittance parameter.

Actually these port parameters can also be represented by using scattering parameters. In that case the noisy two-port network is represented by a noiseless two-port network and the noise scattering parameters referenced to a nominal impedance at each port (Fig. 1.6).

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = [S] \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_{\mathrm{n}1} \\ b_{\mathrm{n}2} \end{bmatrix} \tag{1.15}$$

where $b_{\mathrm{n}1}$ and $b_{\mathrm{n}2}$ can be considered noise waves, and they are related using the correlation matrix, C_S. The correlation matrix C_S is defined by

$$[C_S] = \begin{bmatrix} \langle b_{\mathrm{n}1} b_{\mathrm{n}1}^* \rangle & \langle b_{\mathrm{n}1} b_{\mathrm{n}2}^* \rangle \\ \langle b_{\mathrm{n}2} b_{\mathrm{n}1}^* \rangle & \langle b_{\mathrm{n}2} b_{\mathrm{n}2}^* \rangle \end{bmatrix} \tag{1.16}$$

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

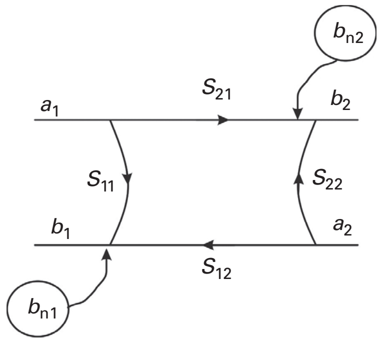

Figure 1.6 A noisy device, S-parameter representation.

and, for a passive two-port network,

$$[C_S] = k_B T \ \Delta f \left((I) - (S)(S)^{T*} \right) \tag{1.17}$$

where (I) is the unit matrix and $(S)^{T*}$ denotes transpose and conjugate.

1.3 Linear FOMs

After having described linear networks, we proceed to explain the corresponding figures of merit (FOMs). We make a distinction between FOMs that are defined on the basis of S-parameters (Section 1.3.1) and those defined on the basis of noise (Section 1.3.2).

1.3.1 Linear network FOMs

1.3.1.1 The voltage standing-wave ratio

The voltage standing-wave ratio (VSWR) is nothing more than the evaluation of the port mismatch. Actually, it is a similar measure of port matching, the ratio of the standing-wave maximum voltage to the standing-wave minimum voltage. Figure 1.7 shows different standing-wave patterns depending on the load.

In this sense it therefore relates the magnitude of the voltage reflection coefficient and hence the magnitude of either S_{11} for the input port or S_{22} for the output port.

The VSWR for the input port is given by

$$\mathrm{VSWR_{in}} = \frac{1 + |S_{11}|}{1 - |S_{11}|} \tag{1.18}$$

and that for the output port is given by

$$\mathrm{VSWR_{out}} = \frac{1 + |S_{22}|}{1 - |S_{22}|} \tag{1.19}$$

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

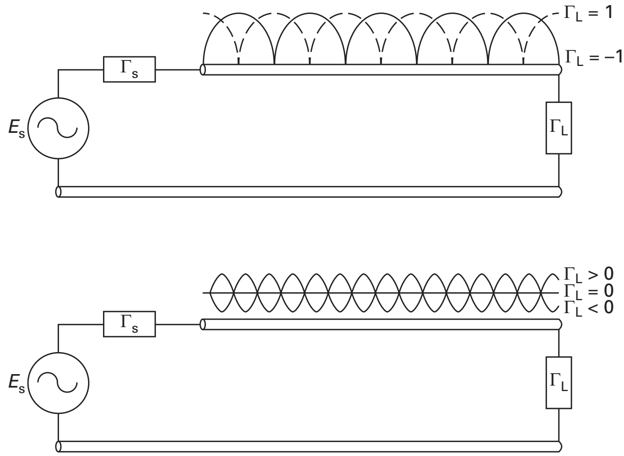

Figure 1.7 The VSWR and standing-wave representation. The standing wave can be seen for different values of the VSWR.

1.3.1.2 Return loss

Other important parameters are the input and output return losses. The input return loss (RL_{in}) is a scalar measure of how close the actual input impedance of the network is to the nominal system impedance value, and is given by

$$RL_{in} = \left|20\log_{10}\left|S_{11}\right|\right| \text{ dB} \tag{1.20}$$

It should be noticed that this value is valid only for a single-port network, or, in a two-port network, it is valid only if port 2 is matched; if not, S_{11} should be exchanged for the input reflection coefficient as presented in Eq. (1.10). As can be seen from its definition, the return loss is a positive scalar quantity.

The output return loss (RL_{out}) is similar to the input return loss, but applied to the output port (port 2). It is given by

$$RL_{out} = \left|20\log_{10}\left|S_{22}\right|\right| \text{ dB} \tag{1.21}$$

1.3.1.3 Gain/insertion loss

Since S_{11} and S_{22} have the meaning of reflection coefficients, their values are always smaller than or equal to unity. The exception is the S_{11} of oscillators, which is larger than unity, because the RF power returned is larger than the RF power sent into the oscillator port.

The S_{21} of a linear two-port network can have values either smaller or larger than unity. In the case of passive circuits, S_{21} has the meaning of loss, and is thus restricted to values smaller than or equal to unity. This loss is usually called the insertion loss. In the case of active circuits, there is usually gain, or in other words S_{21} is larger than unity. In the case of passive circuits, S_{12} is equal to S_{21} because passive circuits are reciprocal. The only exception is the case of ferrites. In the case of active circuits, S_{12} is different from S_{21} and usually much smaller than unity, since it represents feedback, which is often avoided by design due to the Miller effect. The gain or loss is typically expressed in decibels:

$$\text{gain/insertion loss} = \left|20\log_{10}\left|S_{21}\right|\right| \text{ dB} \tag{1.22}$$

1.3.2 Noise FOMs

1.3.2.1 The noise factor

The previous results actually lead us to a very important and key point regarding noisy devices, that is, the FOM called the noise factor (NF), which characterizes the degradation of the signal-to-noise ratio (SNR) by the device itself.

The noise factor is defined as follows.

DEFINITION 1.1 *The noise factor* (F) *of a circuit is the ratio of the signal-to-noise ratio at the input of the circuit to the signal-to-noise ratio at the output of the circuit:*

$$F = \frac{S_\mathrm{I}/N_\mathrm{I}}{S_\mathrm{O}/N_\mathrm{O}} \tag{1.23}$$

where

S_I *is the power of the signal transmitted from the source to the input of the two-port network*

S_O *is the power of the signal transmitted from the output of the two-port network to the load*

N_I *is the power of the noise transmitted from the source impedance* Z_S *at temperature* $T_0 = 290\ K$ *to the input of the two-port network*

N_O *is the power of the noise transmitted from the output of the two-port network to the load*

The noise factor can be expressed as

$$F = \frac{N_\mathrm{ad} + G_\mathrm{A} N_\mathrm{aI}}{G_\mathrm{A} N_\mathrm{aI}} \tag{1.24}$$

where G_A is the available power gain of the two-port network (for its definition, see Section 1.8), N_ad is the additional available noise power generated by the two-port network, and N_aI is the available noise power generated by the source impedance:

$$N_\mathrm{aI} = 4k_\mathrm{B} T_0\, \Delta f \tag{1.25}$$

As can be seen from Eq. (1.24), F is always greater than unity, and it does not depend upon the load Z_L. It depends exclusively upon the source impedance Z_S.

Using reference [3], the noise factor can also be related to the S-parameters by:

$$F = F_\mathrm{min} + 4\frac{R_\mathrm{N}}{Z_0}\frac{|\Gamma_\mathrm{OPT} - \Gamma_\mathrm{s}|^2}{(1-|\Gamma_\mathrm{s}|^2)|1+\Gamma_\mathrm{OPT}|^2} \tag{1.26}$$

where F_min is the minimum noise factor, R_N is called the noise resistance, Γ_OPT is the optimum source reflection coefficient for which the noise factor is minimum.

This formulation can also be made in terms of Y-parameters, and can be expressed as a function of the source admittance Y_S:

$$F = F_{min} + \frac{R_N}{\text{Re}(Y_S)} |Y_S - Y_{OPT}|^2 \tag{1.27}$$

where Y_{OPT} is the optimum source admittance for which the noise factor is minimum.

The terms F_{MIN}, R_N, and Γ_{OPT} (or Y_{OPT}) constitute the four noise parameters of the two-port network. They can be related to the correlation matrices very easily [3].

The noise figure (NF) is simply the logarithmic version of the noise factor, F.

1.3.2.2 Cascade of noisy two-port components

If we cascade two noisy devices with noise factors F_1 and F_2, and with available power gains G_{A1} and G_{A2}, with a source impedance at temperature $T_0 = 290$ K, the additional available noise powers are

$$\begin{aligned} N_{ad1} &= (F_1 - 1)G_{A1}k_B T_0 \, \Delta f \\ N_{ad2} &= (F_2 - 1)G_{A2}k_B T_0 \, \Delta f \end{aligned} \tag{1.28}$$

The available noise power at the output of the second two-port network is

$$N_{aO2} = k_B T_0 \, \Delta f \; G_{A1} G_{A2} + N_{ad1} G_{A2} + N_{ad2} \tag{1.29}$$

The total noise factor is thus

$$F = \frac{N_{aO2}}{k_B T_0 \, \Delta f \; G_{A1} G_{A2}} \tag{1.30}$$

This finally leads to the well-known noise Friis formula,

$$F = F_1 + \frac{F_2 - 1}{G_{A1}} \tag{1.31}$$

In this expression the gain is actually the available power gain of the first two-port network, which depends on the output impedance of the first network. F_1 depends on the source impedance, and F_2 depends on the output impedance of the first two-port network.

The general Friis formula is

$$F = F_1 + \frac{F_2 - 1}{G_{A1}} + \frac{F_3 - 1}{G_{A1} G_{A2}} + \cdots + \frac{F_N - 1}{G_{A1} G_{A2} \ldots G_{AN-1}}$$

1.4 Nonlinear two-port networks

In order to better understand nonlinear distortion effects, let us start by explaining the fundamental properties of nonlinear systems. Since a nonlinear system is defined as a system that is not linear, we will start by explaining the fundamentals of linear systems.

Linear systems are systems that obey superposition. This means that they are systems whose output to a signal composed by the sum of elementary signals can be given as the sum of the outputs to these elementary signals when taken individually.

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

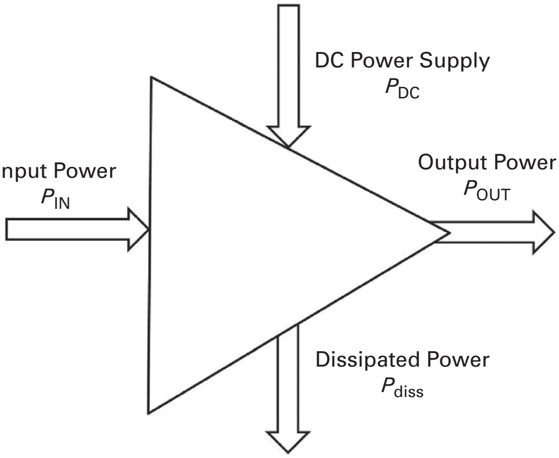

Figure 1.8 The power balance of a nonlinear system.

This can be stated as

$$y(t) = S_L[x(t)] = k_1 y_1(t) + k_2 y_2(t) \tag{1.32}$$

where $x(t) = k_1 x_1(t) + k_2 x_2(t)$, $y_1(t) = S_L[x_1(t)]$, and $y_2(t) = S_L[x_2(t)]$. Any system that does not obey Eq. (1.32) is said to be a nonlinear system. Actually, this violation of the superposition theorem is the typical rule rather than being the exception. For the remainder of this section, we assume the two-port network to be an amplifier.

To better understand this mechanism, consider the general active system of Fig. 1.8, where P_{IN} and P_{OUT} are the input power entering the amplifier and the output power going to the load, respectively; P_{DC} is the DC power delivered to the amplifier by the power supply; and P_{diss} is the total amount of power lost, by being dissipated in the form of heat or in any other form [4].

Using the definition of operating power gain, $G = P_{OUT}/P_{IN}$ (see also Section 1.8.1), and considering that the fundamental energy-conservation principle requires that $P_L + P_{diss} = P_{IN} + P_{DC}$, we can write the operating power gain as

$$G = 1 + \frac{P_{DC} - P_{diss}}{P_{IN}} \tag{1.33}$$

From this equation we can see that, since P_{diss} has a theoretical minimum of zero and P_{DC} is limited by the finite available power from the supply, the amplifier cannot maintain a constant power gain for increasing input power.

This will lead the amplifier to start to deviate from linearity at a certain input power, and thus start to become nonlinear. Figure 1.9 presents this result by sketching the operating power gain of an amplifier versus the input power rise.

1.4.1 Nonlinear generation

In order to evaluate how the inherent nonlinear phenomena can affect amplifiers, let us consider a simple analysis, where we will compare the responses of simple linear and nonlinear systems to typical inputs encountered in wireless technology.

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

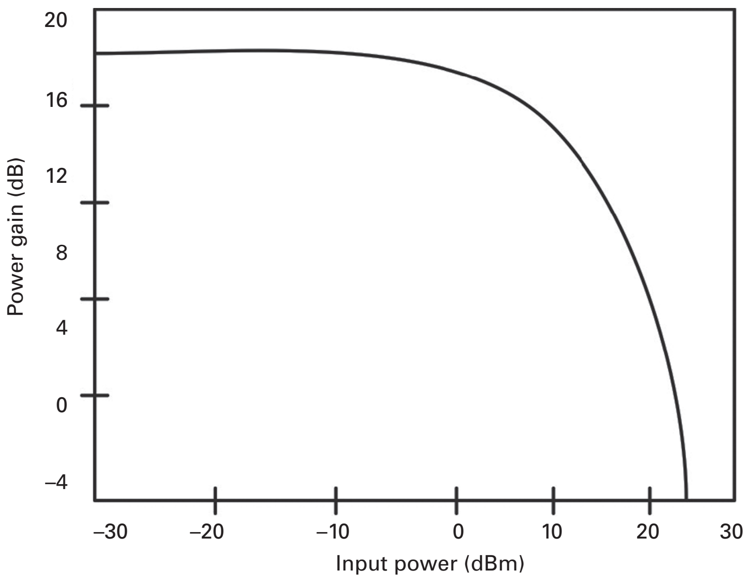

Figure 1.9 The nonlinear behavior of the variation of power gain versus input power.

In wireless systems, which are mainly based on radio-frequency communications, the stimulus inputs are usually sinusoids, with these being amplitude- and phase-modulated by some baseband information signal. Therefore the input signal of these systems can be written as

$$x(t) = A(t)\cos[\omega_c t + \theta(t)] \tag{1.34}$$

where $A(t)$ is the time-dependent amplitude, ω_c is the carrier pulsation, and $\phi(t)$ is the modulated phase.

The simplest form of nonlinear behavior that allows us to mathematically describe the response is that in which the nonlinearity is represented by a polynomial [4]:

$$y_{NL}(t) = a_1 x(t - \tau_1) + a_2 x(t - \tau_2)^2 + a_3 x(t - \tau_3)^3 + \cdots \tag{1.35}$$

In this case the polynomial was truncated to the third order to simplify the calculations, but a higher-degree polynomial can obviously be used. Actually, this type of approximation is the first solution to the more complex nonlinear behavior of an amplifier, and, if the a_s coefficients and τ_s are changed, then many systems can be modeled in this way.

The system actually behaves as a linear one, if $x(t) \gg x(t)^2, x(t)^3$, and $y_{NL}(t) \approx y_L(t) = S_L[x(t)] = a_1 x(t - \tau_1)$.

In this case, the linear response will be

$$y_L(t) = a_1 A(t - \tau_1)\cos[\omega_c t + \theta(t - \tau_1) - \phi_1] \tag{1.36}$$

and the overall nonlinear response can be written as

$$
\begin{aligned}
y_{\mathrm{NL}}(t) = {} & a_1 A(t-\tau_1)\cos[\omega_{\mathrm{c}}t + \theta(t-\tau_1) - \phi_1] \\
& + a_2 A(t-\tau_2)^2 \cos[\omega_{\mathrm{c}}t + \theta(t-\tau_2) - \phi_2]^2 \\
& + a_3 A(t-\tau_3)^3 \cos[\omega_{\mathrm{c}}t + \theta(t-\tau_3) - \phi_3]^3
\end{aligned} \tag{1.37}
$$

Using some trigonometric relations, as presented in [4], the equations can be written as

$$
\begin{aligned}
y_{\mathrm{NL}}(t) = {} & a_1 A(t-\tau_1)\cos[\omega_{\mathrm{c}}t + \theta(t-\tau_1) - \phi_1] \\
& + \frac{1}{2} a_2 A(t-\tau_2)^2 \\
& + \frac{1}{2} a_2 A(t-\tau_2)^2 \cos[2\omega_{\mathrm{c}}t + 2\theta(t-\tau_2) - 2\phi_2] \\
& + \frac{3}{4} a_3 A(t-\tau_3)^3 \cos[\omega_{\mathrm{c}}t + \theta(t-\tau_3) - \phi_3] \\
& + \frac{1}{4} a_3 A(t-\tau_3)^3 \cos[3\omega_{\mathrm{c}}t + 3\theta(t-\tau_3) - 3\phi_3]
\end{aligned} \tag{1.38}
$$

where $\phi_1 = \omega_{\mathrm{c}}\tau_1$, $\phi_2 = \omega_{\mathrm{c}}\tau_2$, and $\phi_3 = \omega_{\mathrm{c}}\tau_3$.

In typical wireless communication systems, the variation of the modulated signals is usually slow compared with that of the RF carrier, and thus, if the system does not exhibit memory effects, one can write

$$
y_{\mathrm{L}}(t) = a_1 A(t)\cos[\omega_{\mathrm{c}}t + \theta(t) - \phi_1] \tag{1.39}
$$

and

$$
\begin{aligned}
y_{\mathrm{NL}}(t) = {} & a_1 A(t)\cos[\omega_{\mathrm{c}}t + \theta(t) - \phi_1] \\
& + \frac{1}{2} a_2 A(t)^2 \\
& + \frac{1}{2} a_2 A(t)^2 \cos[2\omega_{\mathrm{c}}t + 2\theta(t) - 2\phi_2] \\
& + \frac{3}{4} a_3 A(t)^3 \cos[\omega_{\mathrm{c}}t + \theta(t) - \phi_3] \\
& + \frac{1}{4} a_3 A(t)^3 \cos[3\omega_{\mathrm{c}}t + 3\theta(t) - 3\phi_3]
\end{aligned} \tag{1.40}
$$

From Eqs. (1.39) and (1.40) it is clear that the linear and nonlinear responses are significantly different. For instance the number of terms in the nonlinear formulation is quite high compared with the number in the linear formulation.

