

1

Introduction

A fundamental design concern for system designers of wireless communication systems is modeling distortion introduced by the nonlinear behavior of the devices incorporated in the design. Nonlinearity in wireless communication systems usually exists in the RF front ends and is produced by nonlinear devices such as power amplifiers, low-noise amplifiers, mixers, etc. Nonlinearity is responsible for introducing signal components that contribute to degrading system performance in a similar fashion to system noise, interference or channel impairments.

In order to understand the effects of nonlinear distortion on the performance of wireless communication systems, it is important to understand the architecture of these systems and the components that are responsible for introducing nonlinear distortion. On the other hand, it is also important to understand how these systems can be modeled and simulated. Modeling and simulation of nonlinear systems is an important step towards the efficient design of modern communication systems.

In this chapter, an overview of nonlinearity and nonlinear distortion in wireless systems is given. The common sources of nonlinearity in wireless communication system are presented. Then, the concept of nonlinear distortion produced by the nonlinear behavior and its relationship to the performance of wireless systems is given in the subsequent sections. In the last section of this chapter, an overview of the most common modeling and simulations approaches of nonlinear systems and circuits are presented that will serve as an introduction to subsequent chapters that discuss modeling and simulation of nonlinear distortion.

1.1 Nonlinearity in Wireless Communication Systems

Nonlinearity in wireless systems is originated in the nonlinear devices incorporate in the design of the transmitter and the receiver. The main blocks that introduces nonlinear distortion are mixers, power amplifiers in wireless transmitters and low-noise amplifiers in wireless receivers.

1.1.1 Power Amplifiers

Power Amplifiers (PAs) are devices that are used at the end of the transmitter chain in order to produce a signal with a power suitable for transmission through an antenna. In a wireless transmitter, a baseband signal sent to a PA that is connected to an antenna through a matching circuit after being modulated and up-converted. The PA performs power amplification by multiplying the signal by a gain factor that results in an amplified signal whose power is much higher than the input signal. Ideally, a PA has constant gain across all input powers, however, a practical PA has a maximum output power that is determined by the DC input power. Hence, as this limit is approached gradually, the apparent gain of the PA decreases with increasing the input power.

Therefore, depending on the input operating power, the PA is considered linear if it is operated in a power range within the linear amplification range of its characteristics. If the amplifier is operated close to or within the saturation region of its characteristics, then it is considered nonlinear. It is usually desirable to operate the PA near its saturation region in order to obtain maximum power efficiency; however, this means that nonlinear distortion is introduced at the output of the PA, which is undesirable specially when the input signal has a varying amplitude.

PAs are characterized by their gain versus input power curves for single-tone input that are known as Amplitude Modulation–Amplitude Modulation (AM–AM) conversion characteristics. In a typical AM–AM characteristics, the gain of the PA remains constant (output power increases linearly) with increasing input power up to a saturation point where the gain drops (output power remains constant with increasing input power). This is due to the input/output characteristics of the active device (transistor) incorporated in the PA design. Saturation of the PA characteristics is a manifestation of the PA nonlinearity where the output power does not follow the input power by a constant gain. Another manifestation of nonlinearity is the phase characteristics of the PA where the phase of the output signal deviates from the input phase by an angle that depends on the input signal power. These phase characteristics are called the Amplitude Modulation–Phase Modulation (AM–PM) characteristics and they measure the phase distortion introduced by a PA. AM–PM characteristics are mainly due to voltage dependent collector capacitance (caused by a varying depletion layer width) (Cylan, 2005).

The AM–AM and AM–PM characteristics can be formulated by considering a single tone signal of the form

$$x(t) = R \cos(2\pi f_0 t + \psi) \quad (1.1)$$

where R is the signal amplitude, f_0 is the tone frequency and ψ is its phase. The output of nonlinearity can be expressed as

$$y(t) = F[R] \cos(2\pi f_0 t + \psi + \Phi(R)) \quad (1.2)$$

where $F[\cdot]$ is the amplitude distortion as a function of the input amplitude and represent the AM–AM characteristics of the nonlinear device and $\Phi[\cdot]$ is the phase distortion as a function of the input amplitude and represents the AM–PM characteristics. Figure 1.1 shows a typical AM–AM and AM–PM characteristics of a PA.

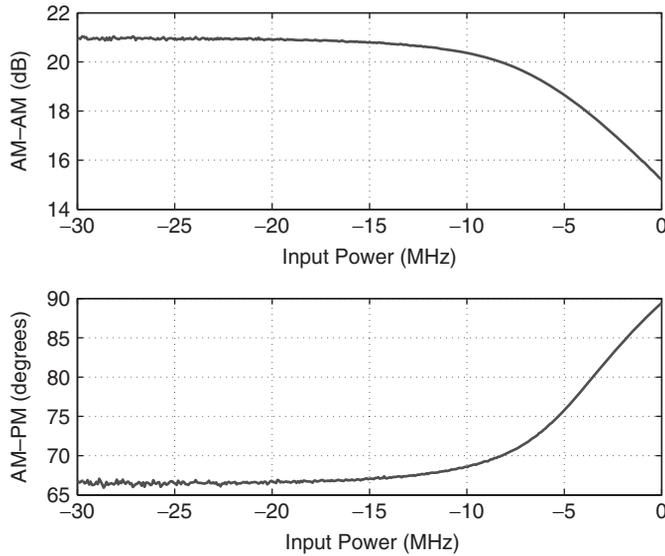


Figure 1.1 AM-AM and AM-PM conversions. Reproduced by permission of © 2002 Artech House.