Moreover, the output of the linear response is a version of the input signal, with the same spectral contents, but with a variation in amplitude and phase as compared with the input, whereas the nonlinear response consists of a panoply of other spectral components, usually called spectral regrowth. Actually, this is one of the properties of nonlinear systems, namely that, in contrast to a linear system, which can only introduce quantitative changes to the signal spectra, nonlinear systems can qualitatively modify spectra, insofar as they eliminate certain spectral components and generate new ones.

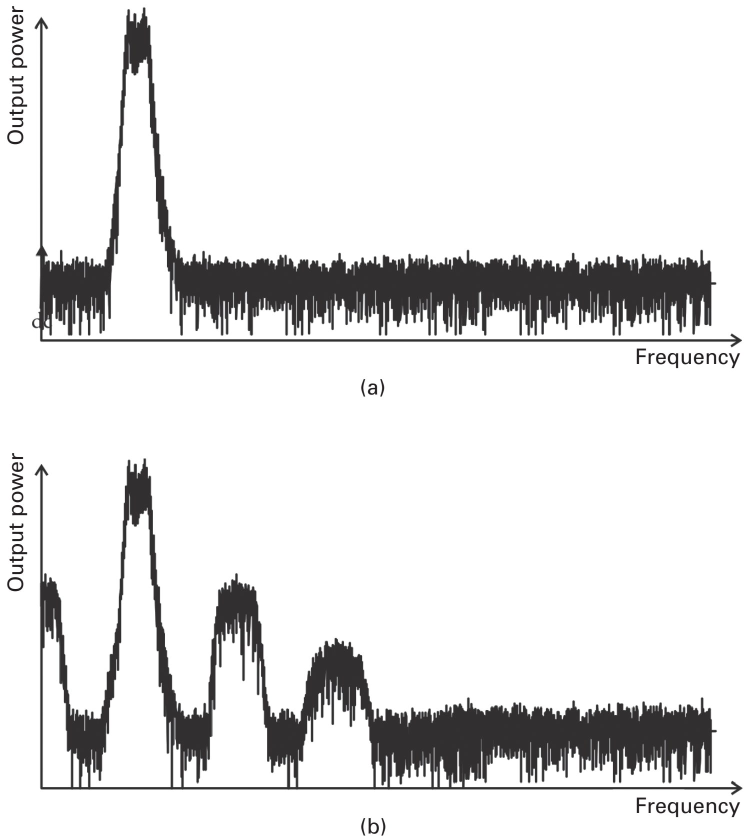

Figure 1.10 Signal probing throughout the wireless system path: (a) the analog signal at point 4; and (b) the signal at point 5, where the generation of nonlinear distortion is visible.

One example of a typical linear component is a filter, since the output of a filter can in principle change exclusively the amplitude and phase of the input signal, while an example of a nonlinearity is a frequency multiplier, where the output spectrum is completely different from the input spectrum.

In a typical wireless system such as the one presented in Fig. 1.1, the most important source of nonlinear distortion is the power amplifier (PA), but all the components can behave nonlinearly, depending on the input signal excitation.

1.4.2 Nonlinear impact in wireless systems

In the previous section the nonlinear generation mechanism was explained using a simple polynomial. In this section we will probe the signal throughout the system presented in Fig. 1.1.

Figure 1.10 presents the spectral content in each of the stages of the wireless system, and allows one to see how the signal changes on traversing the communication path.

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

As can be seen from the images, the starting signal is nothing more than a bit stream arriving at our digital-to-analog converter, point 1 in Fig. 1.1. Then this signal is further converted to analog and filtered out (point 2), up-converted to an IF channel (point 3), amplified and further up-converted to RF (point 4), amplified again (point 5), and transmitted over the air (point 6).

The signal then traverses the air interface and, at the receiver, it is first filtered out (point 7), then amplified using a low-noise amplifier (point 8), and then down-converted to IF (point 9), and to baseband again (point 10), and reconverted to a digital version (point 11).

Since in this section we are looking mainly at the nonlinear behavior of the RF signal, let us concentrate on points 4 and 5. In this case the input signal has a certain spectral shape, as can be seen in Fig. 1.10(a), and, after the nonlinear behavior of the PA, it appears completely different at point 5, where several clusters of spectra appear, Fig. 1.10(b).

The first cluster is centered at DC and, in practical systems, it consists of two forms of distortion, namely the DC value itself and a cluster of very-low-frequency spectral components centered at DC. The DC value distortion manifests itself as a shift in bias from the quiescent point (defined as the bias point measured without any excitation) to the actual bias point measured when the system is driven at its rated input excitation power.

If we look back at Eq. (1.40), we can understand that the DC component comes from all possible mixing, beat, or nonlinear distortion products of the form $\cos(\omega_i t)\ \cos(\omega_j t)$, whose frequency mixing appears at $\omega_x = \omega_i - \omega_j$, where $\omega_i = \omega_j$.

The low-frequency cluster near DC constitutes a distorted version of the amplitude-modulating information, $A(t)$, as if the input signal had been demodulated. This cluster is, therefore, called the baseband component of the output. In spectral terms their frequency lines are also generated from mixing products at $\omega_x = \omega_i - \omega_j$, but now where $\omega_i \neq \omega_j$.

From Fig. 1.10(b) it is clear that there are some other clusters appearing at $2\omega_c$ and $3\omega_c$. These are the well-known second- and third-harmonic components, usually called the harmonic distortion. Actually, they are high-frequency replicas of the modulated signal.

The cluster appearing at $2\omega_c$ is again generated from all possible mixing products of the form $\cos(\omega_i t)\ \cos(\omega_j t)$, but now the outputs are located at $\omega_x = \omega_i + \omega_j$, where $\omega_i = \omega_j$ $(\omega_x = 2\omega_i = 2\omega_j)$ or $\omega_i \neq \omega_j$.

The third-harmonic cluster appears from all possible mixing products of the form $\cos(\omega_i t)\cos(\omega_j t)\cos(\omega_k t)$, whose outputs are located at $\omega_x = \omega_i + \omega_j + \omega_k$, where $\omega_i = \omega_j = \omega_k$ $(\omega_x = 3\omega_i = 3\omega_j = 3\omega_k)$ or $\omega_i = \omega_j \neq \omega_k$ $(\omega_x = 2\omega_i + \omega_k = 2\omega_j + \omega_k)$ or even $\omega_i \neq \omega_j \neq \omega_k$.

The last-mentioned cluster is the one appearing around ω_c. In this scenario the nonlinear distortion appears near the spectral components of the input signal, but is also exactly coincident with them, and thus is indistinguishable from them.

Unfortunately, and in contrast to the baseband or harmonic distortion, which falls on out-of-band spectral components, and thus could be simply eliminated by bandpass filtering, some of these new in-band distortion components are unaffected by any linear operator that, naturally, must preserve the fundamental components. Thus, they constitute the most important form of distortion in bandpass microwave and wireless subsystems. Since this is actually the most important form of nonlinear distortion in narrowband systems, it is sometimes just called "distortion."

In order to clearly understand and identify the in-band-distortion spectral components, they must be first separated in to the spectral lines that fall exactly over the original ones and the lines that constitute distortion sidebands. In wireless systems, the former are known as *co-channel distortion* and the latter as *adjacent-channel distortion*.

Looking back at our formulation Eq. (1.40), all in-band-distortion products share the form of $\cos(\omega_i t)\cos(\omega_j t)\cos(\omega_k t)$, which is similar to the ones appearing at the third harmonic, but now the spectral outputs are located at $\omega_x = \omega_i + \omega_j - \omega_k$. In this case, and despite the fact that both co-channel and adjacent-channel distortion can be generated by mixing products obeying $\omega_i = \omega_j \neq \omega_k$ ($\omega_x = 2\omega_i - \omega_k = 2\omega_j - \omega_k$) or $\omega_i \neq \omega_j \neq \omega_k$, only the mixing terms obeying $\omega_i = \omega_j = \omega_k$ ($\omega_x = \omega_i$) or $\omega_i \neq \omega_j = \omega_k$ ($\omega_x = \omega_i$) fall on top of the co-channel distortion.

1.5 Nonlinear FOMs

Let us now try to identify how we can account for nonlinearity in wireless two-port networks. To this end we will use different signal excitations, since those will reveal different aspects of the nonlinear behavior. We will start first with a single-tone excitation, and then proceed to the best-known signal excitation for nonlinear distortion, namely two-tone excitation, and then the multi-sine excitation figures of merit will also be addressed. Finally, a real modulated wireless signal will be used to define the most important FOMs in wireless systems.

1.5.1 Nonlinear single-tone FOMs

We start by considering that $x(t)$ in Eq. (1.35) is a single sinusoid, $x(t) = A\cos(\omega_c t)$. The output signal is described by Eq. (1.40), and, if we consider that the input signal does not have a phase delay, it can be further simplified to

$$\begin{aligned} y_{NL}(t) = {} & a_1 A\cos[\omega_c t - \phi_1] \\ & + \frac{1}{2}a_2 A^2 \\ & + \frac{1}{2}a_2 A^2 \cos[2\omega_c t - 2\phi_2] \\ & + \frac{3}{4}a_3 A^3 \cos[\omega_c t - \phi_3] \\ & + \frac{1}{4}a_3 A^3 \cos[3\omega_c t - 3\phi_3] \end{aligned} \tag{1.41}$$

In this case the output consists of single-tone spectral components appearing at DC, in the same frequency component as the input signal, and at the second and third harmonics.

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

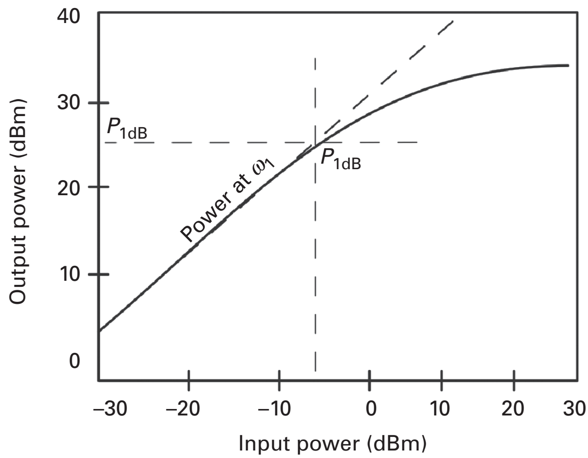

Figure 1.11 AM–AM curves, where the output power is plotted versus the input power increase. The 1-dB-compression point is also visible in the image.

Actually, the output amplitude and phase variation versus input drive manifest themselves as if the nonlinear device could convert input amplitude variations into output amplitude and phase changes or, in other words, as if it could transform possible amplitude modulation (AM) associated with its input into output amplitude modulation (AM–AM conversion) or phase modulation (AM–PM conversion).

AM–AM conversion is particularly important in systems that are based on amplitude modulation, while AM–PM has its major impact in modern wireless telecommunication systems that rely on phase-modulation formats.

If a careful analysis is done at the harmonics, we can also calculate the ratio of the integrated power of all the harmonics to the measured power at the fundamental, a figure of merit named total harmonic distortion (THD).

1.5.1.1 AM–AM

The AM–AM figure of merit describes the relationship between the output amplitude and the input amplitude at the fundamental frequency [4].

Figure 1.11 presents the AM–AM characteristic, where the power of each of the fundamental spectral components is plotted versus its input counterpart. As can be seen, it characterizes the gain compression or expansion of a nonlinear device versus the input drive level.

One of the most important FOMs that can be extracted from this type of characterization is called the 1-dB-compression point, P_{1dB}.

1.5.1.2 The 1-dB-compression point (P_{1dB})

DEFINITION 1.2 *The 1-dB-compression point* (P_{1dB}) *is defined as the output power level at which the signal output is compressed by 1 dB, compared with the output power level that would be obtained by simply extrapolating the linear system's small-signal characteristic.*

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

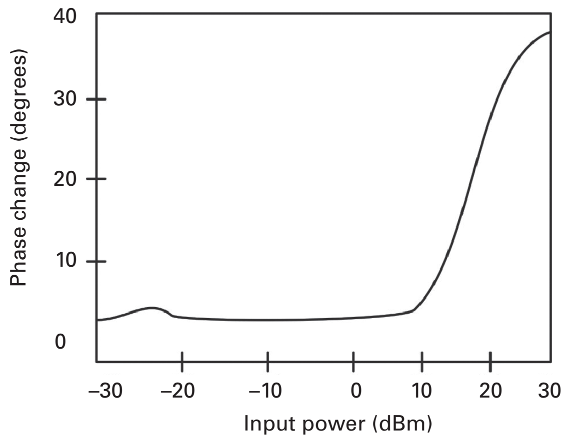

Figure 1.12 AM–PM curves, where the phase delay of the output signal is visible when the input power is varied.

Thus, the $P_{1\mathrm{dB}}$ FOM also corresponds to a 1-dB gain deviation from its small-signal value, as depicted in Fig. 1.9 and Fig. 1.11.

1.5.1.3 AM–PM

Since the co-channel nonlinear distortion actually falls on top of the input signal spectra, Eq. (1.41), the resulting output component at that frequency will be the addition of two vectors: the linear output signal, plus a version of the nonlinear distortion. So vector addition can also determine a phase variation of the resultant output, when the input level varies, as shown in Fig. 1.12.

The change of the output signal phase, $\phi(\omega, A_i)$, with increasing input power is the AM–PM characteristic and may be expressed as a certain phase deviation, in degrees/dB, at a predetermined input power.

1.5.1.4 Total harmonic distortion

The final FOM in connection with single-tone excitation is one that accounts for the higher-order harmonics, and it is called total harmonic distortion (THD).

DEFINITION 1.3 *The total harmonic distortion (THD) is defined as the ratio between the square roots of the total harmonic output power and the output power at the fundamental frequency.*

Therefore the THD can be expressed as

$$\mathrm{THD} = \frac{\sqrt{1/T \int_0^T \left[\sum_{r=2}^{\infty} A_{0_r}(\omega, A_i)\cos[r\omega t + \theta_{0_r}(\omega, A_i)]\right]^2 dt}}{\sqrt{1/T \int_0^T \left[A_{0_1}(\omega, A_i)\cos[\omega t + \theta_{0_1}(\omega, A_i)]\right]^2 dt}} \tag{1.42}$$

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

and, for the polynomial case, we will have

$$\mathrm{THD} = \frac{\sqrt{\frac{1}{8}a_2^2 A_i^4 + (1/32)a_3^2 A_i^6 + \cdots}}{\sqrt{a_1^2 A_i^2/2}} \tag{1.43}$$

1.5.2 Nonlinear two-tone FOMs

A single-tone signal unfortunately is just a first approach to the characterization of a nonlinear two-port network. Actually, as was seen previously, the single-tone signal can be used to evaluate the gain compression and expansion, and harmonic generation, but no information is given about the bandwidth of the signal, or about the distortion appearing in-band.

In order to get a better insight into these in-band-distortion products, RF engineers started to use so-called two-tone excitation signals.

A two-tone signal is composed of a summation of two sinusoidal signals,

$$x(t) = A_1 \cos(\omega_1 t) + A_2 \cos(\omega_2 t) \tag{1.44}$$

Since the input is now composed of two different carriers, many more mixing products will be generated when it traverses the polynomial presented in Eq. (1.35). Therefore, it is convenient to count all of them in a systematic manner. Hence the sine representation will be substituted by its Euler expansion representation:

$$\begin{aligned} x(t) &= A_1 \cos(\omega_1 t) + A_2 \cos(\omega_2 t) \\ &= A_1 \frac{e^{j[\omega_1 t]} + e^{-j[\omega_1 t]}}{2} + A_2 \frac{e^{j[\omega_2 t]} + e^{-j[\omega_2 t]}}{2} \end{aligned} \tag{1.45}$$

This type of formulation actually allows us to calculate all the mixing products arising from the polynomial calculations, since the input can now be viewed as the sum of four terms, each one involving a different frequency. That is, we are assuming that each sinusoidal function involves a positive- and a negative frequency component (i.e., the corresponding positive and negative sides of the Fourier spectrum), so that any combination of tones can be represented as

$$\begin{aligned} x(t) &= \sum_{r=1}^{R} A_r \cos(\omega_r t) \\ &= \frac{1}{2} \sum_{r=-R;r\neq 0}^{R} A_r e^{j\omega_r t} \end{aligned} \tag{1.46}$$

where $r \neq 0$, and $A_r = A_{-r}^*$ for real signals.

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

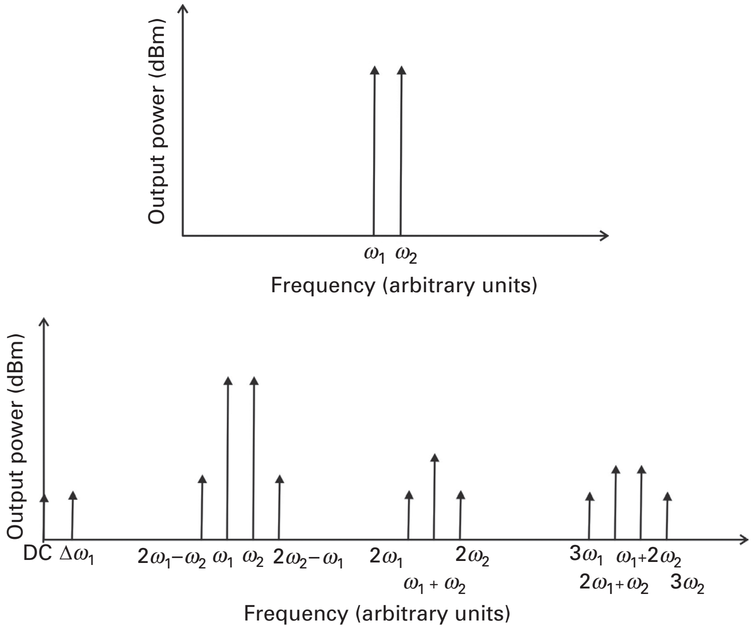

Figure 1.13 The spectrum arrangement of a two-tone signal traversing a nonlinearity, where different spectrum clusters can be seen.

The output of the polynomial model for this type of formulation is now much simpler to develop, and for each mixing value we will have

$$
\begin{aligned}
y_{\mathrm{NL}n}(t) &= \frac{1}{2^n} a_n \left[\sum_{r=-R}^{R} A_r e^{j\omega_r t} \right]^n \\
&= \frac{1}{2^n} a_n \sum_{r_1=-R}^{R} \cdots \sum_{r_n=-R}^{R} A_{r_1} \cdots A_{r_n} e^{j(\omega_{r_1} + \cdots + \omega_{r_n})t}
\end{aligned} \tag{1.47}
$$

The frequency components arising from this type of mixing are all possible combinations of the input ω_r:

$$
\begin{aligned}
\omega_n &= \omega_{r_1} + \cdots + \omega_{r_n} \\
&= m_{-R}\omega_{-R} + \cdots + m_{-1}\omega_{-1} + m_1\omega_1 + \cdots + m_R\omega_R
\end{aligned} \tag{1.48}
$$

where the vector $[m_{-R} \ldots m_{-1} m_1 \ldots m_R]$ is the nth order mixing vector, which must satisfy

$$
\begin{aligned}
\sum_{r=-R}^{R} m_r &= m_{-R} + \cdots + m_{-1} + m_1 + \cdots + m_R \\
&= n
\end{aligned} \tag{1.49}
$$

In the case of a two-tone signal (Fig. 1.13), the first-order mixing, arising from the linear response (coefficient a_1 in the polynomial), will be

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

$$-\omega_2, -\omega_1, \omega_1, \omega_2 \tag{1.50}$$

Adding then the second order (coefficient a_2), we will have

$$-2\omega_2, -\omega_2 - \omega_1, -2\omega_1, \omega_1 - \omega_2, \text{DC}, \omega_2 - \omega_1, 2\omega_1, \omega_1 + \omega_2, 2\omega_2 \tag{1.51}$$

and, for the third-order coefficient a_3,

$$\begin{aligned}
&-3\omega_2, -2\omega_2 - \omega_1, -\omega_2 - 2\omega_1, -3\omega_1,\\
&-2\omega_2 + \omega_1, -\omega_2, -\omega_1, -2\omega_1 + \omega_2,\\
&\ 2\omega_1 - \omega_2, \omega_1, \omega_2, 2\omega_2 - \omega_1,\\
&\ 3\omega_1, 2\omega_1 + \omega_2, \omega_1 + 2\omega_2, 3\omega_2
\end{aligned} \tag{1.52}$$

Obviously there are several ways in which these mixing products can be gathered, and the reader can find the calculations in [4, 5]. If the Euler coefficients corresponding to each mixing product are now added, we obtain the following expression for the output of our nonlinearity when excited by a two-tone signal:

$$\begin{aligned}
y_{\text{NL}}(t) = {} & a_1 A_1 \cos(\omega_1 t - \phi_{1_{10}}) + a_1 A_2 \cos(\omega_2 t - \phi_{1_{10}})\\
& + \frac{1}{2} a_2 (A_1^2 + A_2^2)\\
& + a_2 A_1 A_2 \cos[(\omega_2 - \omega_1)t - \phi_{2_{-11}}]\\
& + a_2 A_1 A_2 \cos[(\omega_1 + \omega_2)t - \phi_{2_{11}}]\\
& + \frac{1}{2} a_2 A_1^2 \cos(2\omega_1 t - \phi_{2_{20}}) + \frac{1}{2} a_2 A_2^2 \cos(2\omega_2 t - \phi_{2_{02}})\\
& + \frac{3}{4} a_3 A_1^2 A_2 \cos[(2\omega_1 - \omega_2)t - \phi_{3_{2-1}}]\\
& + \left(\frac{3}{4} a_3 A_1^3 + \frac{6}{4} a_3 A_1 A_2^2\right) \cos(\omega_1 t - \phi_{3_{10}})\\
& + \left(\frac{3}{4} a_3 A_2^3 + \frac{6}{4} a_3 A_2 A_1^2\right) \cos(\omega_2 t - \phi_{3_{01}})\\
& + \frac{3}{4} a_3 A_1 A_2^2 \cos[(2\omega_2 - \omega_1)t - \phi_{3_{-12}}]\\
& + \frac{1}{4} a_3 A_1^3 \cos(3\omega_1 t - \phi_{3_{30}})\\
& + \frac{3}{4} a_3 A_1^2 A_2 \cos[(2\omega_1 + \omega_2)t - \phi_{3_{21}}]\\
& + \frac{3}{4} a_3 A_1 A_2^2 \cos[(\omega_1 + 2\omega_2)t - \phi_{3_{12}}]\\
& + \frac{1}{4} a_3 A_2^3 \cos(3\omega_2 t - \phi_{3_{03}})
\end{aligned} \tag{1.53}$$

where $\phi_{1_{10}} = \omega_1 \tau_1$, $\phi_{1_{01}} = \omega_2 \tau_1$, $\phi_{2_{-11}} = \omega_2 \tau_2 - \omega_1 \tau_2$, $\phi_{2_{20}} = 2\omega_1 \tau_2$, $\phi_{2_{11}} = \omega_1 \tau_2 + \omega_2 \tau_2$, $\phi_{2_{02}} = 2\omega_2 \tau_2$, $\phi_{3_{2-1}} = 2\omega_1 \tau_3 + \omega_2 \tau_3$, $\phi_{3_{10}} = \omega_1 \tau_3$, $\phi_{3_{01}} = \omega_2 \tau_3$, $\phi_{3_{-12}} = 2\omega_2 \tau_3 - \omega_1 \tau_3$, $\phi_{3_{30}} = 3\omega_1 \tau_3$, $\phi_{3_{21}} = 2\omega_1 \tau_3 + \omega_2 \tau_3$, $\phi_{3_{12}} = \omega_1 \tau_3 + 2\omega_2 \tau_3$, and $\phi_{3_{03}} = 3\omega_2 \tau_3$.

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

If we look exclusively at the in-band distortion, the output components will be

$$\begin{aligned} y_{\mathrm{NL}_{\text{in-band}}}(t) = {} & a_1 A_1 \cos\left(\omega_1 t - \phi_{1_{10}}\right) + a_1 A_2 \cos\left(\omega_2 t - \phi_{1_{10}}\right) \\ & + \frac{3}{4} a_3 A_1^2 A_2 \cos\left[(2\omega_1 - \omega_2)t - \phi_{3_{2-1}}\right] \\ & + \left(\frac{3}{4} a_3 A_1^3 + \frac{6}{4} a_3 A_1 A_2^2\right) \cos\left(\omega_1 t - \phi_{3_{10}}\right) \\ & + \left(\frac{3}{4} a_3 A_2^3 + \frac{6}{4} a_3 A_2 A_1^2\right) \cos\left(\omega_2 t - \phi_{3_{01}}\right) \\ & + \frac{3}{4} a_3 A_1 A_2^2 \cos\left[(2\omega_2 - \omega_1)t - \phi_{3_{-12}}\right] \end{aligned} \tag{1.54}$$

From this equation it is clear that the in-band distortion in this case is much richer than that in the single-sinusoid case. With the two-tone excitation one can identify the linear components arising from the a_1 terms and the nonlinear components arising from the a_3 terms.

For the nonlinear components, two further distinctions can be made, since two terms will fall in frequency sidebands, namely the cases of $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$, and two other terms will fall right on top of the input signal at ω_1 and ω_2.

The terms falling in the sidebands are normally called intermodulation distortion (IMD). Actually, every nonlinear mixing product can be denominated as an intermodulation component since it results from intermodulating two or more different tones. But, although it cannot be said to be universal practice, the term IMD is usually reserved for those particular sideband components.

This form of distortion actually constitutes a form of adjacent-channel distortion.

The terms that actually fall on top of ω_1 and ω_2 are known as the co-channel distortion, and in fact, if we look carefully, we see that they can actually be divided into two separate forms. For instance, for ω_1,

$$\begin{aligned} y_{\mathrm{NL}_{\text{co-channel}}}(t) & = \left(\frac{3}{4} a_3 A_1^3 + \frac{6}{4} a_3 A_1 A_2^2\right) \cos\left(\omega_1 t - \phi_{3_{10}}\right) \\ & = \frac{3}{4} a_3 A_1^3 \cos\left(\omega_1 t - \phi_{3_{10}}\right) + \frac{6}{4} a_3 A_1 A_2^2 \cos\left(\omega_1 t - \phi_{3_{10}}\right) \end{aligned} \tag{1.55}$$

which corresponds to a term that depends only on A_1^3, which is perfectly correlated with the input signal at A_1, and another term that falls on top of ω_1 but depends also on the A_2 term, meaning that it can be uncorrelated with the input signal.

Actually, the correlated version of the output signal $\frac{3}{4} a_3 A_1^3 \cos(\omega_1 t - \phi_{3_{10}})$ is the term that is responsible for the compression or expansion of the device gain. This is similar to what was previously said regarding single-tone excitations, which we called AM–AM and AM–PM responses.

The other term, $\frac{6}{4} a_3 A_1 A_2^2 \cos(\omega_1 t - \phi_{3_{10}})$, which also includes a contribution from A_2 and can be uncorrelated with the input signal, is actually the worst problem in terms of communication signals. It is sometimes referred to as distortion noise. In wireless communications it is this type of nonlinear distortion that can degrade, for instance, the

Table 1.1 Two-tone nonlinear distortion mixing products up to third order

Mixing product frequency	Output amplitude	Result
ω_1	$(1/2)a_1 A_1$	Linear response
ω_2	$(1/2)a_1 A_2$	Linear response
$\omega_1 - \omega_1$	$(1/2)a_2 A_1^2$	Change in DC bias point
$\omega_2 - \omega_2$	$(1/2)a_2 A_2^2$	Change in DC bias point
$\omega_2 - \omega_1$	$(1/2)a_2 A_1 A_2$	Second-order mixing response
$2\omega_1$	$(1/4)a_2 A_1^2$	Second-order harmonic response
$\omega_1 + \omega_2$	$(1/2)a_2 A_1 A_2$	Second-order mixing response
$2\omega_2$	$(1/4)a_2 A_2^2$	Second-order harmonic response
$2\omega_1 - \omega_2$	$(3/8)a_3 A_1^2 A_2$	Third-order intermodulation distortion
$\omega_1 + \omega_2 - \omega_2$	$(3/4)a_3 A_1 A_2^2$	Cross-modulation response
$\omega_1 + \omega_1 - \omega_1$	$(3/8)a_3 A_1^3$	AM–AM and AM–PM response
$\omega_2 + \omega_2 - \omega_2$	$(3/8)a_3 A_2^3$	AM–AM and AM–PM response
$\omega_2 + \omega_1 - \omega_1$	$(3/4)a_3 A_1^2 A_1^2$	cross-modulation response
$2\omega_2 - \omega_1$	$(3/8)a_3 A_1 A_2^2$	Third-order intermodulation distortion
$3\omega_1$	$(1/8)a_3 A_1^3$	Third-order harmonic response
$2\omega_1 + \omega_2$	$(3/8)a_3 A_1^2 A_2$	Third-order mixing response
$2\omega_2 + \omega_1$	$(3/8)a_3 A_1 A_2^2$	Third-order mixing response
$3\omega_2$	$(1/8)a_3 A_2^3$	Third-order harmonic response

error-vector magnitude (the definition of which will be given later) in digital communication standards.

Table 1.1 summarizes the above definitions by identifying all of the distortion components falling on the positive side of the spectrum which are present in the output of our third-degree polynomial subjected to a two-tone excitation signal.

Following this nonlinear study for a two-tone signal excitation, some figures of merit can be defined.

1.5.2.1 The intermodulation ratio

The intermodulation ratio (IMR) is, as the name states, the ratio between the power corresponding to the output that appears exactly at the same positions as the input spectral components (these components will be called from now on the power at the fundamental frequency) and the power corresponding to the intermodulation power.

DEFINITION 1.4 *The intermodulation ratio (IMR) is defined as the ratio between the fundamental and intermodulation (IMD) output powers:*

$$\mathrm{IMR} = \frac{P_{\mathrm{out}_{\mathrm{fund}}}}{P_{\mathrm{IMD}}} \tag{1.56}$$

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

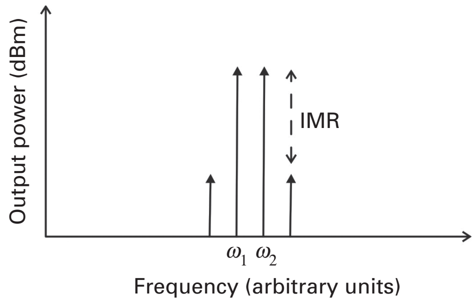

Figure 1.14 IMR definition in a two-tone excitation.

It should be noticed here that the output power at the fundamental frequency already includes some nonlinear distortion that appears at the same frequency as the input. This was seen in Section 1.5.2 as the co-channel distortion responsible for the AM–AM curves for the case of single-tone excitation, and consequently for the gain compression and expansion. On considering Fig. 1.14 and Eq. (1.56), it is clear that the intermodulation ratio refers only to the in-band nonlinear distortion, not to the harmonic content. This measure is usually described in dBc, meaning decibels below carrier. It should also be pointed out here that the upper IMD and the lower IMD may be different, which is called IMD asymmetry [6]. The IMR must then be defined as upper or lower.

In order to observe the results for the two-tone case, as was stated above, the in-band distortion is

$$\begin{aligned} y_{\mathrm{NL}_{\text{in-band}}}(t) &= a_1 A_1 \cos\left(\omega_1 t - \phi_{1_{10}}\right) + a_1 A_2 \cos\left(\omega_2 t - \phi_{1_{10}}\right) \\ &+ \frac{3}{4} a_3 A_1^2 A_2 \cos\left[(2\omega_1 - \omega_2)t - \phi_{3_{2-1}}\right] \\ &+ \left(\frac{3}{4} a_3 A_1^3 + \frac{6}{4} a_3 A_1 A_2^2\right) \cos\left(\omega_1 t - \phi_{3_{10}}\right) \\ &+ \left(\frac{3}{4} a_3 A_2^3 + \frac{6}{4} a_3 A_2 A_1^2\right) \cos\left(\omega_2 t - \phi_{3_{01}}\right) \\ &+ \frac{3}{4} a_3 A_1 A_2^2 \cos\left[(2\omega_2 - \omega_1)t - \phi_{3_{-12}}\right] \end{aligned} \tag{1.57}$$

which has terms that are clearly co-channel distortion, and thus will add to the output linear response, namely those appearing at

$$\begin{aligned} y_{\mathrm{NL}_{\text{in-band-co-channel}}}(t) &= a_1 A_{i1} \cos\left(\omega_1 t - \phi_{1_{10}}\right) + a_1 A_{i2} \cos\left(\omega_2 t - \phi_{1_{10}}\right) \\ &+ \left(\frac{3}{4} a_3 A_{i1}^3 + \frac{6}{4} a_3 A_{i1} A_{i2}^2\right) \cos\left(\omega_1 t - \phi_{3_{10}}\right) \\ &+ \left(\frac{3}{4} a_3 A_{i2}^3 + \frac{6}{4} a_3 A_{i2} A_{i1}^2\right) \cos\left(\omega_2 t - \phi_{3_{01}}\right) \end{aligned} \tag{1.58}$$

In this case the IMD power and the fundamental linear output power will be

$$P_{\rm IMD}(2\omega_1-\omega_2)=\frac{1}{T_{2\omega_1-\omega_2}}\int_0^{2\omega_1-\omega_2}\left[\frac{3}{4}a_3A_{i1}^2A_{i2}\cos\left[(2\omega_1-\omega_2)t-\phi_{3_{2-1}}\right]\right]^2dt$$
$$=\frac{9}{32}a_3^2A_{i1}^4A_{i2}^2$$
$$P_{\rm fund}(\omega_1)=\frac{1}{T_{\omega_1}}\int_0^{\omega_1}\left[a_1A_{i1}\cos\left(\omega_1t-\phi_{1_{10}}\right)\right.$$
$$\left.+\left(\frac{3}{4}a_3A_{i1}^3+\frac{6}{4}a_3A_{i1}A_{i2}^2\right)\cos\left(\omega_1t-\phi_{3_{10}}\right)\right]^2dt$$
$$=\frac{1}{2}a_1^2A_{i1}^2+\frac{1}{2}\left(\frac{3}{4}a_3A_{i1}^3+\frac{6}{4}a_3A_{i1}A_{i2}^2\right)^2$$
$$+a_1A_{i1}\left(\frac{3}{4}a_3A_{i1}^3+\frac{6}{4}a_3A_{i1}A_{i2}^2\right)\cos\left(\phi_{1_{10}}-\phi_{3_{10}}\right)\qquad(1.59)$$

The same calculations should be done for the IMD at $2\omega_2-\omega_1$, resulting in an IMR for two tones as

$$\mathrm{IMR}_{2\mathrm{t_{low}}}=\frac{32}{9a_3^2A_{i1}^4A_{i2}^2}$$
$$\times\left[\frac{1}{2}a_1^2A_{i1}^2+\frac{1}{2}\left(\frac{3}{4}a_3A_{i1}^3+\frac{6}{4}a_3A_{i1}A_{i2}^2\right)^2\right.$$
$$\left.+a_1A_{i1}\left(\frac{3}{4}a_3A_{i1}^3+\frac{6}{4}a_3A_{i1}A_{i2}^2\right)\cos(\phi_{1_{10}}-\phi_{3_{10}})\right]\qquad(1.60)$$

Nevertheless, for low-power signals the nonlinear contribution to the co-channel distortion is insignificant, and thus sometimes only the linear output power is considered when calculating the overall IMR, which will be

$$\mathrm{IMR}_{2\mathrm{t_{low}}}=\frac{\frac{1}{2}a_1^2A_{i1}^2}{(9/32)a_3^2A_{i1}^4A_{i2}^2}\qquad(1.61)$$

For equal input signal amplitude in both tones, $A_i=A_{i1}=A_{i2}$, the IMR value will finally be

$$\mathrm{IMR}_{2\mathrm{t}}=\frac{\frac{1}{2}a_1^2A_i^2}{(9/32)a_3^2A_i^6}$$
$$=\frac{16a_1^2}{9a_3^2A_i^4}\qquad(1.62)$$

1.5.2.2 Underlying linear gain

Actually it should also be mentioned here that sometimes a figure of merit called underlying linear gain (ULG) is defined. This gain accounts for the overall output signal that is correlated with the input signal, thus the ULG is given by

$$\mathrm{ULG}=\frac{P_{\mathrm{out_{fund}}}}{P_{\mathrm{in_{fund}}}}\qquad(1.63)$$

For a two-tone excitation it will be

$$\mathrm{ULG}_{2\mathrm{t_{low}}} = \frac{\frac{1}{2}a_1^2 A_{i1}^2 + \frac{1}{2}\left(\frac{3}{4}a_3 A_{i1}^3\right)^2 + a_1 A_{i1}\left(\frac{3}{4}a_3 A_{i1}^3\right)\cos\left(\phi_{1_{10}} - \phi_{3_{10}}\right)}{\frac{1}{2}A_{i1}^2} \tag{1.64}$$

In the case of linear systems, this gain reduces to the linear gain, that is LG $= \frac{1}{2}a_1^2 A_{i1}^2 / \left(\frac{1}{2}A_{i1}^2\right) = a_1^2$. Actually, with the ULG we are accounting for the AM–AM impact on the overall gain.