1.1.1.1 Memory Effects

The AM-AM and AM-PM characteristics define the PA linearity at a single frequency. However, the PA characteristics may differ at different frequencies. This is due to the memory effects of the PA that are usually caused by the time constants in the biasing circuit or are due to the impedance mismatch with the amplifier. Figure 1.2 shows the

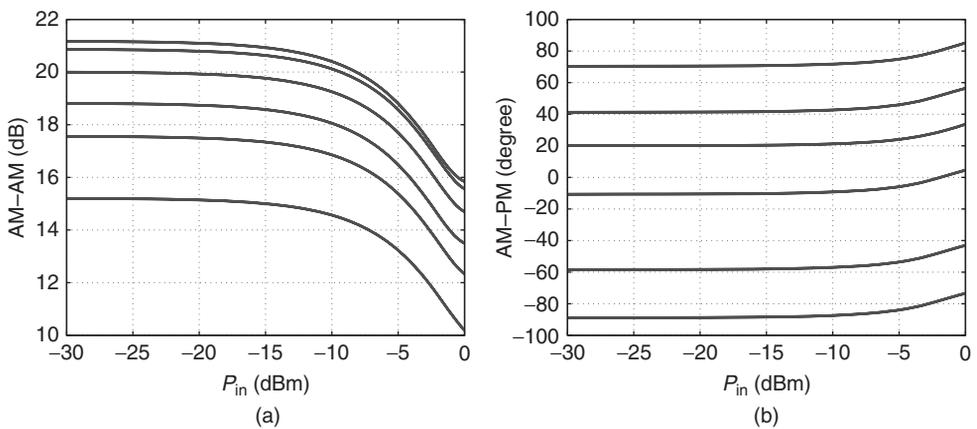


Figure 1.2 AM-AM conversion at different frequencies.

AM–AM and AM–PM of a PA at different frequencies. In fact, the existence of the AM–PM characteristics indicates that the PA has memory since a memoryless PA has only AM–AM characteristics. However, if the time constants of the memory are smaller than the maximum signal envelope frequency, the system is called a quasi-memoryless system. In general, if the bandwidth of the PA is much larger than the modulation bandwidth of the signal, the PA can be considered as memoryless. Memory effects can manifest themselves as hysteresis in the time domain, which causes the intermodulation products of a two-tone test to be asymmetrical in the frequency domain.

The most common approaches to modeling the wideband behavior (i.e. memory effects) of nonlinear amplifiers are those based on Volterra series analysis. Volterra series analysis represents an analytical approach to modeling nonlinearity since it represents nonlinearity in a similar way in which Taylor series does for analytic functions. A general Volterra series model of a nonlinear system is described by the following functional expansion of continuous functions (Lunsford, 1993):

$$y(t) = \sum_{n=1}^{\infty} F_n(x(t)) = \sum_{n=0}^{\infty} y_n(t) \quad (1.3)$$

where $F_n(x(t))$ is the Volterra functional and is defined as

$$F_n(x(t)) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\lambda_1, \dots, \lambda_n) \prod_{i=1}^n x(\lambda_i) d\lambda_i \quad (1.4)$$

where $h_n(\lambda_1, \dots, \lambda_n)$ is the n -dimensional Volterra kernel that can be symmetric without loss of generality and leads to a unique set of Volterra kernels. Other models for nonlinear systems with memory are discussed in Chapter 3.

1.1.1.2 PA Classes

PA designs fall into a number of classes according to their biasing conditions. Different classes have different nonlinear characteristics and hence, different power efficiencies. The most common PA classes are Class A, B, AB, C, D, E and F, where Class A, B and AB being linear (with low power efficiency) whereas the others are nonlinear (with high power efficiency). The PA class is distinguished by its biasing conditions that are chosen to give a desired PA linearity at the expense of its efficiency (Briffa, 1996; Cripps, 2000).

Higher positive bias voltages provide a higher linear region of operation. However, positive bias voltages (such as class A operations) lead to higher conduction angles, which means low power efficiency. On the other hand, negative bias voltages (such as with class C operation) lead to lower conduction angles but this requires a higher input drive level to maintain power efficiency. This makes the PA operate near saturation, which means nonlinear operation. Figure 1.3 shows typical RF output power and power efficiency of a PA versus conduction angle where different classes are defined (Cripps, 2000).

1.1.2 Low-Noise Amplifiers (LNAs)

The low-noise amplifier is the first block in any RF receiver and is responsible for the amplification of the received signal, which usually has very low power. An LNA in an

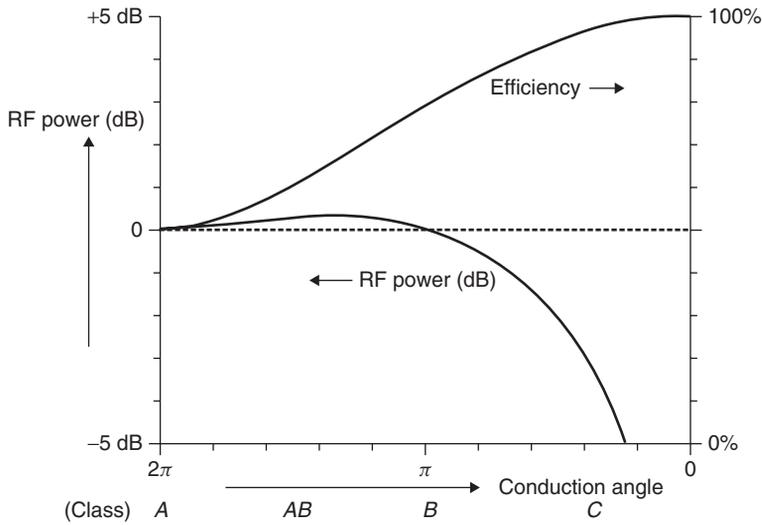


Figure 1.3 PA efficiency vs. conduction angle (Cripps, 2000). Reproduced by permission of © 2000 Artech House.

RF receiver, therefore, plays an important role in the quality of the reception process, and hence the design of LNAs represents a limiting factor in the performance of the overall communication system.

Since an LNA in an RF wireless receiver is required to amplify weak signals in the presence of noise, the design of an LNA requires a tradeoff of a number of factors such as linearity, noise performance, stability, power consumption and complexity. Linearity and noise performance are responsible for determining the dynamic range of the amplifier where the minimum and maximum allowed signals levels are determined. An LNA is designed with its gain compression (due to nonlinearity) determined by the maximum received signal level expected in a certain application. On the other hand, an LNA is designed with a noise performance such that its added noise is below the minimum expected received signal level. Therefore, the main difference between LNAs and PAs is that in PA design noise performance is not an issue since PAs usually operate with input signal powers that are much higher than the inherent noise of the PA circuit. This means that the LNA nonlinearity is linked to the noise performance of its circuit where a tradeoff must be made. On the other hand, the interaction of received signal with channel noise by the nonlinear amplification in an LNA represents a source of distortion introduced inside the band of the received signal (Chabrak, 2006; Gharaibeh, 2009).