1.5.2.3 Intercept points

If the output power at the fundamental and that at the IMD spectral components are plotted versus input power for traditional nonlinear components, the results seen in Fig. 1.15 are observed. In this case the fundamental output power will start first with a linear progression with the input power. That is, a 1-dB increase in input power will impose a 1-dB increase in output power. Next it will start compressing or expanding the growth slope accordingly to Eq. (1.59). The IMD power will start at a lower level than the fundamental power, since it depends on a third-order polynomial. Then it will rise at a slope of 3 dB for each additional 1 dB at the input, corresponding to the third-order polynomial. Finally, for higher values of output power, the IMD will compress or expand according to higher orders of distortion. If the linear output response on the one hand and the third-order small-signal response on the other hand are extrapolated, it gives rise to a FOM called the third-order intercept point (IP_3). This FOM allows wireless engineers to calculate the small-signal nonlinear response very efficiently. Actually, this is very important for obtaining the amount of nonlinear distortion that arises from an interferer at the wireless system receiver.

DEFINITION 1.5 *The third-order intercept point (IP_3) is a fictitious point that is obtained when the extrapolated 1-dB/dB slope line of the output fundamental power intersects the extrapolated 3-dB/dB slope line of the IMD power.*

From a mathematical point of view, we will have to calculate the input or output power at which they intercept, thus, referring to Eq. (1.59),

$$\frac{1}{2}a_1^2 A_i^2 = \frac{9}{32}a_3^2 A_i^6 \rightarrow A_i^2 = \frac{4}{3}\frac{a_1}{a_3} \tag{1.65}$$

$$\mathrm{IP}_3 = P(\omega_1) = \frac{2}{3}\frac{a_1^3}{a_3} \tag{1.66}$$

In this case the output IP_3 was calculated, but in certain cases it is preferable to calculate the input IIP_3; see Fig. 1.15.

It should be mentioned that, despite their rarely being seen, some other intercept figures of merit could be defined for fifth-order (IP_5) or seventh-order (IP_7) distortion.

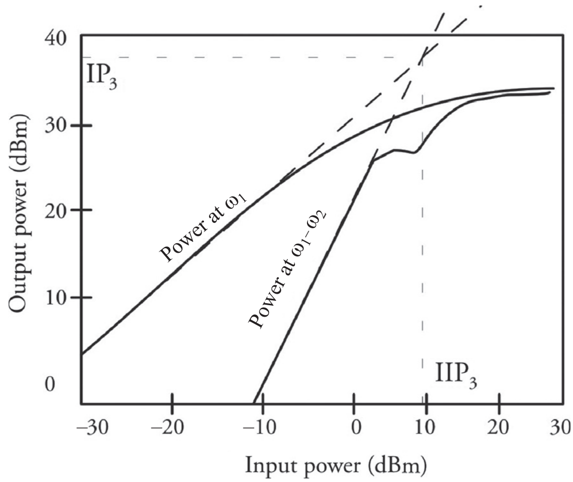

Figure 1.15 The definition of IP3. The extrapolated IP3 can be seen as the intercept of the third-order and fundamental power rises.

IP_3 can be further used to calculate the intermodulation power at any input power, if restricted to the small-signal region. It is then possible to relate the IMR to IP_3 by [4]

$$IP_{3_{dB}} = P_{fund_{dB}} + \frac{1}{2} IMR_{dB} \tag{1.67}$$

or

$$IMR_{dB} = 2\left(IP_{3_{dB}} - P_{fund_{dB}}\right) \tag{1.68}$$

These equations were calculated for a two-tone input signal with equal amplitudes in both tones, and $P_{fund_{dB}}$ is the output power at a specific tone.

1.5.2.4 Nonlinear distortion in the presence of dynamic effects

It should be stated here that certain nonlinear DUTs present what are called memory effects. These effects are a representation of dynamics in the nonlinear generation mechanism, as is discussed in [6]. The dynamic effects can mask the real intermodulation distortion in the DUT, since the lower sideband and higher sideband of the two-tone analysis can be different. Most often the dynamics arise from a baseband component being mixed with the fundamentals. So the dynamic effects can be measured by exciting these baseband components.

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

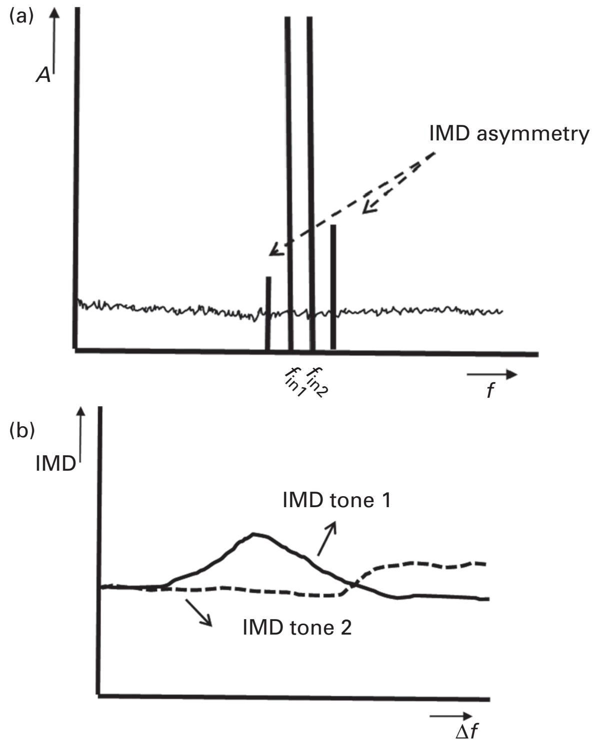

Figure 1.16 The impact of two-tone nonlinear dynamics: (a) two-tone IMD measurement when memory effects are visible; and (b) two-tone IMD variation with tone spacing.

These phenomena can be measured in the laboratory by exciting the nonlinear device with a signal that covers most of the baseband spectrum behavior. One way to achieve that is by using a two-tone signal and varying the tone spacing between the tones. If the IMD is measured with varying tone separation, then the impact of the baseband envelope behavior will be seen in the IMD variation, as shown in Fig. 1.16.

1.5.3 FOMs for nonlinear continuous spectra

New wireless communication standards, mainly digital wireless communications, have great richness in spectral content, which means that the single-tone and two-tone figures of merit have become obsolete, and are not able to characterize important aspects of that type of transmission.

Thus microwave and wireless system engineers have started to use other forms of excitation and other types of FOM to account for nonlinear distortion in digital wireless communication signals. Moreover, the nonlinear nature of RF components means that there will be a close relationship between the usefulness of a certain characterization technique and the similarity of the test signal to the real equipment's excitation.

In this sense engineers are considering other forms of excitation, including digitally modulated carriers with pseudo-random baseband signals, multi-tones (more than two

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

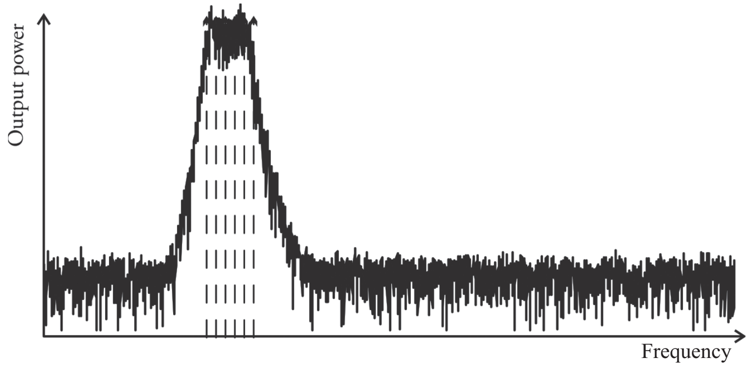

Figure 1.17 A typical input signal in a wireless communication system.

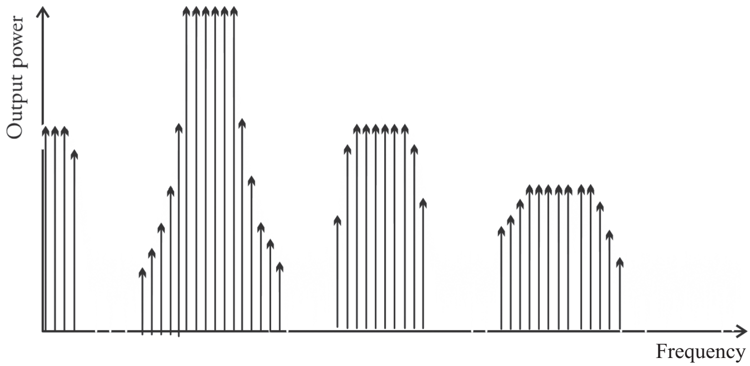

Figure 1.18 The nonlinear output response of a multi-sine signal excitation.

tones, usually called multi-sines), and band-limited noise [7]. In this section we will address figures of merit developed for rich spectra, continuous or not.

In Fig. 1.17 we can see the typical input in a wireless communication system, for which the spectrum is actually continuous. Microwave engineers sometimes use similar signals in the laboratory for mimicking this type of spectral richness using a multi-sine signal, Fig. 1.18, which allows a much simpler analysis of the output signal.

Figures 1.18 and 1.19 present the output of a typical signal like this, where the baseband, IMD, second harmonic and third harmonic are evident. This is similar to what we had previously seen in a two-tone excitation, but now each cluster is much richer.

In this sense the out-of-band distortion is dealt with by deploying the same line of thought as that which was used for the single-tone and the two-tone signal, meaning that in traditional wireless signals these are eliminated using output filtering. So the typical FOM for wireless systems usually refers mainly to the in-band components.

Figure 1.20 presents the in-band distortion that can be seen at the output of a nonlinear system when it is excited by a bandlimited continuous spectrum.

It is evident that we can identify the co-channel distortion that falls on top of the linear output signal, corresponding to a linear complex gain multiplication of the input

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

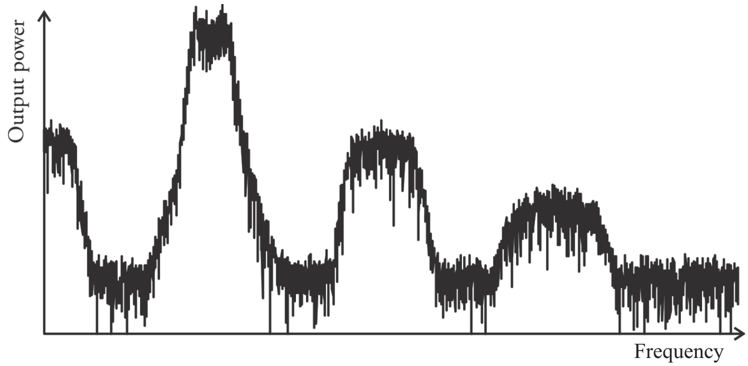

Figure 1.19 The nonlinear output response of a rich input spectrum.

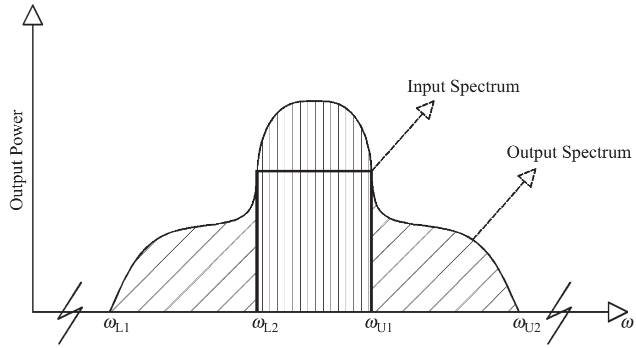

Figure 1.20 The in-band nonlinear output response of a rich input spectrum.

spectrum. We can also see a sideband on each side, corresponding to what is called spectral regrowth. It arises from the nonlinear odd-order terms, similarly to the IMD tones appearing in a two-tone excitation. Having this signal in mind, we can now define some important figures of merit for characterization of nonlinear rich spectra.

1.5.3.1 The multi-sine intermodulation ratio

The multi-sine intermodulation ratio (M-IMR) is actually a generalization of the IMR concept introduced in Section 1.5.2.1. As can be seen from Fig. 1.21, this figure of merit can be defined as follows.

DEFINITION 1.6 *The multi-sine intermodulation ratio (M-IMR) is defined as the ratio of the common fundamental power per tone, $P_{\text{fund}_{\text{tone}}}$, to the power of the ω_r distortion component present in the lower or upper adjacent bands, $P_{\text{L/U}}(\omega_\text{r})$.*

In mathematical terms it is nothing other than

$$\text{M-IMR} = \frac{P_{\text{fund}_{\text{tone}}}}{P_{\text{L/U}}(\omega_\text{r})} \tag{1.69}$$

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

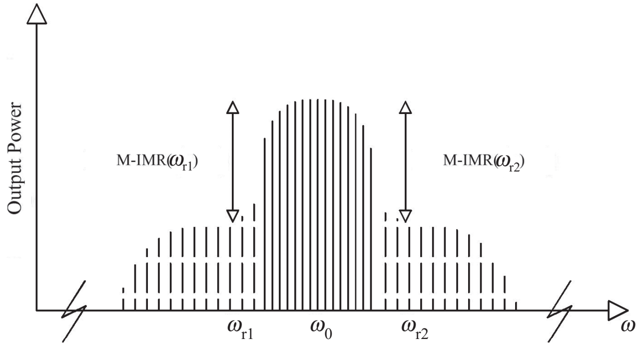

Figure 1.21 The definition of the multi-sine intermodulation ratio.

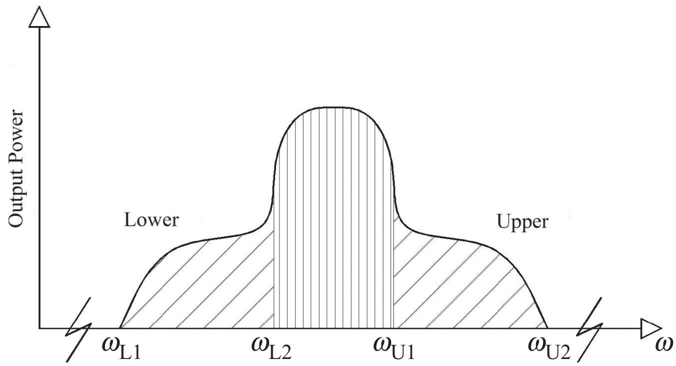

Figure 1.22 The definition of the adjacent-channel power ratio.

1.5.3.2 The adjacent-channel power ratio

The FOM known as the M-IMR allows engineers to measure and account for each tone in the multi-sine approach, but when the signal is continuous the user should account not for each sine, but for all the spectral power that is being created by the nonlinearity. This part of the spectrum is called adjacent-channel distortion, and is composed of all distortion components falling on the adjacent-channel location. Actually, in typical communication scenarios it can be a source of interference with adjacent channels. For accounting for this type of distortion, and mainly the amount of power being regrown, several FOMs can be defined.

DEFINITION 1.7 *The total adjacent-channel power ratio ($ACPR_T$) is the ratio of the total output power measured in the fundamental zone, P_{fund}, to the total power integrated in the lower, P_{LA}, and upper, P_{UA}, adjacent-channel bands.*

Figure 1.22 shows this FOM and how it is calculated. Mathematically it can be described as

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

$$\mathrm{ACPR_T} = \frac{P_{\mathrm{fund}}}{P_{\mathrm{LA}} + P_{\mathrm{UA}}} = \frac{\int_{\omega_{\mathrm{L}_2}}^{\omega_{\mathrm{U}_1}} S_{\mathrm{o}}(\omega)d\omega}{\int_{\omega_{\mathrm{L}_1}}^{\omega_{\mathrm{L}_2}} S_{\mathrm{o}}(\omega)d\omega + \int_{\omega_{\mathrm{U}_1}}^{\omega_{\mathrm{U}_2}} S_{\mathrm{o}}(\omega)d\omega} \tag{1.70}$$

where $S_{\mathrm{o}}(\omega)$ is the power spectral density.

Sometimes it is also interesting to address only a specific part of the spectral regrowth, and in that situation we can define the upper or lower ACPR value. This FOM can be defined as follows.

DEFINITION 1.8 *The upper or lower adjacent-channel power ratio ($ACPR_L$ or $ACPR_U$) is the ratio between the total output power measured in the fundamental zone, P_{fund}, and the lower or upper adjacent-channel power, P_{LA} or P_{UA}.*

Mathematically,

$$\begin{aligned} \mathrm{ACPR_L} &= \frac{P_{\mathrm{fund}}}{P_{\mathrm{LA}}} = \frac{\int_{\omega_{\mathrm{L}_2}}^{\omega_{\mathrm{U}_1}} S_{\mathrm{o}}(\omega)d\omega}{\int_{\omega_{\mathrm{L}_1}}^{\omega_{\mathrm{L}_2}} S_{\mathrm{o}}(\omega)d\omega} \\ \mathrm{ACPR_U} &= \frac{P_{\mathrm{fund}}}{P_{\mathrm{UA}}} = \frac{\int_{\omega_{\mathrm{L}_2}}^{\omega_{\mathrm{U}_1}} S_{\mathrm{o}}(\omega)d\omega}{\int_{\omega_{\mathrm{U}_1}}^{\omega_{\mathrm{U}_2}} S_{\mathrm{o}}(\omega)d\omega} \end{aligned} \tag{1.71}$$

Sometimes it is preferable to consider not all of the adjacent-channel power, but only a piece of it. That happens because in continuous spectra it is usually difficult to define where the adjacent-channel spectrum starts and ends, mainly due to the roll-off of the system filters. Thus the industry refers to this FOM as the spot ACPR, this being defined as follows.