Noise performance of an LNA is quantified in terms of the Noise Figure (NF) that measures the amount of noise that is added by the LNA. NF of an LNA is defined as the ratio of the Signal-to-Noise Ratio at the input (SNR_i) to the Signal to Noise Ratio at the output (SNR_o). However, the nonlinearity of LNAs hinders the simple evaluation of their NF. The interaction of signal and noise by nonlinearity results in many output signal components that may be correlated to each other and this complicates the evaluation of the SNR_o . On the other hand, since nonlinear distortion adds to system noise, NF of nonlinear

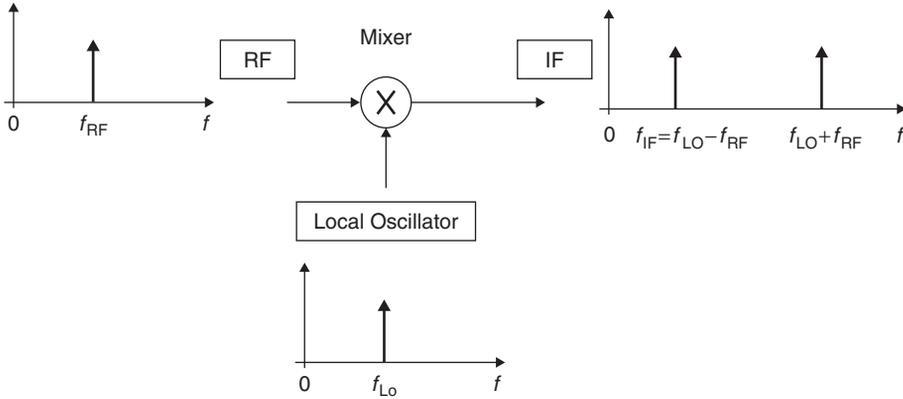


Figure 1.4 A conceptual diagram of an ideal mixer (Cripps, 2000).

amplifiers is highly dependent on the input power because signal distortion increases as the amplifier is driven close to saturation.

1.1.3 Mixers

A mixer is a product device used for frequency up- and down-conversion process at both the transmitter and receiver. A mixer is therefore, inherently nonlinear for normal operation, however, this nonlinearity may result in unwanted effects. In general, the main challenge in the design of mixers is to maintain a second-order nonlinearity needed for the mixing process especially with direct conversion transmitters or receivers. However, this is not easy to achieve as mixers usually exhibit higher-order nonlinearity in addition to the desired second order nonlinearity (Sheng *et al.*, 2003; Terrovitis and Meyer, 2000). Figure 1.4 shows a conceptual diagram of an ideal mixer that operates as downconverter. An ideal mixer is a three port system where the Radio Frequency port (RF) and the Local Oscillator port (LO) are its inputs, while the Intermediate Frequency port (IF) is its output. In this mode of operation and with single-tone inputs, the mixer multiplies the RF-signal with the LO-signal where two sinusoids are generated at the IF output port; one at the sum frequency and another at the difference frequency. A mixer is non ideal if it comprises higher-order nonlinearities (above 2), which means that its output will consist of other spectral components than the two produced by the ideal mixer (Vandermot *et al.*, 2006).

1.2 Nonlinear Distortion in Wireless Systems

As discussed in the previous sections, nonlinearity in wireless communication systems is originated in the PAs, LNAs and mixers. This nonlinearity limits the delivered output power because of the compression nonlinear characteristics and also introduces unwanted signal components at the output of the nonlinear device. These unwanted signal components are called “nonlinear distortion” that is manifested as harmonics at multiples of the fundamental frequencies when the input signal consists of discrete tones and, as spectral regrowth when the input signal spectrum has a finite bandwidth.

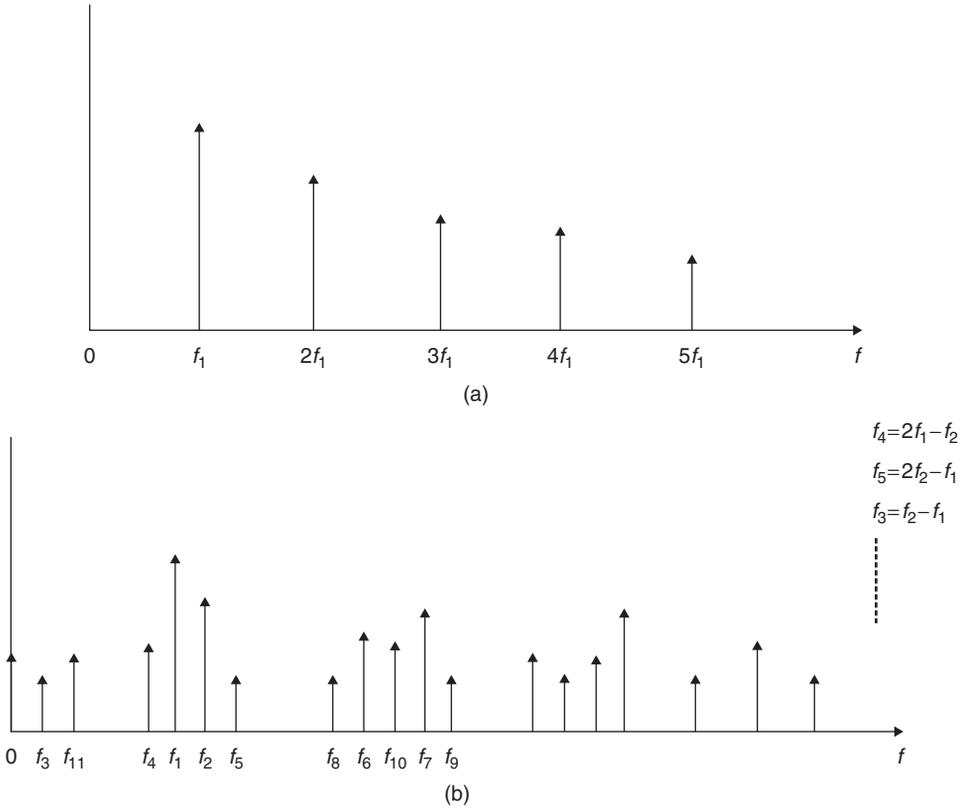


Figure 1.5 Spectra of the nonlinear response to (a) single-tone signal and (b) two-tone signal (Rhyne and Steer, 1987).

Understanding nonlinear distortion is usually based on studying the response of nonlinearity to single or multiple sinusoids (tones) (Gharaibeh *et al.*, 2006; Pedro and de-Carvalho, 2001; Rhyne and Steer, 1987; Vandermot *et al.*, 2006). Figure 1.5 shows the nonlinear response to single- and two-tone signals (Rhyne and Steer, 1987). With a single-tone signal, nonlinearity results in compression of the signal amplitude and the introduction of harmonics at multiples of the fundamental frequency. With two-tone signals nonlinearity results in additional signal components called intermodulation products (IP) that fall very close to the fundamental tones and represent the interaction between the input sinusoids by nonlinearity. In the case of multitone signals, nonlinear distortion is manifested as a large number of intermodulation products that fall within and outside the bandwidth of the input signal. The distortion components that fall inside the bandwidth of the input signal are termed “in-band distortion” and those that fall outside the bandwidth of the signal are termed “out-of-band distortion”. The nonlinear distortion of multiple tone signals is usually used as an approximation of the nonlinear distortion of finite-bandwidth signals which can be approximated by a large number of tones with different amplitudes, phases and frequencies.

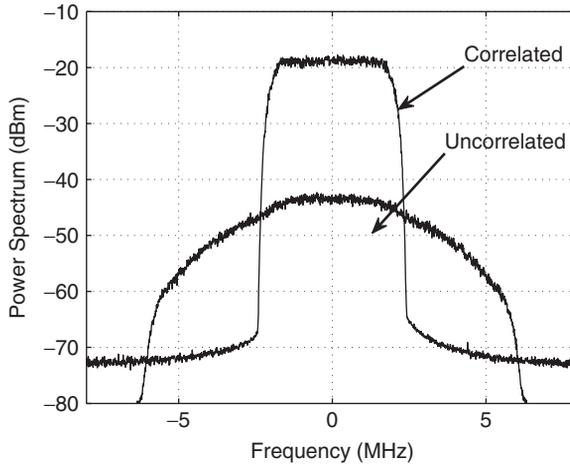


Figure 1.6 Spectrum of the nonlinear response to a digitally modulated signal partitioned into linear and spectral regrowth components.

With digitally modulated signals, nonlinearity results in two main impairments to the output spectrum. The first is gain compression that results in reduced output power relative to linear operation of the nonlinear device, and the second is called spectral regrowth that consists of distortion components that appear inside and outside the bandwidth of the input signal. Figure 1.6 shows the power spectrum at the output of a nonlinearity partitioned into linear with gain compression and spectral regrowth components. The in-band portion of spectral regrowth results in the degradation of system performance while the out-of-band component results in the degradation of the performance of other systems that use adjacent frequency channels.

In the following subsections, the main impairments that nonlinear distortion cause to wireless communications systems are presented.

1.2.1 Adjacent-Channel Interference

Adjacent-Channel Interference (ACI) is a well-known manifestation of nonlinear behavior in wireless systems. ACI is an irreducible out-of-band component of spectral regrowth that is responsible for introducing interference in the adjacent frequency channels in systems that use Frequency Division Multiplexing (FDM) as shown in Figure 1.7. In old cellular systems, such as analogue systems and Global System for Mobile (GSM) systems, frequency planning was required to cope with interference so that the same frequency channels are not used in neighboring cells. Modern wireless systems use Code Division Multiple Access (CDMA) technology where the concept of clusters is not used, and hence the same channels are reused in adjacent cells. That is, channels adjacent to a main channel are used for communication in the same cell. Distortion introduced in an adjacent channel is therefore more significant than with other systems.

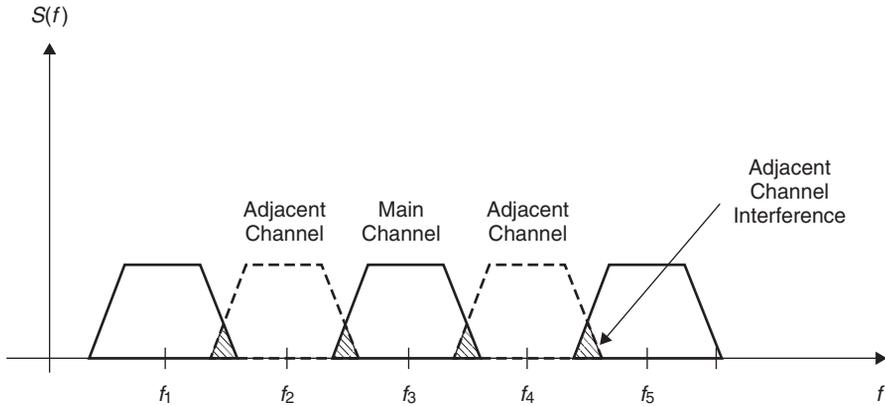


Figure 1.7 Adjacent channel interference in FDM systems (Rhyne and Steer, 1987).

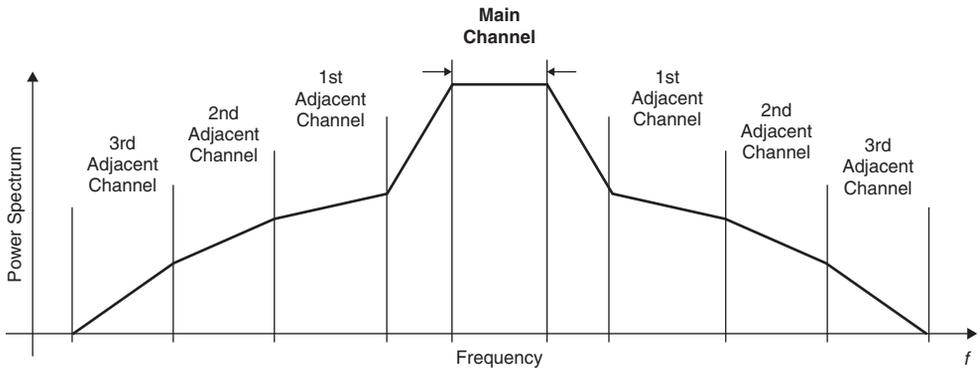


Figure 1.8 Definition of ACPR.

Therefore, it is necessary for the system design to impose a maximum level on the amount of distortion that can be introduced in adjacent cells. Since it is not possible to design for absolute distortion levels, relative levels of power in an adjacent channel to power radiated in the main channel is specified. The most commonly used measure of this distortion is called Adjacent Channel Power Ratio (ACPR) that refers to the ratio of the power in the main channel to the power in one of the adjacent channels (See Figure 1.8). The definition of ACPR is standard dependent and will be discussed in Chapter 7.