DEFINITION 1.9 *The spot adjacent-channel power ratio ($ACPR_{SP_L}$ or $ACPR_{SP_U}$) is the ratio of the total output power measured in the fundamental zone, P_{fund}, to the power integrated in a band of predefined bandwidth and distance from the center frequency of operation $P_{\mathrm{SP_{L/U}}}$.*

Mathematically it can be described as

$$\begin{aligned} \mathrm{ACPR_{SP_L}} &= \frac{P_{\mathrm{fund}}}{P_{\mathrm{SP_L}}} = \frac{\int_{\omega_{\mathrm{L}_2}}^{\omega_{\mathrm{U}_1}} S_{\mathrm{o}}(\omega)\mathrm{d}\omega}{\int_{\omega_{\mathrm{NBL}_1}}^{\omega_{\mathrm{NBL}_2}} S_{\mathrm{o}}(\omega)d\omega} \\ \mathrm{ACPR_{SP_U}} &= \frac{P_{\mathrm{fund}}}{P_{\mathrm{SP_U}}} = \frac{\int_{\omega_{\mathrm{L}_2}}^{\omega_{\mathrm{U}_1}} S_{\mathrm{o}}(\omega)d\omega}{\int_{\omega_{\mathrm{NBU}_1}}^{\omega_{\mathrm{NBU}_2}} S_{\mathrm{o}}(\omega)d\omega} \end{aligned} \tag{1.72}$$

Figure 1.23 shows this figure of merit and how it is calculated.

1.5.3.3 Co-channel distortion FOMs

As was seen previously, nonlinear distortion generates adjacent-band spectral regrowth, but also co-channel distortion, which imposes a strong degradation of the signal-to-noise ratio. Unfortunately, co-channel distortion falls exactly on top of the input signal spectrum, and thus on top of the linear output signal.

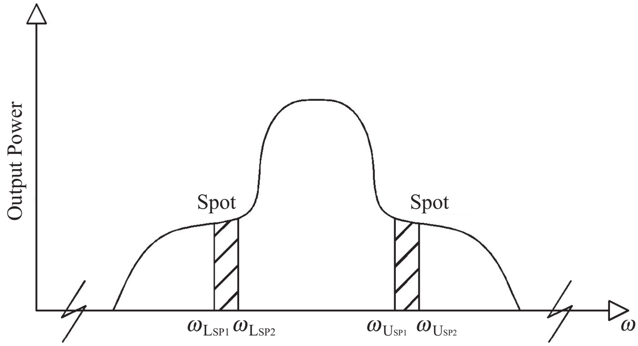

Figure 1.23 The definition of the adjacent-channel spot power ratio.

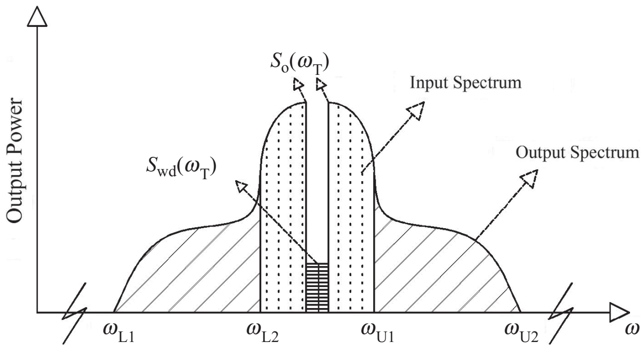

Figure 1.24 The definition of the noise power ratio.

This type of distortion can be accounted for using a FOM known as the co-channel power ratio (CCPR), which allows a correct measure of the nonlinear distortion falling inside the band. However, because this form of distortion is intricately mixed with the fundamental zone of much higher amplitude, the measurement of this type of distortion is quite difficult. That is why wireless system engineers came up with a FOM called the noise power ratio (NPR) that actually allows one to characterize the co-channel distortion in an indirect way. The basic idea is mainly to open a notch in the input signal spectrum, and to account for the noise in that notch hole both at the input and at the output, Fig. 1.24. We will see in future chapters how to measure this type of co-channel distortion, but let us now define these co-channel FOMs.

The NPR can be defined as follows.

DEFINITION 1.10 *The noise power ratio (NPR) is defined as the ratio of the output power spectral density function measured in the vicinity of the test window position,* ω_T, $S_o(\omega_T)$*, to the power spectral density observed within that window,* $S_{wd}(\omega_T)$.

Mathematically it is expressed by

$$\mathrm{NPR}(\omega_T) = \frac{S_o(\omega_T)}{S_{wd}(\omega_T)} \tag{1.73}$$

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

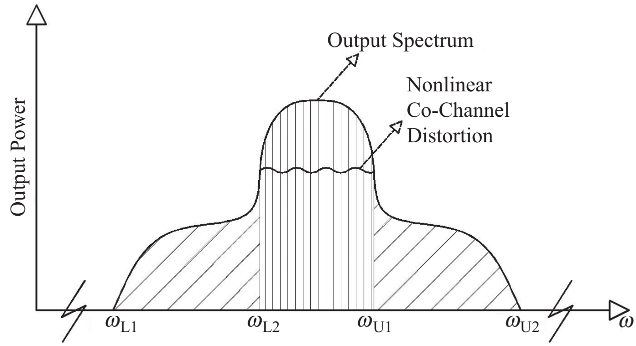

Figure 1.25 The definition of the co-channel power ratio.

The co-channel power ratio (Fig. 1.25) can be defined as follows.

DEFINITION 1.11 *The co-channel power ratio (CCPR) is defined as the ratio of the integrated output power measured in the fundamental zone,* P_{fund}*, to the total integrated co-channel perturbation,* $P_{\text{co-channel}}$.

Mathematically it is expressed as

$$\text{CCPR} = \frac{P_{\text{fund}}}{P_{\text{co-channel}}} = \frac{\int_{\omega_{L_2}}^{\omega_{U_1}} S_o(\omega)d\omega}{\int_{\omega_{L_2}}^{\omega_{U_1}} S_{\text{co-channel distortion}}(\omega)d\omega} \tag{1.74}$$

1.6 System-level FOMs

Regarding general figures of merit that can be applied to generic RF and wireless components and circuits, some system-level FOMs will now be presented. These FOMs relate more generally not to a single component or circuit, but rather to a bigger system. In that sense they are normally stated as high-level quantities, and most of the time they are related to the information being sent over the communication channel and thus not necessarily to any spectral or time characteristics.

Some of these FOMs include the error-vector magnitude and the bit error rate, which are FOMs that can be measured at a high level of abstraction and that depend not on a single component but rather on the activity of the complete system.

1.6.1 The constellation diagram

In a digital radio, the evaluation of the transmitted signals is fundamental. This characterization can be done by referring to the constellation diagram. Let us explain the concept of a constellation diagram. In a digital modulated RF signal we can describe

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

the input and output signal as a sine wave that is in phase or in quadrature phase arrangement:

$$
\begin{aligned}
x(t) &= A(t)\cos(\omega t + \theta(t)) \\
&= A(t)\frac{e^{j(\omega t+\theta(t))} + e^{-j(\omega t+\theta(t))}}{2} \\
&= \mathrm{Re}\left[A(t)e^{-j(\omega t+\theta(t))}\right] \\
&= \mathrm{Re}\left[A(t)e^{-j\theta(t)}e^{-j\omega t}\right] \\
&= \mathrm{Re}\left[\tilde{x}(t)e^{-j\omega t}\right] \\
&= \mathrm{Re}\left\{[I(t) + jQ(t)]e^{-j\omega t}\right\} \\
&= I(t)\cos(\omega t) + Q(t)\sin(\omega t)
\end{aligned}
\tag{1.75}
$$

By representing a wireless signal as a complex number and modulating a cosine and sine carrier signal with the real and imaginary parts, respectively, the symbol can be sent with orthogonal carriers on the same frequency.

These carriers are often referred to as quadrature carriers. If the wireless system uses a coherent detector, then it is possible to independently demodulate these carriers. This principle is actually the base for quadrature modulation.

Actually the phase and quadrature information is normally called the complex envelope of the signal, and is represented in several ways. One possibility is to include a two-time-domain graph with $I(t)$ and $Q(t)$ plotted over time; the other possibility is using a constellation diagram, where the phase and quadrature values are plotted over each other in a complex graph representation (see Fig. 1.26).

Since the symbols are represented as complex numbers, they can be visualized as points on the complex plane. The real-number and imaginary-number coordinate axes are often called the in-phase axis, or I-axis, and the quadrature axis, or Q-axis.

Plotting several symbols in a scatter diagram produces the constellation diagram. The points on a constellation diagram are called constellation points. They are a set of modulation symbols comprising the modulation alphabet.

Thus a constellation diagram is nothing more than a representation of these $I(t)$ and $Q(t)$ in a complex diagram, plotting each pattern in phase ($I(t)$) and quadrature ($Q(t)$). Actually, it displays the signal as a two-dimensional scatter diagram in the complex plane. Sometimes the plot is displayed only at the symbol-sampling instants, but the overall trajectory is also very important for understanding certain aspects of the system behavior.

In a pure quadrature phase-shift keying (QPSK) signal the constellation diagram will be that presented in Fig. 1.27(a) if the represented symbols are plotted only at the symbol rate, but it will be that presented in Fig. 1.27(b) if the overall trajectories of $I(t)$ and $Q(t)$ are plotted.

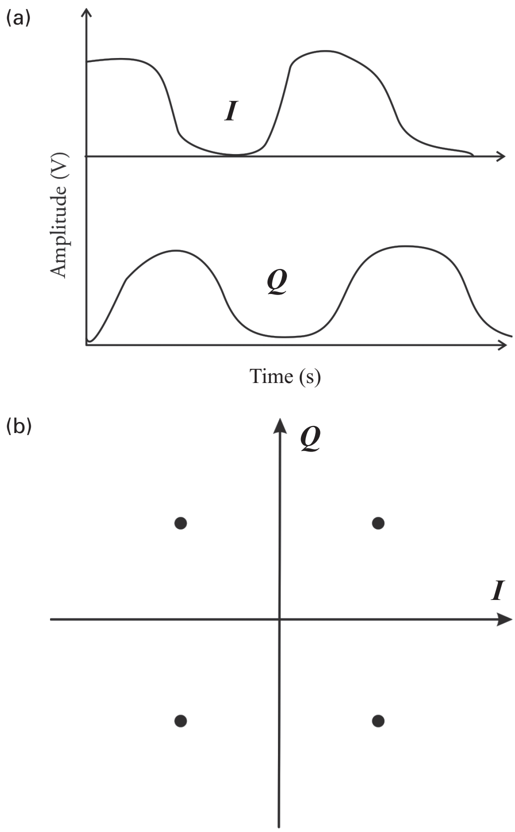

Figure 1.26 The $I(t)$ and $Q(t)$ representation: (a) time-domain waveforms and (b) the constellation diagram.

Constellation diagrams can be used to recognize the type of interference and distortion in a signal.

This tool is very important, since it allows engineers to observe the transmitted constellation points and to compare them with the received ones. In this way, they are able to identify the similarities and differences between them, thereby accounting for the signal degradation from a point of view that is actually the ultimate one, meaning that it relates directly to the information transmission quality.

The system itself can degrade the transmitted signal by adding noise to the signal, or can degrade it due to nonlinear distortion, as was seen in Section 1.4.

In terms of transmission degradation, engineers will seek in the constellation diagram a deviation of the actual received signal from the transmitted one, and will calculate this difference using some form of Euclidean distance. Thus the receiver will demodulate the received signal incorrectly if the corruption has caused the received symbol to move closer to another constellation point than the one transmitted.

This is actually called maximum-likelihood detection. The use of the constellation diagram allows a straightforward visualization of this process.

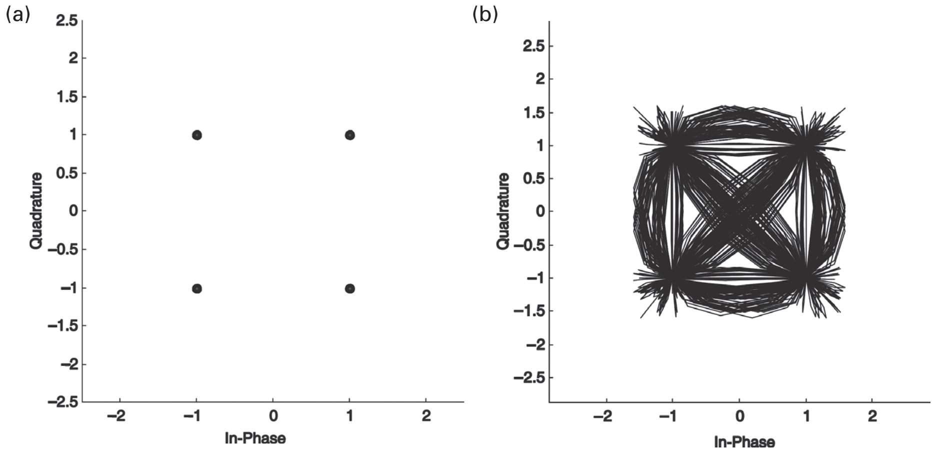

Figure 1.27 The QPSK constellation diagram: (a) Sampled at the symbol rate; and (b) the overall trajectory.

1.6.2 The error-vector magnitude

Considering the constellation-diagram approach, the first system-level FOM to be presented is the error-vector magnitude (EVM). The EVM is a measure that is used to evaluate the performance of an RF system in digitally modulated radios.

An ideal signal sent by an ideal transmitter without any interference will have all constellation points precisely at the ideal locations. However, if the signal is interfered with by different propagation-channel imperfections, such as noise, nonlinear distortion, phase noise, adjacent-channel interference, etc., the symbols and thus the constellation points will deviate from the ideal locations.

The EVM is actually accounting for the errors in the points in a constellation diagram. It is nothing more than a measure of how far the points are from their ideal locations.

DEFINITION 1.12 *The error-vector magnitude (EVM) is a vector (geometric) in the I–Q plane between the ideal constellation point and the point received by the receiver. It can also be stated as the difference between actually received symbols and ideal symbols. The average power of the error vector, normalized with respect to the signal power, is the EVM. For the percentage format, the root-mean-square (*rms*) average is used.*

In mathematical terms the EVM is expressed as a percentage:

$$\mathrm{EVM}_{[\mathrm{RMS}]} = \sqrt{\frac{\sum_{k=1}^{N} |Z_{\mathrm{c}}(k) - S(k)|^2}{\sum_{k=1}^{N} |S(k)|^2}} \tag{1.76}$$

$$\mathrm{EVM}_{[\%]} = \mathrm{EVM}_{[\mathrm{RMS}]} \times 100 \tag{1.77}$$

where N is the number of received symbols, $Z_{\mathrm{c}}(\,)$ is the actual received symbol, and $S(\,)$ is the ideal symbol that should be received (Fig. 1.28).

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

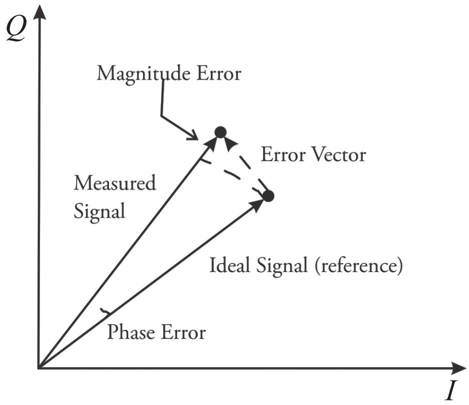

Figure 1.28 Calculation of the error-vector magnitude.

The EVM can also be obtained from the signal-to-noise ratio, and some relationships have been developed to compare these two quantities [8]:

$$\mathrm{EVM}_{[\mathrm{RMS}]} = \sqrt{\frac{1}{\mathrm{SNR}}} \tag{1.78}$$

This equation is very important, since it states that, by accounting for the noise degradation or the nonlinear distortion degradation inside the band, it is possible to account for the EVM of an overall system.

1.6.3 The peak-to-average power ratio

Another important FOM that actually cannot be attributed to the system itself, but rather must be attributed to the signal, is the one that accounts for the relationship between the peak power and the average power of the signal. Some authors [9] have used the peak-to-average power ratio (PAPR), as the ratio of the average power that would result if the envelope were sustained at its peak magnitude to the average power in the N-sinusoid sum. The PAPR thus has the mathematical form

$$\mathrm{PAPR} = \frac{\max|x(t)^2|}{[1/(NT)]\int_0^{NT} |x(t)|^2\, dt} \tag{1.79}$$

A version of the PAPR for a sampled signal can also be used. This is defined as

$$\mathrm{PAPR_s} = \frac{\max|x_k^2|}{E|(x(k)|^2\, dt} \tag{1.80}$$

where $E(\,)$ is the expectancy of x.

1.7 Filters

Filters, combiners, power dividers, isolators, and other linear components can be characterized using the linear FOMs presented in Section 1.3. In all these situations the

Table 1.2 A filter datasheet of electrical characteristics (guaranteed over $-50\,°C$ to $+90\,°C$ operating temperature)

Part number	Frequency band [MHz]	Insertion loss (dB)	Maximum VSWR (dB)	Typical attenuation (dB)
Filter 1	800–1000	0.35 typical (0.5 max.)	1.5	30 at $2F_0$
Filter 2	865–985	0.34 typical (0.5 max.)	1.4	27 at $2F_0$
Filter 3	1700–1900	0.37 typical (0.5 max.)	1.6	40 at $2F_0$

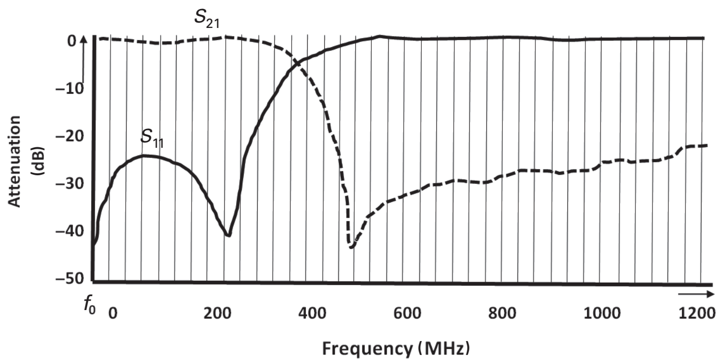

Figure 1.29 Measured S-parameters of a low-pass filter.

main FOMs to consider are the S-parameters, and correspondingly the insertion loss and VSWR as well as the bandwidth.

In order to better understand a typical datasheet of a filter, consider Table 1.2. On this example datasheet, several filters, all bandpass filters, are presented, each with their respective bandwidth, insertion loss, in this case near 0.5 dB in the band of interest, and VSWRs on the order of 1.4, which corresponds to an impedance mismatch of near $\Gamma_{in} = 0.17$ and to an impedance of 35 Ω to 70 Ω in a 50-Ω environment.

The bandwidth of a filter is usually determined by the lower and upper frequencies at which the in-band insertion loss has dropped with a certain dB value, typically 3 dB.