1.2.2 Modulation Quality and Degradation of System Performance

As discussed above, in-band distortion is responsible for the degradation of system performance that is manifested as the degradation of Signal-to-Noise Ratio (SNR) and ultimately system Bit Error Rate (BER). For linear modulation schemes, the nonlinear behavior

is manifested as compression and rotation of the signal constellation, which results in increasing system probability of error (bit error rate).

A basic approach to quantify nonlinear distortion in digital communication systems is to consider a memoryless nonlinearity which can be characterized by the AM–AM and AM–PM characteristics as in Equation (1.2). Therefore, for a digitally modulated signal of the form

$$x(t) = r(t) \cos(2\pi f_0 t + \psi(t)) \quad (1.5)$$

where $r(t)$ is a time-varying signal amplitude and represents the amplitude modulation of the signal; and $\psi(t)$ is the time-varying phase of the signal that represents phase modulation, the output of the nonlinearity can be expressed as

$$y(t) = F[r(t)] \cos(2\pi f_0 t + \psi(t) + \Phi(r(t))) \quad (1.6)$$

where $F[\cdot]$ is the amplitude distortion as a function of the input amplitude and represents the AM–AM characteristics of the nonlinear device and $\Phi[\cdot]$ is the phase distortion as a function of the input amplitude and represents the AM–PM characteristics. Therefore, with nonlinear amplification, the AM–AM distortion corrupts the envelope of the signal and the AM–PM distortion corrupts its phase. Figure. 1.9(a) shows the constellation diagram of a 16 QAM signal with Raised-Cosine (RC) pulse shaping and Figure. 1.9(b) shows the constellation diagram after amplification by a nonlinear amplifier. It is clear that the constellation diagram becomes distorted and this distortion appears as a compression and rotation of the constellation point. Consequently the received signal will be harder to detect and this results in degradation of the system BER (Briffa, 1996).

It is important to note that systems which use linear modulation schemes (QPSK, QAM, etc.), which have become common in communication systems and standards due to their higher spectral efficiency, are more susceptible to nonlinear amplification than systems which use modulation schemes with constant envelopes such as FSK (frequency modulation) or its variants. If the signal has time-varying amplitude then its instantaneous

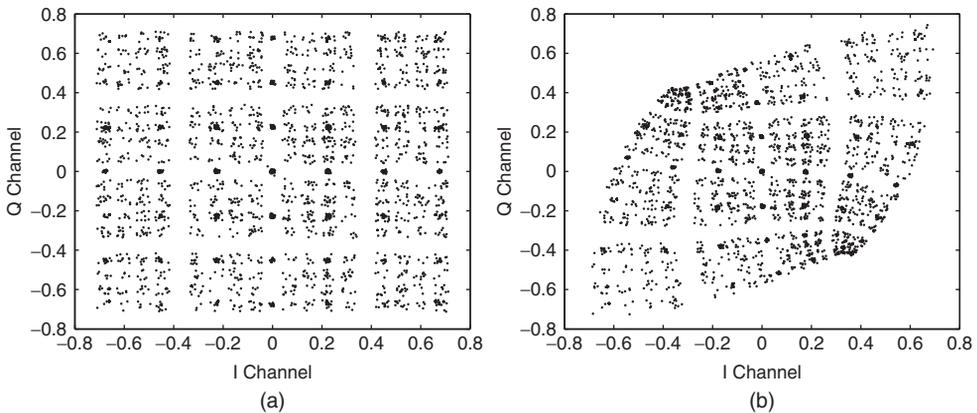


Figure 1.9 Signal constellation (a) before and (b) after nonlinear amplification.

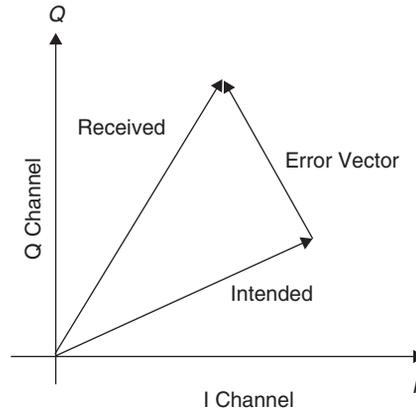


Figure 1.10 Definition of EVM.

input power changes continuously. This means that the signal at the PA output is distorted if the amplifier gain are not linear. If the signal has a constant envelope, then the PA linearity is not an important issue because the instantaneous input power stays constant and hence, there are no gain and phase variations for a specific operating point (Cylan, 2005).

Nonlinear distortion in wireless transmitters is usually quantified by the Error Vector Magnitude (EVM) which is a measure of the departure of signal constellation from its ideal reference because of nonlinearity. EVM is defined as the distance between the desired and actual signal vectors (error vector) as shown in Figure 1.10, normalized to a fraction of the signal amplitude. The actual value of the constellation point can deviate from the ideal value significantly, depending on PA nonlinearity. EVM quantifies in-band distortion which causes high bit error rates during reception of the transmitted data. Therefore, EVM specifications must be fulfilled in order to have a good quality of communication. Another measure of the fidelity of the signal under nonlinear amplification in some systems, like CDMA systems, is the waveform quality factor (ρ). The waveform quality factor is a measure of the correlation between a scaled version of the input and the output waveforms (Aparin, 2001).

1.2.3 Receiver Desensitization and Cross-Modulation

In a mobile receiver, the interaction of multiple signals by nonlinearity results in the problem of receiver desensitization. For example, one of the stringent requirements in second-generation CDMA receiver design (such as the IS-95 system) is the reception of a CDMA channel in the presence of a single-tone jammer (Aparin, 2003; Gharaibeh and Steer, 2005; IS95 Standard, 1993). The single tone jammer models, for example, a narrow-band Advanced Mobile Phone Service (AMPS) signal transmitted from a nearby AMPS base station or any other type of jamming or interference. In this scenario, desensitization of the single tone is a measure of the receiver's ability to receive a CDMA signal at its assigned frequency and in the presence of a single-tone jammer at a given frequency offset

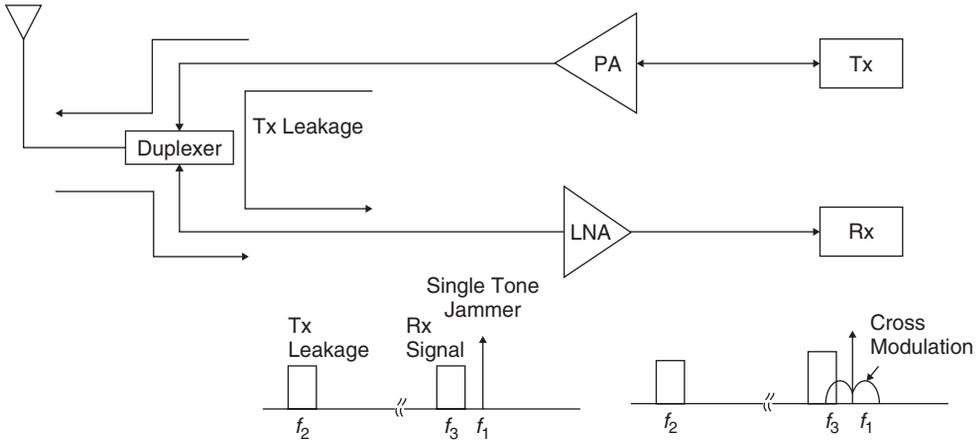


Figure 1.11 Receiver desensitization in reverse link CDMA system (Gharaibeh, 2004).

from the CDMA signal center frequency. This situation is shown in Figure 1.11 where the poor isolation between the transmitter (Tx) and receiver (Rx) ports of the duplexers causes a stronger Tx leakage to the Rx input. The interference introduced by the jammer results from cross-modulation of the jammer and transmitter leakage where Tx leakage modulation is transferred to the jammer by the Rx nonlinearities causing widening the jammer spectrum, as shown in Figure 1.11. This process results in a significant amount of interference in the Rx channels adjacent to the jammer reducing the Rx sensitivity (Aparin, 2003; Gharaibeh and Steer, 2005).

1.3 Modeling and Simulation of Nonlinear Systems

1.3.1 Modeling and Simulation in Engineering

System modeling and simulation are inevitable tools in modern engineering design. In addition to optimizing system design, the role of modeling and simulation is to allow concepts to be explored even before enabling technologies have been brought to maturity. In order to analyze the concept of system modeling and system simulations, it is important to define the concepts of a system, a model and simulation individually.

A system is defined as a collection of objects that interact together to achieve the objective of which the system is designed for. Systems produce outputs that depend on their inputs and on their contents, both of which may change with time and space. A model of a system is a mathematical representation that provides a simplification of the actual system in order to provide understanding and at the same time preserve the ability to predict the system output. The simplification of reality is usually done on two main aspects of the system. The first is the simplification of the details of the system, that is, the components of the system. The second is the simplification associated with the dynamics of the system that represent the cause–effect characteristics of the system that are separated in time and/or space (Jeruchem *et al.*, 2000).

Modeling of systems is usually faced with the question of how much the model represents the actual system given that it is a simplification of reality. The other issue that may arise at the same time is related to the amount of simplification that is allowed in order to develop a good model. The answers to these questions, in fact, depend on the level of understanding that we expect from the model. Too much simplification leads to a weak representation of reality and hence, the model does not provide enough understanding. On the other hand, less simplification means that more details are included in the model (in order to enhance the representation of reality) but this results in a complicated model that also does not promote understanding. Therefore, developing a model is a tradeoff between complexity and accuracy (IThink, 2009).

In general, there are two main types of modeling in engineering: physical modeling and behavioral modeling. A physical model is based on knowledge of the system components that together make up a real system and also, on knowledge of the rules that describe their interaction. Examples of such models are circuit-level models that are used in engineering design processes. On the other hand, behavioral (black box) models are based on the input/output measurements data and hence, their accuracy depends on the quality of measurements (Arabi, 2008; Jeruchem *et al.*, 2000). In general, physical models are more accurate than behavioral models since they involve more details on the system, while behavioral models are simpler and easier to simulate.

Simulation is defined as implementing a system model in a computer program that runs over time to achieve two main goals; the first is to study the interactions among system components; and secondly to study dynamics of the system. In other words, a simulation algorithm aims at studying how the system output changes when it is driven by different inputs and/or studying how the system behaves when the interactions among its components are defined differently. One of the benefits of simulation is that it provides time and space compression of the system where the behavior of the system over a long time period can be viewed with much shorter period by simulation, provided that the model is accurate (Frevert *et al.*, 2005; IThink, 2009; Jeruchem *et al.*, 2000).

Modeling and simulation are interrelated activities in the sense that when a model is used in simulation of a system, the simulator may produce results that indicate that the model is incorrect when compared to reality. This means that the model needs to be updated in order to achieve a closer representation of the actual system. This process continues until a good model and a good simulation algorithm are achieved that eventually lead to the desired representation of the real system. On the other hand, the level of details included in the model has a major influence on the number of computations needed to simulate the system. Therefore, the complexity of the simulation algorithm is always proportional to the complexity of the model since increasing the level of the model details or increasing the model dynamics means that more computations are needed to simulate the system.

Modeling and simulation are engineering disciplines that usually develop through practice rather than by studying their methodology. It is always the combination of the knowledge about the problem and the model developer practice that makes the process of modeling and simulation successful. In most engineering processes, the model developer simulates the model in order to get insight into the validity of the model. The developer then revises the model and simulates it again until an acceptable level of understanding about the system is obtained (Frevert *et al.*, 2005).

1.3.2 Modeling and Simulation for Communication System Design

As discussed in the previous section, modeling and simulation of systems are activities that accompany system design process. The design process is classified into several design levels that start at the “system level” and ends at the “gate level” in digital systems or the “circuit level” in analogue system; a process usually known as the top-down approach, see Figure 1.12 (Frevert *et al.*, 2005). System-level design of modern communication system usually follows the “specification approach” that is based on three main components; a set of specifications an algorithm that models the proposed system and a simulation tool. System-level design using this approach starts by setting up a list of specifications for the intended system and then developing an algorithm based on these specifications that achieves the desired performance criteria. A simulator is then used to evaluate the selected algorithm and to provide insight into which parameters of the algorithm need to be adapted in order to achieve the predefined performance criteria. On the other hand, a simulator provides insight on the mechanics of the proposed algorithm where its behavior is studied over the evolution of time. This process is repeated until an optimal algorithm that meets all the system specifications is reached.

In communication system design these algorithms mainly aim at transmitting data from the signal source at point A to a sink at point B. Algorithms that are specified at this level may be for example (Frevert *et al.*, 2005):

- data structure and protocol;
- Forward Error Correction techniques (FEC);
- modulation techniques (QPSK, QAM, GMSK, OFDM);
- channel equalization and synchronization.