Another important FOM is the one corresponding to the out-of-band attenuation. In this sample case, it is specified at the double frequency of F_0, which is the bandwidth's upper frequency limit. These frequencies are indicated in Fig. 1.29, showing the measured S-parameters of a low-pass filter.

1.8 Amplifiers

Amplifiers have been the main components in any RF/microwave radio design, since they allow one to increase the signal level in order to be able to transmit or receive the communication signal with a certain signal-to-noise ratio value.

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

Depending on the section of the system where the amplifier has to be inserted, the amplifier can have different functions, for instance consider Fig. 1.1. If the amplifier is inserted into the receiver as its first stage, then the focus will be on the low-noise behavior, and therefore it is a low-noise amplifier (LNA). On the other hand, if the amplifier is to be inserted into the transmitter as its last stage, the amplifier is focused on increasing the output power, since the main objective of the transmitter is to achieve a high power level in order to fulfill the link budget that the system engineer has designed. In this case the amplifier is called a power amplifier (PA). In other sections of the transceiver chain, we can see some generic amplifiers (GAs) that are usually designed to maximize gain, and not necessarily for high power or low noise. This category of amplifier includes the variable-gain amplifiers (VGAs). Their objective is to reduce distortion in subsequent stages, by dynamically optimizing the gain in order to maintain a constant output amplitude.

In the next sections, the FOMs applicable to amplifiers are described. A distinction is made among linear and noise FOMs, which are of interest primarily for GAs and LNAs but also for PAs; nonlinear FOMs, which are primarily applicable to PAs; and transient FOMs, which are specific for VGAs. As an example, Table 1.3 presents a typical datasheet of an amplifier. The linear FOMs calculable from S-parameters are the gain as function of frequency, gain variation over temperature, input and output return loss, and reverse isolation. The noise figure is listed as well. The nonlinear FOMs listed in this example datasheet are the 1-dB-compression point, saturated power, and third-order intercept point. An amplifier datasheet always includes also the DC operating conditions. The temperature dependency is especially important for PAs.

1.8.1 Linear and noise FOMs

Since amplifiers are two-port networks, the linear FOMs described in Section 1.3 are generally applicable.

First of all, an RF engineer should be aware of any mismatch in the input and output connections, in order to guarantee that the power loss in these connections is minimal. This can be expressed by the VSWR, defined in Section 1.3.1.1, or by the return loss, RL, defined in Section 1.3.1.2.

The FOM related to noise is the noise figure, NF, which was defined in Section 1.3.2.1. The lower the noise figure, the better the performance of the low-noise amplifier. Nevertheless, even when not optimized, the noise figure is usually listed as well in the datasheets of the other types of amplifiers. It is a measure for the noise added by the amplifier in the system, which is of importance to evaluate the overall noise, as expressed by the Friis formula (1.31).

The aim of an amplifier is to amplify the input signal, so the gain is its most characteristic FOM. Before elaborating on the gain, it is also important to note that an amplifier should act as an isolator, or at least as a strong attenuator, in the reverse direction. Or, in other words, an excitation at its output, e.g., caused by a malfunctioning subsequent block in the chain, should not propagate to its input because this may damage the

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

Table 1.3 Amplifier datasheet, electrical characteristics (specified at 25 °C and 50 mA)

Parameter		Minimum	Typical	Maximum	Units
Frequency range		DC		10	GHz
Gain	$f = 0.1$ GHz $f = 1$ GHz $f = 2$ GHz $f = 6$ GHz $f = 10$ GHz	11.5 10 	12.5 12.3 11.5 11.1 10.8	13 12.9 	dB
Gain versus temperature	$f = 0.1$ GHz $f = 1$ GHz $f = 2$ GHz $f = 6$ GHz $f = 10$ GHz		0.001 0.001 0.0015 0.0019 0.0022	0.002 0.003 0.0035 0.0038 0.004	dB/ °C
Input return loss	$f = 0.1$ GHz $f = 1$ GHz $f = 2$ GHz $f = 6$ GHz $f = 10$ GHz		30 25 20 19 17		dB
Output return loss	$f = 0.1$ GHz $f = 1$ GHz $f = 2$ GHz $f = 6$ GHz $f = 10$ GHz		26 23 21 16 15		dB
Reverse isolation	$f = 2$ GHz	12	18		dB
Output 1-dB-compression point	$f = 0.1$ GHz $f = 1$ GHz $f = 2$ GHz $f = 6$ GHz $f = 10$ GHz		15 15 15 15 11		dBm
Saturated output power (at 3 dB compression)	$f = 0.1$ GHz $f = 1$ GHz $f = 2$ GHz $f = 6$ GHz $f = 10$ GHz		16 16 16 15 14		dBm
Output IP_3	$f = 0.1$ GHz $f = 1$ GHz $f = 2$ GHz $f = 6$ GHz $f = 10$ GHz	24 24 24 24 23	30 30 30 29 27		dBm
Noise figure	$f = 0.1$ GHz $f = 1$ GHz $f = 2$ GHz $f = 6$ GHz $f = 10$ GHz		4 4.2 4.2 4.4 4.6	5 5 5 5.2 5.5	dB
Group delay	$f = 2$ GHz		60		ps
Operating current			50		mA
Operating voltage		3	3.3	4	V
Voltage variation with temperature			−2		mV/ °C
Voltage variation with current			9		mV/mA
Thermal resistance			180		°C/W

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

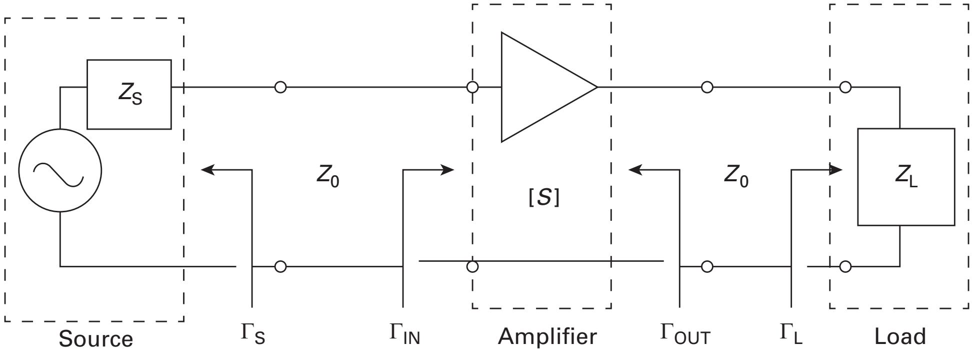

Figure 1.30 Power gain definitions.

preceding blocks. The corresponding FOM is the isolation. It is related to S_{12} since this is the reverse transmission coefficient (see Section 1.2):

$$\text{isolation} = \left|20 \log_{10} |S_{12}|\right| \text{ dB} \tag{1.81}$$

So intuitively one may think that the gain of an amplifier is expressed by its S_{21}. It is a more complex matter, though. In fact, there are three power gain definitions, which are the transducer power gain G_T, the operating power gain G, and the available power gain G_A. The three definitions are applicable to any amplifier, but a given type of power gain may be more suitable for particular cases, as we will describe below [1, 2].

Before introducing the definitions of these power gains, we first should introduce some other definitions, explained by means of Fig. 1.30.

Figure 1.30 shows a microwave amplifier that is excited by a source with source impedance Z_S and terminated with a load Z_L. In a system design, the load represents the input impedance of the circuit block following the amplifier, such as the antenna in a transmitter. Similarly, Z_S may represent the output impedance of the circuit block preceding the amplifier. In the subsequent equations, the amplifier is represented by its S-parameters. The source impedance Z_S corresponds to the reflection coefficient Γ_S, and similarly the load Z_L corresponds to the reflection coefficient Γ_L.

Two other reflection coefficients are indicated in Fig. 1.30, namely Γ_{IN} and Γ_{OUT}. The reflection coefficient Γ_{IN} is the input reflection coefficient of the amplifier followed by the load. Similarly, the reflection coefficient Γ_{OUT} is the output reflection coefficient of the amplifier preceded by the source impedance. Γ_{IN} and Γ_{OUT} can be expressed in terms of the blocks' parameters, as follows (see also Eq. (1.10)):

$$\Gamma_{IN} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \tag{1.82}$$

$$\Gamma_{OUT} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S}$$

Note that Γ_{IN} becomes independent of the load if the amplifier is unilateral. This means that there is no feedback from output to input, which corresponds to S_{12} equal to zero. The measure is the isolation FOM (Eq. (1.81)). Similarly, Γ_{OUT} is independent of the reflection coefficient of the source network, Γ_{S}, if the amplifier is unilateral. Also, Γ_{IN} reduces to S_{11} if Γ_{L} is equal to zero, or in other words Z_{L} is equal to 50 Ω. Similarly, Γ_{OUT} reduces to S_{22} if Γ_{S} is equal to zero, or, in other words, the source impedance is 50 Ω.

Before proceeding to the power gain definitions, we still need to define several powers. The power available from the source is denoted as P_{AVS}. The power effectively going into the amplifier is P_{IN}. P_{IN} is lower than P_{AVS} if there is a mismatch between the source and the input of the amplifier. If there is no mismatch, P_{IN} is equal to P_{AVS}. Similarly, at the output P_{AVN} stands for the output power available from the amplifier, while P_{L} is the power delivered to the load. P_{L} is smaller than P_{AVN} unless there is no mismatch between the amplifier output and the load.

The first power gain definition is the transducer power gain (G_{T}).

DEFINITION 1.13 *The transducer power gain (G_{T}) is the ratio of the power delivered to the load to the power available at the source.*

G_{T} is expressed by the following equation:

$$G_{\mathrm{T}} = \frac{P_{\mathrm{L}}}{P_{\mathrm{AVS}}}$$

$$= \frac{(1-|\Gamma_{\mathrm{S}}|)^2}{|1-\Gamma_{\mathrm{S}}\Gamma_{\mathrm{IN}}|^2}\,|S_{21}|^2\,\frac{(1-|\Gamma_{\mathrm{L}}|)^2}{|1-\mathrm{S}_{22}\Gamma_{\mathrm{L}}|^2} \tag{1.83}$$

$$= G_{\mathrm{S}}G_0G_{\mathrm{L}} \tag{1.84}$$

The expression has been written as a product of three factors in order to be able to better evaluate the contribution of each block. The first factor, G_{S}, relates to the interaction between the source network and the input of the amplifier, G_0 stands for the contribution of the amplifier itself, and G_{L} is related to the interaction between the output of the amplifier and the load. If the source and load are perfect, meaning $Z_{\mathrm{S}} = Z_{\mathrm{L}} = 50\ \Omega$, or $\Gamma_{\mathrm{S}} = \Gamma_{\mathrm{L}} = 0$, then G_{T} reduces to $|S_{21}|^2$. In other words, the power gain definitions take into account the mismatches between the amplifier and its preceding and following blocks. In the case of G_{T}, the mismatches both at the input and at the output are taken into account. The transducer power gain is the best applicable in the real situation when one wants to know how much power effectively gets delivered to the load when the source generates a certain power level. As we will see next, the operating power gain stresses the mismatch at the output, while the available power gain relates to the mismatch at the input. So the aim of the various power gain definitions is to express the actual power gain when the amplifier is embedded in a system. In the limit when there are no mismatches at input and output, the three power gains reach their maximal values and are equal to each other, $G_{\mathrm{T,max}} = G_{\mathrm{A,max}} = G_{\mathrm{max}}$.

The next definition is the operating power gain G.

DEFINITION 1.14 *The operating power gain (G) is the ratio of the power delivered to the load to the power going into the amplifier.*

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

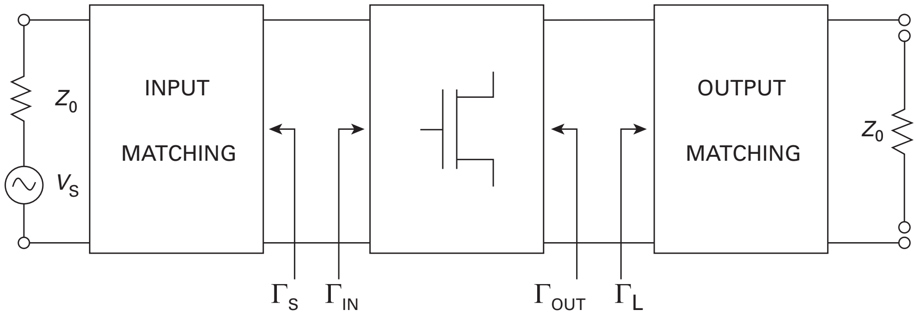

Figure 1.31 The general scheme for amplifier design.

The corresponding expression is

$$G = \frac{P_L}{P_{IN}} = \frac{1}{1 - |\Gamma_{IN}|^2} |S_{21}|^2 \frac{1 - |\Gamma_L|^2}{|1 - S_{22}\Gamma_L|^2} \quad (1.85)$$

We note that G is a function of Γ_L, while it is independent of Γ_S. In systems, Z_{load} is usually close to 50 Ω, and therefore the differences between the various power gain values are small.

However, the definitions of power gain are generally applicable to two-port networks, and are therefore also made use of during amplifier design. In such cases, Fig. 1.31 is appropriate. The amplifier is now represented as a cascade of three blocks: a central block that includes the active part, namely one or more transistors enabling the amplification, while the other two blocks are the passive input and output matching networks. Note that an actual amplifier design may be more complicated than this general scheme. Here again, the central block is represented by its S-parameters. The input matching network is represented by its reflection coefficient Γ_S, and similarly the output matching network is represented by its reflection coefficient Γ_L. It is assumed in the calculations that the source has no mismatch, or, in other words, that its impedance is Z_0. The same applies to the output, where it is assumed that the terminating load is a Z_0 impedance. So the operating power gain is typically of interest for power amplifiers, because in such designs the load Z_L is optimized for high power. Usually this value is not matched to the output impedance of the central active block, and therefore it reduces the operating power gain.

Finally, the available power gain G_A is defined as follows.

DEFINITION 1.15 *The available power gain (G_A) is the ratio of the power available from the amplifier to the power available from the source.*

The corresponding expression is

$$G_A = \frac{P_{AVN}}{P_{AVS}} = \frac{(1 - |\Gamma_S|)^2}{|1 - S_{11}\Gamma_S|^2} |S_{21}|^2 \frac{1}{(1 - |\Gamma_{OUT}|)^2} \quad (1.86)$$

We note that G_A is a function of Γ_S while it is independent of Γ_L. So, in amplifier design (according to Fig. 1.31), the available power gain is typically of interest for low-noise amplifiers, because in such designs Γ_S is designed to be Γ_{OPT} in order to achieve low-noise performance (see Section 1.3.2). Usually this value is not matched to the input impedance of the central active block, and therefore the available power gain is compromised.

The next FOMs gain in importance, due to the global drive for less energy consumption. They are a measure of the efficiency, namely how efficiently the amplifier converts the supplied DC power and RF input power into RF output power. We make the distinction between efficiency (η) and power added efficiency (PAE). Note that we are referring now to the actual use of the amplifier, or, in other words, to the scheme in Fig. 1.30.

DEFINITION 1.16 *The efficiency (η) of an amplifier is defined as the ratio between the output power at the fundamental frequency and the supplied DC power.*

The corresponding expression is

$$\eta = \frac{P_L}{P_{DC}} \tag{1.87}$$

This efficiency is also called the drain efficiency if the active part consists of FETs (or derived devices such as HEMTs), or the collector efficiency if the amplifier is based on BJTs (or derived devices such as HBTs). It is named in this way because this definition does not take into account the RF input power injected in to the amplifier.

The second FOM in terms of efficiency is the power added efficiency or PAE. It is more complete than η because it does take into account the RF input power. The definition is as follows.

DEFINITION 1.17 *The power added efficiency (PAE) of an amplifier is defined as the net increase in RF power divided by the DC power supplied.*

The corresponding expression is

$$\begin{aligned} \text{PAE} &= \frac{P_L - P_{IN}}{P_{DC}} \\ &= \frac{P_L}{P_{DC}}\left(1 - \frac{1}{G}\right) \\ &= \eta\left(1 - \frac{1}{G}\right) \end{aligned} \tag{1.88}$$

As we can deduce from Eq. (1.88), the PAE converges to η in the case of high power gain levels.

1.8.2 Nonlinear FOMs

Here again, since amplifiers are two-port networks, the nonlinear FOMs as described in Section 1.5 are generally applicable. The most common FOMs that are listed in

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

amplifier datasheets are the 1-dB-compression point P_{1dB}, the third-order intercept point IP_3, the saturated output power P_{sat}, the efficiency, and the power added efficiency, of which P_{1dB} and IP_3 have already been described extensively in Section 1.5.

When considering Fig. 1.11, we see that the output power saturates at high input power. The reason, as already explained in Section 1.4, is that the output power cannot be higher than the supplied power, this being the sum of P_{DC} and P_{IN}. Since the power-transfer characteristic at high input powers is not perfectly flat in practical cases, the saturated power is usually determined at a certain compression point, beyond the 1-dB-compression point. A typical approach is to take the output power at the 3-dB-compression point as P_{sat}.

1.8.3 Transient FOMs

The final set of FOMs is related to the transient behavior of amplifiers. It primarily applies to VGAs, which are amplifiers whose gain can be varied by means of a DC control signal. Depending on the design architecture, this control signal can be part of the DC bias supply feeding the transistors within the VGA, or can be applied to a DC bias-dependent component in the matching network (e.g., a diode). In certain cases a VGA can also be composed of a GA followed by a variable attenuator. So the gain of the GA is constant and the attenuation is being changed. Figure 1.32 illustrates how the gain changes on varying the control voltage.

In addition to the amplifier FOMs already discussed, the range of variability of the amplifier gain and the speed of change are key points when describing VGAs. The latter means that the FOMs become dependent on the dynamic behavior of the control signal. For this reason, some typical FOMs from control theory can be applied, such as the slew rate, rise time, settling time, ringing, and overshoot.

1.8.3.1 Slew rate

DEFINITION 1.18 *The slew rate of an amplifier is the maximum rate of change of the output, usually quoted in volts per second.*

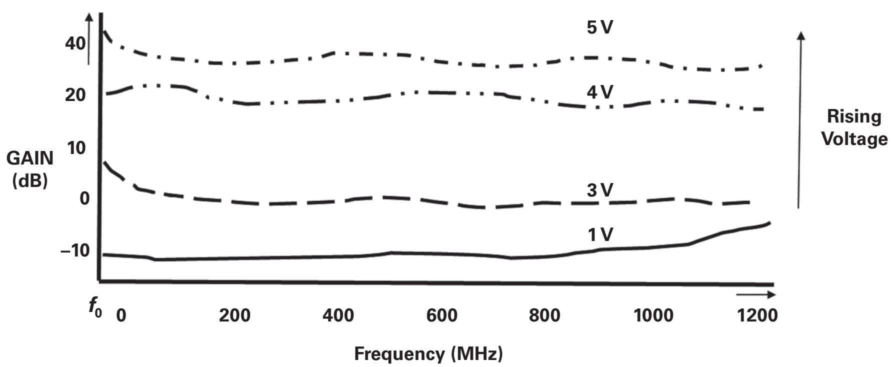

Figure 1.32 A variable-gain amplifier controlled by a DC control bias voltage.