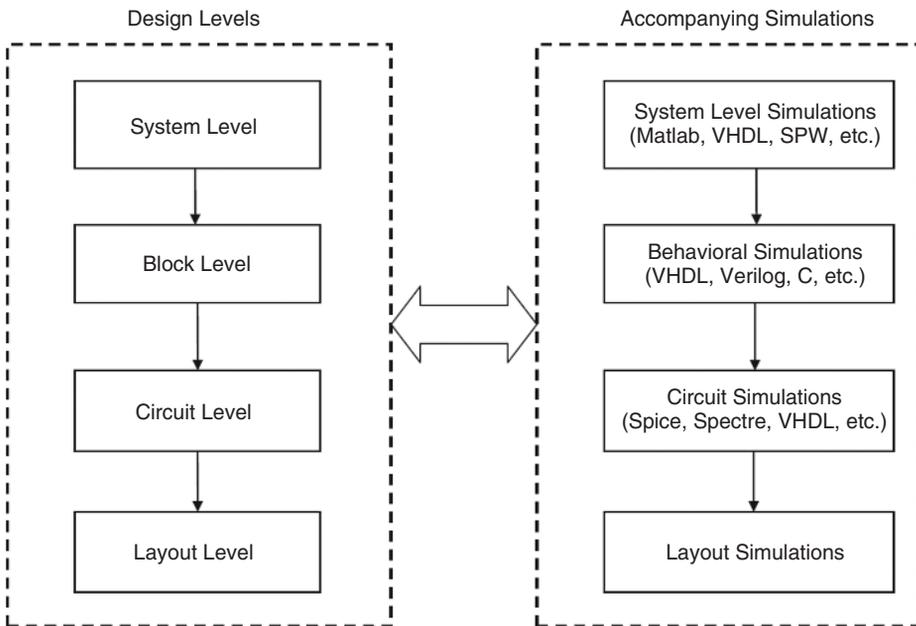


Figure 1.12 Design flow and the accompanying simulations (Frevert *et al.*, 2005).

Examples of system-level simulators are CoCentric System Studio[®], MATLAB[®], and SPW[®]. A commonly encountered example of system-level simulation of communication systems is the simulation of the degradation of the postulated system given a communication link budget. System BER is usually simulated versus SNR and distortion parameters that are set according to an initial link budget. If the result of the simulation says that the performance measure cannot be met, then the link budget or the distortion parameters must be modified. This process is repeated until the performance measure is met (Frevert *et al.*, 2005).

System-level design verified by simulation is then used as an input to the following design steps where the system is broken down into many components, each of which is designed individually by the aid of simulations. Therefore, system-level design and simulation enables the building blocks of the system to be identified. The next step towards complete system design is to design these building blocks. This design level is called “block-level” design and consists of detailed descriptions of each block which, altogether, lead to the functionalities identified by the system-level design. In communication systems, these blocks can be analogue or digital or mixed blocks and can be designed individually using block-level simulators that enable the verification of the functionality of each block. Examples of block level simulators are VHDL-AMS, Verilog-AMS and ISS (Frevert *et al.*, 2005).

The next level of system design is called the gate level in digital subsystems or the circuit level in analogue subsystems. In this level, gates or circuits are represented by netlists that contain gates or analogue elements. Simulation at this level is done using circuit level simulators such as VHDL and Verilog for digital circuits or SPICE and Spectre for analogue circuits. Circuit simulation enables the design at the block level to be evaluated and modified to meet the performance measures of each of the block (Frevert *et al.*, 2005).

1.3.3 Behavioral Modeling of Nonlinear Systems

Behavioral modeling of nonlinear systems refers to a class of modeling techniques where a nonlinear system is dealt with as a black-box. The black-box model is developed using a set of Input-Output (I/O) measurements and then by using system optimization theory, a function or a set of functions along with their parameters are extracted. Unlike circuit-level modeling, no information is required about the circuit topology, however, intuitive or a priori knowledge about the system helps in developing the model. For example, in modeling nonlinear power amplifiers, a priori knowledge is available since nonlinearity originates from active devices, which is well understood in circuit theory. Therefore, this nonlinearity takes input-output characteristics that are well known a priori and then the modeling problem reduces to estimating system parameters using classical parameter estimation theory such as least squares and its variants.

Different behavioral modeling approaches have been developed in the literature for modeling nonlinear systems. Perhaps the most common are those based on the Volterra and Wiener theory of nonlinear system identification (Bendat, 1990; Chen, 1989; Jeruchem *et al.*, 2000; Schetzen, 1981; Shi and Sun, 1990). A detailed discussion on different modeling approaches for nonlinear systems is presented in Chapter 3.

Table 1.1 Periodic steady-state methods for various nonlinear circuits (Mayaram, 2000)

Simulation Method	Nonlinear Device
Time Domain	Amplifiers, Mixers, Oscillators
Frequency Domain	Mildly Nonlinear Amplifiers, Mixers, Oscillators
Mixed Time–Freq	Mixers, Switched Capacitor
Envelope	Oscillators, PLLs
Linear Time-Varying	Mixers, Switched Capacitor

1.3.4 Simulation of Nonlinear Circuits

Simulation methods of nonlinear circuits are usually divided into main categories: time domain methods and frequency-domain methods. In time-domain methods, a circuit is simulated by solving the system differential equations numerically with a time discretization method (Kundert *et al.*, 1989). On the other hand, frequency-domain methods use the Fourier analysis to solve the circuit equations in the frequency domain. The two methods produce a periodic steady-state solution to the circuit equations. The main difference between these two methods is that time-domain methods can deal with strongly nonlinear circuits and discontinuities while frequency-domain methods deal with circuit components that can be characterized in the frequency domain.

Other approaches include mixed frequency–time approaches and circuit envelope methods which use a combination of both time and frequency methods to simulate a circuit and aim at reducing the drawbacks of each method in specific applications. In the following, a brief description of these methods is presented. Table 1.1 shows the applications of each of the presented simulation methods.

1.3.4.1 Time-Domain Methods

Time-domain methods are used to find the steady-state solution of the underlying differential equations of a circuit assuming that the solution is periodic. In these methods, a set of nodal voltages are determined such that $\mathbf{v}(T) = \mathbf{v}(0)$, where T is the period and $v(0)$ is the initial condition that forces the solution to be periodic.