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

Many amplifiers are ultimately slew-rate-limited (typically by the impedance of a drive current having to overcome capacitive effects at some point in the circuit), which sometimes limits the full-power bandwidth to frequencies well below the amplifier's small-signal frequency response.

1.8.3.2 Rise time

DEFINITION 1.19 *The rise time (t_r) of an amplifier is the time taken for the output to change from 10% to 90% of its final level when driven by a step input.*

1.8.3.3 Settling time and ringing

DEFINITION 1.20 *The settling time is the time taken for the output to settle to within a certain percentage of the final value (for instance 0.1%).*

The next definition is ringing. Ringing is the result of overshoot caused by an under-damped circuit.

DEFINITION 1.21 *Ringing refers to an output variation that cycles above and below an amplifier's final value, and leads to a delay in reaching a stable output.*

1.8.3.4 Overshoot

DEFINITION 1.22 *The overshoot is the amount by which the output exceeds its final, steady-state value, in response to a step input.*

Table 1.4 presents a typical datasheet of a VGA. The main difference from a typical amplifier datasheet is related to the range of gain variability and the speed of this change. In case of the example shown in Fig. 1.32, we see that the response time (10% to 90%) is 25 μs and the control range is 30 dB.

1.9 Mixers

As we can deduce from Fig. 1.1, mixers are essential components in wireless transceivers since they up-convert the modulated signal from baseband to RF for wireless transmission, and then also down-convert the received RF signal back to baseband. There are several up-conversion/down-conversion configurations (e.g., homodyne, superheterodyne, ...), among which the most important ones will be described in Chapter 2 in connection to the internal architecture of measurement instrumentation. There are also various design topologies (e.g., single-ended, balanced, double-balanced, ...), the description of which is beyond the scope of this book. The note to make in connection with FOMs, though, is that mixers can be based either on diodes (passive mixers) or on transistors (active mixers). The choice depends on the requirements of the particular transceiver design.

Figure 1.33 represents a down-converting mixer. A mixer is a three-port circuit, with as ports the radio-frequency (RF) port, the local oscillator (LO) port, and the intermediate-frequency (IF) port. The RF signal enters the RF port and gets down-converted to baseband, and then the corresponding baseband (or IF) signal exists at the IF port. The mixer is driven into nonlinear operation by a pump signal, this being

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

Table 1.4 The variable-gain amplifier. The datasheet for a broadband amplifier. The electrical characteristics are response time (10% to 50%) 25 μs, and control voltage 0 to 5 V.

Parameter		Value	Units
Frequency	f_L f_U	10 1200	MHz
Gain	Minimum Typical Flatness Control range	24 34 ±1.5 30	dB
Maximum power	Output 1-dB CP Input (no damage)	+13 +10	dBm
Dynamic range	NF (typical) IP_3 (typical)	15 +25	dB dBm
VSWR	In Out	2.2 2.0	dB
DC power	Voltage Current	15 170	V mA

Maximum ratings

Operating temperature	−20 °C to 71 °C
Storage temperature	−55 °C to 100 °C
DC voltage	17 V

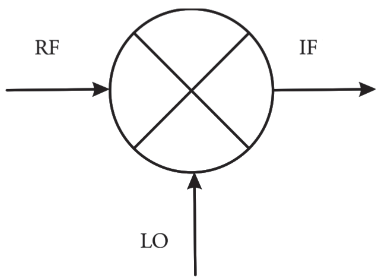

Figure 1.33 The general mixer symbol, where the three ports are identified.

the local oscillator. In the case of an up-converting mixer, the IF port is the input and the RF port acts as the output.

Table 1.5 shows an example datasheet of a mixer. The mixer FOMs will be clarified in the next sections. We make the distinction between two-port and three-port FOMs. Throughout the FOM definitions and expressions, we assume that the mixer is a down-converter. Similar expressions for an up-converting mixer can easily be deduced.

1.9.1 Two-port FOMs

In terms of FOMs, a mixer is often considered as a two-port circuit, because the operation of interest is happening between the RF and IF ports. The modulated signal is

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

Table 1.5 A mixer datasheet

Parameter		Value	Units
Frequency	RF/LO IF	250–3250 DC–800	MHz
Conversion loss	Typical Maximum	6.5 8.5	dB
Isolation LO/RF	Typical Minimum	30 15	dB
Isolation LO/IF	Typical Minimum	10 5	dB
Isolation RF/IF	Typical Minimum	30 15	dB
LO power	Nominal	+17	dBm
1-dB-compression point	Typical	+10	dBm
Input IP_3	Typical	+18	dBm
RF input power	Maximum	100	mW
Impedance	Nominal	50	Ω
Operating temperature		−40 to 85	°C

down-converted between the RF port and IF port, or up-converted between the IF and RF ports, while the signal entering the LO port is kept constant at a high power level. For this reason, the concepts introduced in Section 1.5 are applicable to mixers.

The first FOM is the conversion loss L.

DEFINITION 1.23 *The conversion loss (L) of a down-converting mixer is the ratio of the RF input power and the IF output power.*

Mathematically,

$$L = \frac{P_{\mathrm{AVS,RF}}}{P_{\mathrm{AVN,IF}}} \tag{1.89}$$

where $P_{\mathrm{AVS,RF}}$ is the power of the modulated input signal at the RF carrier frequency, and $P_{\mathrm{AVN,IF}}$ is the output power of the down-converted signal at IF.

In the case of a diode-based mixer, there is always a conversion loss since diodes have no gain. Even using transistors as the nonlinear element in the mixer, there is a conversion loss or at most a small conversion gain. The reason is that the transistors have to be operated in a strongly nonlinear condition, e.g., near to pinch-off operating conditions, which corresponds to low gain. Owing to losses in the matching networks, the overall conversion is often a loss.

The FOMs which are a measure for the linearity of the mixer are the input 1-dB-compression point $\mathrm{InP}_{1\mathrm{dB}}$ and the input third-order intercept point IIP3. These two FOMs have already been introduced in Sections 1.5.1 and 1.5.2, but in the case of

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

mixers they are usually referred to the input, as opposed to what is done in the case of amplifiers, where $P_{1\text{dB}}$ and IP3 are usually referred to the output.

Mixers also have a noise contribution. The corresponding FOM is the single-sideband noise figure (SSBNF). It is defined in an analogous way to the noise figure NF of a linear two-port device (see Section 1.3.2), but now taking into account that the frequency at the output of a mixer is different from the frequency at the input of the mixer.

DEFINITION 1.24 *The single-sideband noise figure (SSBNF) for a mixer is the decibel value of the signal-to-noise ratio at the input of the mixer at RF divided by the signal-to-noise ratio at the output of the mixer at IF.*

Mathematically, the noise factor for a mixer F_{mixer} is

$$
\begin{aligned}
F_{\text{mixer}} &= \frac{S_{\text{I,RF}}/N_{\text{I,RF}}}{S_{\text{O,IF}}/N_{\text{O,IF}}} \\
&= \frac{N_{\text{O,IF}}}{N_{\text{I,RF}}L}
\end{aligned}
\tag{1.90}
$$

where $S_{\text{I,RF}} = P_{\text{AVS,RF}}$ is the signal power at the input of the mixer at RF, $S_{\text{O,IF}} = P_{\text{AVS,IF}}$ is the signal power at the output of the mixer at IF, $N_{\text{I,RF}}$ is the available noise power at the input of the mixer at RF, $N_{\text{O,IF}}$ is the available noise power at the output of the mixer at IF, and L is the conversion loss.

Consequently, the noise figure is given by

$$
\text{SSBNF} = 10\log_{10} F_{\text{mixer}} \tag{1.91}
$$

1.9.2 Three-port FOMs

The next set of FOMs consists of FOMs that are related to the three-port nature of the mixer.

In general, a mixer has mismatches at each of its ports, like any other microwave circuit. As in the case of linear two-port networks (see Section 1.2), the port mismatches of a mixer are represented by the VSWR or return loss RL. Even though the mixer is a nonlinear circuit, and thus the reflection coefficients at each of the three ports may be dependent on the input power, a constant value corresponding to normal operating conditions is listed in mixer datasheets.

The aim of a down-converting mixer is to convert the signal at RF to a signal at IF. Any other frequency components in the signal's spectrum are unwanted, and therefore should be avoided. A distinction is made between on the one hand the leakage of the LO and IF signals to the other ports (see later) and on the other hand the harmonics and intermodulation products generated due to the nonlinear operating mode of the mixer. Owing to the small frequency offset between the RF and LO signals, the out-of-band unwanted spectral components can easily be filtered out. Commercial mixers usually have such filters included in the package already.

Since the LO power is high in order to drive the mixer in nonlinear operation, one should avoid having part of this signal leak to the RF and IF ports, because such signal

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

may damage the circuit block preceding or following the mixer. Similarly, one wants to avoid having part of the IF output power couple back to the RF port. The possibility of leakage from the IF port to the LO port is usually ignored, since the IF power is very small compared with the LO power. In the case of passive circuits and amplifiers, such unwanted coupling between ports is called isolation (see Section 1.8.1). In mixer datasheets, both "isolation" and "leakage" are used. The various leakages are defined as follows.

DEFINITION 1.25 *The LO/RF leakage of a mixer is equal to the ratio of the power at LO frequency at the RF port and the LO power.*

Mathematically,

$$\text{LO/RF leakage} = \frac{P_{\text{RF,LOfreq}}}{P_{\text{LO}}} \tag{1.92}$$

with P_{LO} the input power at the LO port and at LO frequency, and $P_{\text{RF,LOfreq}}$ the output power at the RF port and at LO frequency.

DEFINITION 1.26 *The LO/IF leakage of a mixer is equal to the ratio of the power at LO frequency at the IF port and the LO power.*

Mathematically,

$$\text{LO/IF leakage} = \frac{P_{\text{IF,LOfreq}}}{P_{\text{LO}}} \tag{1.93}$$

with $P_{\text{IF,LOfreq}}$ the output power at the IF port and at LO frequency.

DEFINITION 1.27 *The RF/IF leakage of a mixer is equal to the ratio of the power at IF frequency at the RF port and the IF power.*

Mathematically,

$$\text{IF/RF leakage} = \frac{P_{\text{RF,IFfreq}}}{P_{\text{IF}}} \tag{1.94}$$

with P_{IF} the input power at the IF port and at IF, and $P_{\text{RF,IFfreq}}$ the output power at the RF port and at IF.

The final thing to note is that all of these FOMs are temperature-dependent.

1.10 Oscillators

1.10.1 Oscillator FOMs

Oscillators can come in various forms, so we will include in this chapter various of these flavors. They can be free-running oscillators, voltage-controlled oscillators, and synthesized ones. For all of those types, the oscillator characteristic to be considered as the first and most important FOM is the frequency of operation.

Actually, oscillators are produced to generate a typical signal waveform. In the case of microwave and RF circuits, this waveform is most of the time a sinusoidal signal.

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

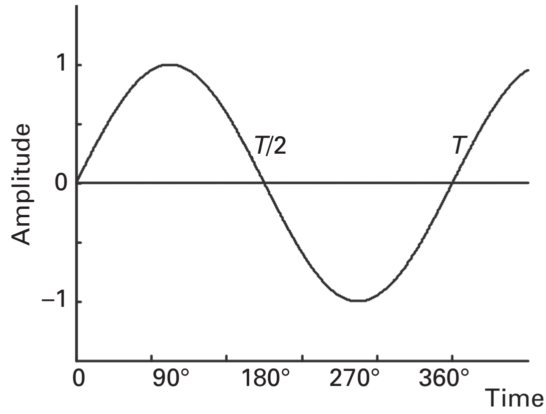

Figure 1.34 A sine-wave oscillator, where the period definition is marked.

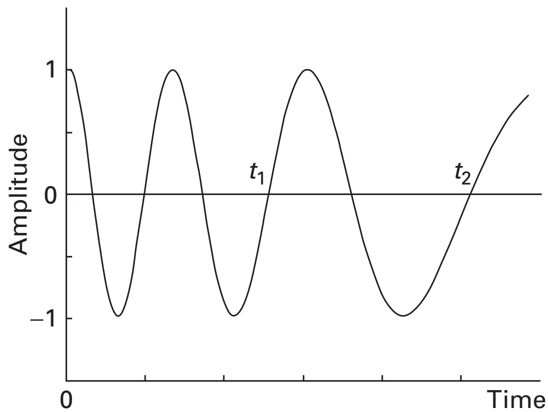

Figure 1.35 Frequency stability, where the change in signal frequency over time can be seen.

A sine-wave signal generator produces nothing other than a voltage that changes as a function of time in a sinusoidal manner, as shown in Fig. 1.34.

In this sine wave we can define a frequency, which is the number of cycles per second at which the waveform repeats itself, and the amplitude of the waveform. Mathematically, it is represented by

$$V(t) = A \sin\left(2\pi \frac{t}{T} + \theta\right) = A \sin(2\pi f + \theta) \tag{1.95}$$

It is expected that an oscillator is a pure sine wave, but most of the time this is not true. The generator is corrupted by other factors that will have a significant impact on the output waveform. For instance, on considering Fig. 1.35, it is clear that an important characteristic of the oscillator is the frequency stability, that is how well the frequency is maintained over time. In fact the term frequency stability encompasses the concepts

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

of random noise, intended and incidental modulation, and any other fluctuations of the output frequency of a device.

1.10.1.1 Frequency stability

DEFINITION 1.28 *Frequency stability is in general the degree to which an oscillating source produces the same frequency value throughout a specified period of time. It is implicit in this general definition of frequency stability that the stability of a given frequency decreases if the wave shape of the signal is anything other than a perfect sine function.*

We can also further present the short-term and long-term stability, which are usually expressed in terms of parts per million per hour, day, week, month, or year. Long-term stability represents phenomena caused by the aging of circuit elements and of the material used in the frequency-determining element. Short-term stability relates to frequency changes of duration less than a few seconds about the nominal frequency. The reader is directed to [10] for more information.

Since we are dealing with an electronic generator, that is, one based on a strong nonlinearity, the generation of harmonics, as explained in Section 1.4.1, is also very important, since it can create harmonic components in the output signal.

1.10.1.2 Phase noise

The presence of noise in the electronic components can also create another non-ideal behavior of the sine wave. This is most of the time accounted for using the FOM called phase noise (Fig. 1.36), since the noise will behave as a corrupted phase in the output signal. In this case the output sine wave can be represented as

$$V(t) = [A + \epsilon(t)]\sin(2\pi f + \phi(t)) \tag{1.96}$$

This phase-noise FOM is the term most widely used to describe the frequency stability's characteristic randomness. There are also other terms, such as spectral purity, which refers to the ratio of signal power to phase-noise sideband power.

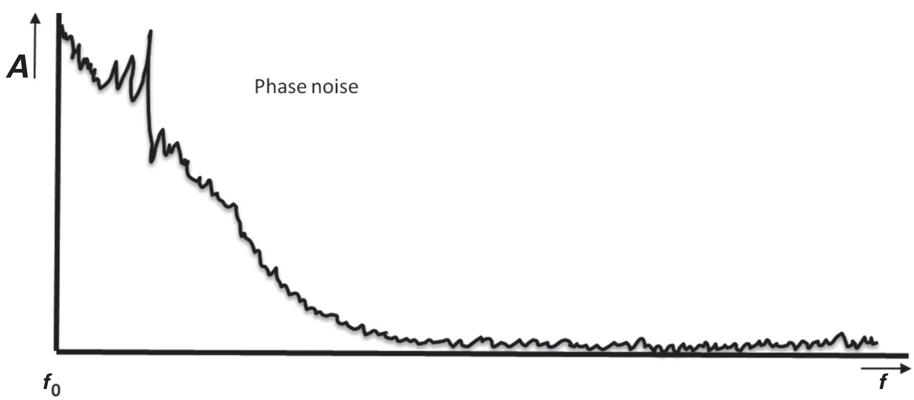

Figure 1.36 Phase-noise representation.

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

Table 1.6 An oscillator data sheet

Parameter	Test condition	Minimum	Typical	Maximum	Units
Nominal frequency[a]	LVDS/CML/LVPECL	10		945	MHz
	CMOS	10		160	
Temperature stability	$T_A = -40$ to $+85\,^{\circ}$C	−20 −50 −100		+20 +50 +100	ppm
Absolute pull range	±25		±345		ppm
Aging	Frequency drift over first year			±3	ppm
	Frequency drift over 15-year life			±10	
Power-up time[b]				10	ms

[a] Nominal output frequency set by $V_{CNOM} = V_{DD}/2$.
[b] Time from power-up or tri-state mode to f_0.

For a correct evaluation of frequency stability, and thus of phase noise, we should calculate the power spectral density of the waveform. In this case, for the waveform presented by Eq. (1.96), the power spectral density is

$$G_{\text{sideband}} = \int_{-\infty}^{+\infty} S_g(f)df \tag{1.97}$$

where $S_g(f)$ represents the two-sided spectral density of fluctuations of the output waveform.

Actually Glaze, in his chapters in [10], discussed a possible definition of frequency stability that relates the sideband power of phase fluctuations to the carrier power level. This quantity is called $\pounds(f)$.

DEFINITION 1.29 *$\pounds(f)$ is defined as the ratio of the power in one sideband, referred to the input carrier frequency on a per-hertz-of-bandwidth spectral-density basis, to the total signal power, at Fourier frequency difference δf from the carrier, per device. In fact, it is a normalized frequency-domain measure of phase-fluctuation sidebands, expressed in decibels relative to the carrier per hertz:*

$$\pounds(f) = \frac{\text{power density (one phase-modulation sideband)}}{\text{carrier power}} \tag{1.98}$$

On looking at a typical oscillator datasheet from a manufacturer, it is clear that most of the FOMs explained here are represented in the datasheet, as depicted in Table 1.6.

1.11 Frequency-multiplier FOMs

Frequency multipliers are circuits that convert an input signal at frequency f_0 into an output signal at a frequency that is a multiple of f_0. Their use in wireless transceivers is usually in combination with oscillators. The higher the required LO frequency, the more

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

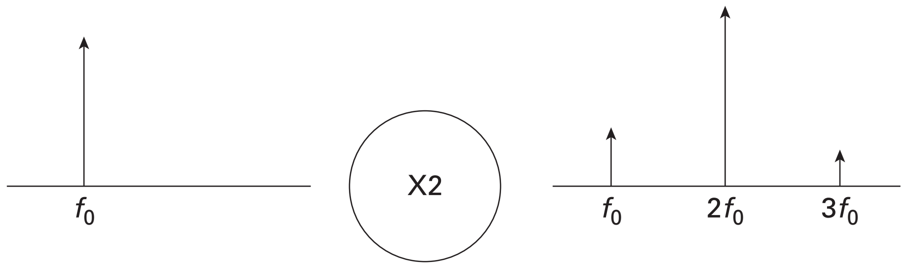

Figure 1.37 Frequency doubling, showing the input and output spectrum components.

difficult it is to design and fabricate oscillators with low phase noise. So the approach used is to take a very good oscillator at a lower LO frequency and then up-convert this frequency by means of a frequency multiplier. Typical multiplication factors in practical designs are in the range 2–4, because the higher the multiplication factor the higher the conversion loss (see later for the definition). Using a frequency multiplier does increase the phase noise, by a factor equal to the multiplication factor, but the resulting phase noise is still typically lower than that of an oscillator at the higher frequency.

Figure 1.37 shows schematically a frequency doubler. In this example, the input signal is a single-tone signal at frequency f_0, and the intended output signal is at $2f_0$. Since the output signal of a linear circuit is by definition at the same frequency as the frequency of the input signal (see also Section 1.2), any frequency-converting circuit (frequency multipliers but also mixers) is a nonlinear circuit, and therefore unwanted spectral components are generated at the output as well, e.g., at f_0 and $3f_0$. The latter are characterized by the frequency-multiplier FOMs which we will define next.

The first FOM of a frequency multiplier is the conversion loss, which is defined similarly to that in the case of mixers (see Section 1.9.1). The definition, with n the multiplication factor, is as follows.

DEFINITION 1.30 *The conversion loss (L) of an n times frequency multiplier is the ratio of the input power at frequency f_0 and the output power at frequency nf_0.*

Mathematically,

$$L = \frac{P_{\mathrm{AVS},f_0}}{P_{\mathrm{AVN},nf_0}} \tag{1.99}$$

As in the case of mixers, there is usually no conversion gain, even when the frequency multiplier is based on transistors.

At the output, we want to have a clear output signal at the multiplied frequency. So both the f_0 spectral component and unwanted harmonics $(mf_0|_{m\neq n})$ should be suppressed. This is achieved by design, e.g., by using a circuit architecture that automatically suppresses the f_0 component, and/or by incorporating dedicated filters in the circuit's package. The corresponding FOMs are the fundamental and harmonic rejections.

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

Table 1.7 A frequency-multiplier datasheet

Multiplication factor	Frequency (GHz) f_1 In	Frequency (GHz) f_2 Out	Input power Min.	Input power Max.	Conversion loss Typ.	Conversion loss Max.	Harmonic output f_1 Typ.	Harmonic output f_1 Min.	Harmonic output f_3 Typ.	Harmonic output f_3 Min.	Harmonic output f_4 Typ.	Harmonic output f_4 Min.
			13	16	11.5	15	30	18	35	23	25	15
2	5–8	10–16	10	13	15	18	21	14	30	18	20	13
	8–10	16–20	13	16	12	15	33	20	27	17	50	35
			10	13	15	18.5	30	16	23	16	40	30

Min., minimum; Max., maximum; Typ., typical.

DEFINITION 1.31 *The fundamental rejection is the ratio of the output power at frequency nf_0 and the output power at frequency f_0.*

Mathematically,

$$\text{fundamental rejection} = \frac{P_{\mathrm{AVN},nf_0}}{P_{\mathrm{AVN},f_0}} \tag{1.100}$$

The rejection is usually expressed in dBc, which is the difference, expressed in decibels, relative to the wanted signal, or carrier.

DEFINITION 1.32 *The harmonic rejection is the ratio of the output power at frequency nf_0 and the output power at frequency $f_{m0}|_{m\neq n\neq 0}$.*

Mathematically,

$$\text{harmonic rejection} = \frac{P_{\mathrm{AVN},nf_0}}{P_{\mathrm{AVN},mf_0}\big|_{m\neq n\neq 0}} \tag{1.101}$$

Finally, as with all electronic circuits, the characteristics of a frequency multiplier are temperature dependent.

An example datasheet of a frequency doubler is presented in Table 1.7.

1.12 Digital converters

Finally, this chapter would not be complete if digital converters were not included.

Digital converters, either analog to digital (ADC) or digital to analog (DAC), are becoming a key component in radio and wireless communication circuits and systems. Actually, the advent of software-defined radio (SDR) and/or cognitive radio (CR) is moving this technology faster to higher frequencies and thus to new characterization procedures.

An ADC converts an analog signal to digital quantities, by operation in two axes over the continuous time-domain analog signal. These two axes correspond to a sampling in time and a sampling in amplitude, usually called quantization [11].

Sampling in time corresponds to sampling the time-domain signal at some discrete points with a sampling frequency that should obey the Nyquist frequency. Sampling in

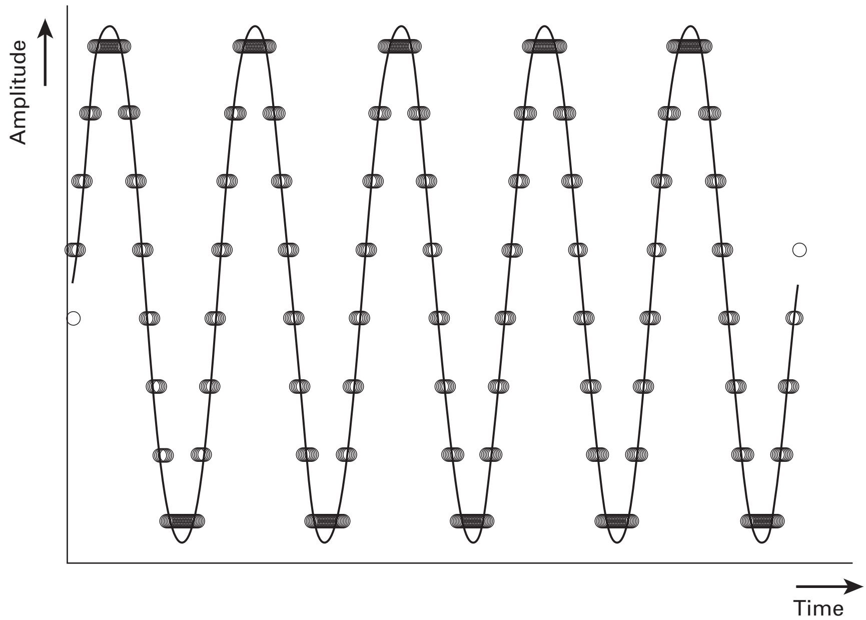

Figure 1.38 Sampling and quantization procedures.

amplitude corresponds to sampling the amplitude axis at discrete points also. Quantization and sampling at the same time corresponds to picking up the sampled points both in time and amplitude and constraining them to fit a pre-determined matrix. This fact leads to several important FOMs, since the time sample can and will generate aliasing errors, while the quantization will generate a minimum-noise floor that will severely degrade the output digital signal, by adding quantization noise to it. This is illustrated in Fig. 1.38.

The study of sampling and quantization noise is beyond the scope of this book, but the reader is directed to reference [11] for more information.

Traditional ADC/DAC applications are at low frequencies, and thus some of the characterizations and FOMs used are low-frequency-related ones. For instance, we can have offset errors, gain errors, integral nonlinearities, differential nonlinearities, and special FOMs related to the quantization noise.

1.12.1 Figures of merit

The more traditional FOMs can be defined as follows.

DEFINITION 1.33 *The gain error is the difference between the measured and ideal full-scale input voltage range of the ADC.*

DEFINITION 1.34 *The offset error is the DC offset imposed on the input signal by the ADC, reported in terms of LSB (codes).*

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

1.12.1.1 ADC time behavior

Some of the FOMs related to the time behavior of the ADC include the following:

DEFINITION 1.35 *The aperture uncertainty is related to the signal jitter, which is the sample-to-sample variation in aperture delay.*

DEFINITION 1.36 *The Encode pulse width/duty cycle. Pulse width high stands for the minimum amount of time the Encode pulse should be left in the Logic 1 state to achieve the rated performance, while pulse width low is the minimum time the Encode pulse should be left in the low state.*

DEFINITION 1.37 *The maximum conversion rate is the maximum Encode rate at which the image spur calibration degrades by no more than 1 dB.*

DEFINITION 1.38 *The minimum conversion rate is the minimum Encode rate at which the image spur calibration degrades by no more than 1 dB.*

DEFINITION 1.39 *The output propagation delay is the delay between a differential crossing of one Encode and another Encode (or zero crossing of a single-ended Encode).*

DEFINITION 1.40 *The pipeline latency is the number of clock cycles by which the output data lags relative to the corresponding clock cycle.*

DEFINITION 1.41 *The signal-to-noise ratio for ADCs (SNR_{ADC}) is the ratio of the RMS signal amplitude (set at 1 dB below full scale) to the RMS value of the sum of all other spectral components, excluding harmonics and DC.*

For an ideal ADC one can represent this value as

$$\mathrm{SNR}_{\mathrm{ADC}} = 6.02N + 1.76\,\mathrm{dB} \tag{1.102}$$

where N is the number of bits of the ADC.

DEFINITION 1.42 *The effective number of bits (ENOB) corresponds to the number of bits that we can have when considering not the ideal but the measured SNR.*

The ENOB is calculated from the measured SNR as follows:

$$\mathrm{ENOB} = \frac{\mathrm{SNR}_{\mathrm{MEASURED}} - 1.76}{6.02} \tag{1.103}$$

1.12.1.2 ADC nonlinear behavior

Some FOMs are also related to the nonlinear behavior of the ADC.

DEFINITION 1.43 *The differential nonlinearity is the type of nonlinearity that corresponds to the deviation of any code width from an ideal one-least-significant-bit (LSB) step.*

DEFINITION 1.44 *The integral nonlinearity is the deviation of the transfer function from a reference line measured in fractions of 1 LSB using a best straight line determined by a least-square-curve fit.*

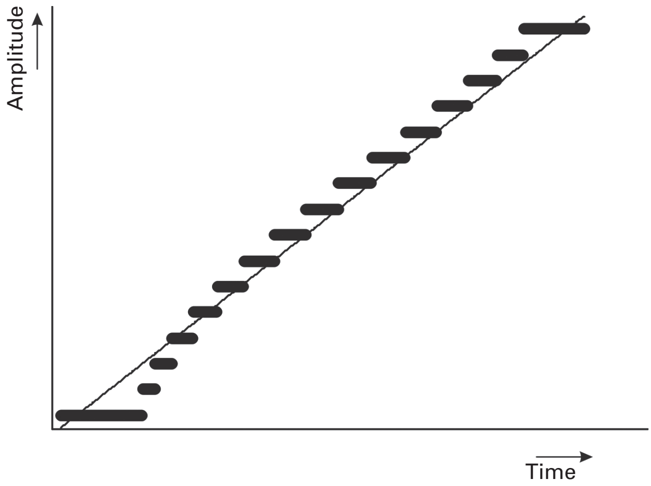

Figure 1.39 The impact of nonlinear behavior on the ADC quantization curves.

These two FOMs can be seen in Fig. 1.39.

From a wireless point of view, nonlinearity can also be characterized by using typical measures of intermodulation and harmonics as previously presented in Section 1.5.

DEFINITION 1.45 *The second-harmonic distortion is the ratio of the RMS signal amplitude to the RMS value of the second-harmonic component, reported in dBFS (dBFS means dB full scale, which means x dB below the full scale).*

DEFINITION 1.46 *The third-harmonic distortion is the ratio of the RMS signal amplitude to the RMS value of the third-harmonic component, reported in dBFS.*

DEFINITION 1.47 *The signal-to-noise-and-distortion (SINAD) ratio is the ratio of the RMS signal amplitude (set 1 dB below full scale) to the RMS value of the sum of all other spectral components, including harmonics, but excluding DC and image spur.*

DEFINITION 1.48 *The two-tone intermodulation distortion rejection is the ratio of the RMS value of either input tone to the RMS value of the worst third-order intermodulation product, reported in dBc.*

Sometimes the characterization of analog converters also includes several FOMs related to any spurious signals that are visible at the output signal. Those FOMs include the following.

DEFINITION 1.49 *The spurious-free dynamic range (SFDR) is the ratio of the RMS signal amplitude to the RMS value of the peak spurious spectral component, except the image spur. The peak spurious component may, but need not, be a harmonic. It can be*

https://doi.org/10.1017/CBO9780511791512.003 Published online by Cambridge University Press

Table 1.8 An ADC datasheet (for $V_A = 3.7$ V, $V_C = 1.5$ V, ENCODE = 400 MSPS, and $0\,°C \le T \le 60\,°C$, unless specified otherwise)

Parameter		Case T	Minimum	Typical	Maximum	Units
Dynamic performance						
SNR						
Analog input	10 MHz	Full	62	64		dBFS
	70 MHz	Full	61.5	63.5		dBFS
at −1.0 dBFS	128 MHz	Full	61.5	63.5		dBFS
	175 MHz	Full	61.5	63.5		dBFS
SINAD ratio						
Analog input	10 MHz	Full	59	63.5		dBFS
	70 MHz	Full	58.5	63		dBFS
at −1.0 dBFS	128 MHz	Full	57.5	61.5		dBFS
	175 MHz	Full	55	60		dBFS
SFDR						
Analog input	10 MHz	Full	69	85		dBFS
	70 MHz	Full	69	80		dBFS
at −1.0 dBFS	128 MHz	Full	66	72		dBFS
	175 MHz	Full	62	68		dBFS
Image Spur						
Analog input	10 MHz	Full	60	75		dBFS
	70 MHz	Full	60	72		dBFS
at −1.0 dBFS	128 MHz	Full	60	66		dBFS
	175 MHz	Full	57	63		dBFS
Offset spur						
Analog input at −1.0 dBFS		60 °C		65		dBFS
Two-tone IMD						
F_1, F_2 at −6 dBFS		60 °C		−75		dBc
Analog input						
Frequency range		Full	10		175	MHz
Digital Input DR_EN						
Minimum time, low		Full	5.0			ns
Switching specifications						
Conversion rate		Full	396	400	404	MSPS
Encode pulse width high		60 °C		1.25		ns
Encode pulse width low		60 °C		1.25		ns

reported in dBc (that is, it degrades as the signal level is lowered) or dBFS (always related back to converter full-scale).

Other FOMs consider the VSWR, IMR, ACPR, NPR, etc. which were presented in Sections 1.3 and 1.5 on linear and nonlinear FOMs.

Table 1.8 presents a typical ADC datasheet. From the datasheet it is clear that the FOMs of an ADC related to the RF part are focused specifically on the signal-to-noise ratio (SNR), or, if we include distortion, the signal-to-noise-and-distortion (SINAD) ratio, or, if we include all types of possible spurious we can have the spurious–free dynamic range, which is related not only to the noise, but also to the maximum signal allowed before clipping. All these characteristics are expressed in dBFS.

Problems

1.1 Prove the expressions for Γ_{IN} and Γ_{OUT} in Eq. (1.82).

1.2 Knowing that a receiver has a sensitivity of −100 dBm, what should be the maximum power an out-of-band interferer can have in order to maintain signal quality? Consider that the receiver has a gain of 10 dB, that the useful signal has a bandwidth of 100 KHz, an NF of 2 dB, 1-dB-compression point of 30 dBm, and an IP_3 of 47 dBm, and that signals out of band have an extra 50 dB rejection due to the use of an input filter.

1.3 In your view, what is preferable in a transceiver: to have high gain and low NF at the first stage of the receiver, or to have high gain and high IP_3?

1.4 In a receiver chain, the first block is a cable, followed by an amplifier and a mixer. Considering that the loss in the cable is 3 dB, the amplifier gain is 10 dB with an NF of 2 dB, and the mixer has an insertion loss of 10 dB, calculate the overall NF.

1.5 Explain why we use a low-noise amplifier in the receiver chain and a power amplifier in the transmitter chain.

1.6 What is the main impact of phase noise?

1.7 How can you evaluate the impact of phase noise on digital systems?

1.8 What is the difference between gain and underlying linear gain?

1.9 Comment on the relationship between the 1-dB-compression point and IP_3.

1.10 Explain why the EVM can be calculated using the SNR, not the ACPR.

1.11 What are the main limitations of digital converters?

1.12 If we have a system with NPR = 10 dB, what will the EVM be?

References

[1] M. B. Steer, *Microwave and RF Design: A Systems Approach*. Herndon, VA: SciTech Publishing, 2010.

[2] D. M. Pozar, *Microwave Engineering*. New York: John Wiley, 2005.

[3] S. Maas, *Noise in Linear and Nonlinear RF and Microwave Circuits*. Norwood, MA: Artech House, 2005.

[4] J. C. Pedro and N. B. Carvalho, *Intermodulation Distortion in Microwave and Wireless Circuits*. New York: Artech House, 2003.

[5] S. Maas, *Nonlinear Microwave and RF Circuits*. Norwood, MA: Artech House, 2003.

[6] N. B. Carvalho and J. C. Pedro, "A comprehensive explanation of distortion side band asymmetries," *IEEE Trans. Microwave Theory Tech.*, vol. 50, no. 9, pp. 2090–2101, Sep. 2002.

[7] N. Carvalho, K. Remley, D. Schreurs, and K. Gard, "Multisine signals for wireless system test and design," *IEEE Microwave Mag.*, vol. 9, no. 3, pp. 122–138, Jun. 2008.

[8] K. Gharaibeh, K. Gard, and M. Steer, "Accurate estimation of digital communication system metrics – SNR, EVM and ϱ in a nonlinear amplifier environment," in *ARTFG Microwave Measurements Conference*, Honolulu, Hawaii, Jun. 2004.

[9] F.-L. Luo, *Digital Front-End in Wireless Communications and Broadcasting, Circuits and Signal Processing*. Cambridge: Cambridge University Press, 2011.

[10] D. Sullivan, D. Allan, D. Howe, and F. Walls (eds.), *Characterization of Clocks and Oscillators*. Boulder, CO: National Institute of Standards and Technology, 1990.

[11] H. H. Nguyen and E. Shwedyk, *A First Course in Digital Communications*. Cambridge: Cambridge University Press, 2009, Chapter 4.