The solution of the circuit differential equations is usually done using iterative approaches such as the Newton shooting method that is used by most commercial circuit simulators such as **Spice**[®] and **Spectre**[®]. Since this method is iterative, the time needed to reach a solution depends on a prescribed error tolerance.

There are a number of issues associated with time-domain transient simulations. First, time-domain methods have limitations on the size of the circuits to be simulated because large circuits tend to produce large matrices that need to be efficiently manipulated by computers. Therefore, time-domain methods are usually useful for circuits with less than 300 nodes (Mayaram, 2000). Secondly, time-domain methods usually require excessive computational times because they deal with the absolute bandwidth of the signal rather than the baseband bandwidth as in envelope simulations. Finally, time-domain methods

have limitations when used for simulation of distortion in nonlinear circuits where special care must be taken when choosing the error tolerance used in obtaining the iterative solution of the circuit differential equations (Mayaram, 2000).

1.3.4.2 Harmonic Balance (HB) Method

Harmonic Balance (HB) is a frequency-domain method for simulation of mildly nonlinear analogue circuits. HB is based on finding the frequency spectrum of the resulting currents in the circuit given a periodic voltage excitation when the circuit is in its periodic steady state. HB methods are useful in computing quantities that define nonlinear distortion such as third-order Intermodulation Distortion (IMD3), third order Input Intercept Point (IIP3) and Total Harmonic Distortion (THD). HB methods can also be used to perform nonlinear noise analysis of analogue circuits and also to compute harmonics, phase noise, and amplitude limits of oscillator (Agilent, 2006). An example of circuit simulators based on HB is Agilent ADS.

In harmonic balance, the system of nonlinear equations is formulated in both time and frequency domains where the linear contributions are calculated in the frequency domain and the nonlinear contributions in the time domain. Therefore, given an input voltage excitation to the circuit in the frequency domain $V(\omega)$, the resulting current $I(\omega) = F(V(\omega))$ in the linear part of the circuit is found by solving circuit equations using the Kirchhoff Current Law (KCL) at each node. KCL is applied for a number of independent frequencies (harmonic) and the current is calculated at those harmonics by solving a system of linear equations. (Gilmore, 1991; Rabaie *et al.*, 1988).

For the nonlinear part of the circuit, the current–voltage calculations are performed in the time domain where the time-domain representation of the input voltage excitation is found using the Fourier series as

$$v(t) = a_0 + \sum_i a_i \cos(\omega_i t) + \sum b_i \sin(\omega_i t) \quad (1.7)$$

where $\omega_i = i\omega_0$ and ω_0 is the fundamental frequency of the input signal. The time-domain waveform is then applied to the nonlinear part of the circuit and the current response $i(t) = f(v(t))$ is determined in the time domain. The time-domain current is then transformed to frequency domain using Fourier transform.

In a next step the current found in the nonlinear part of the circuit is converted to the frequency domain using Fourier transform and then the frequency spectrum of all the currents at a node is balanced at each frequency. This results in a set of linear simultaneous equations that can be solved in the frequency domain. The resulting currents are then converted to the time domain using the inverse Fourier transform.

Harmonic balance simulations represent an acceptable frequency-domain approach for simulating the response of amplifiers, mixers when stimulated by multisine signal sources. However, as the number of input tones increases, the simulation time increases. Hence, HB simulations are not usually practical when simulating non-periodic digitally modulated waveforms since these cannot be represented by discrete tones or need a large number of tones to approximate their continuous spectrum, which means very long simulation times (Yap, 1997).

1.3.4.3 Mixed Frequency–Time Methods

Combinations of both basic methods; the time-domain and frequency-domain methods, result in a new method called the mixed time–frequency method which result from a combination of the circuit envelope method and the quasi-periodic shooting method.

The basic idea of mixed frequency–time methods is that a quasi-periodic signal such as a modulated signal or a multitone signal can be recovered by simulating a finite number of periods of the carrier signal distributed evenly over the period of the modulation signal. For example consider, a two-tone signal of the form (Kundert, 2003)

$$v(t) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} V_{nm} e^{j2\pi(nf_1+mf_2)t} \quad (1.8)$$

where f_1 and f_2 are the fundamental frequencies. Now, if these frequencies are designated as a carrier (f_1) and an envelope of a modulated signal (f_2), then the signal $v(t)$ represents a quasi-periodic signal. The response of the circuit is also quasi-periodic and can then be represented by a Fourier series as (Kundert, 2003)

$$i(t) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} I_{nm} e^{j2\pi(nf_1+mf_2)t} \quad (1.9)$$

In mixed frequency–time simulation, the objective is to find the discrete response $i_n = i(nT_1)$ of the circuit instead of the continuous waveform $i(t)$ obtained from sampling $i(t)$ at the carrier frequency, where $T_1 = 1/f_1$. The key idea with mixed frequency–time simulations is that sampling the quasi-periodic two-tone signal at the carrier frequency (f_1) results in a sampled waveform that is periodic at the modulation frequency. This means that if K harmonics are needed to represent the modulation signal (envelope) then the quasi-periodic signal can be recovered by knowing only $(2K + 1)$ cycles of the carrier evenly distributed over the period of the modulation (envelope) signal. Hence, the simulation time is determined by the number of harmonics needed to represent the envelope waveform and not the carrier, which means less simulation time than time-domain methods (Kundert, 2003).

One application of mixed frequency–time methods is the distortion analysis of switched-capacitor circuits where the path of the signal is linear but the clock causes switching. This is because the period of the clock is usually orders of magnitude smaller than the time interval of interest and hence, classical circuit simulation algorithms become extraordinarily computationally expensive (Mayaram, 2000).

1.3.4.4 Envelope Methods

Time-domain methods are not suitable for simulation of high-frequency modulated signals because the simulation must follow the fast varying carrier signal. This results in prohibitively large number of time steps and hence, long simulation times and large memory requirements. Due to the fact that in modulated signals a high-frequency carrier is used to carry slowly varying information signals where the carrier does not contain any information, simulation of modulated signals in RF circuits can be done on the envelope level rather than simulating the RF modulated carrier (Berenji *et al.*, 2006).

Circuit envelope simulation methods applies time-domain analysis techniques on top of the harmonic balance simulations where the solution to the circuit equations is represented as the sum of harmonics with time-varying complex envelopes (Yap, 1997):

$$i(t) = \Re \left(\sum_{k=1}^N i_k(t) e^{j\omega_k t} \right) \quad (1.10)$$

where $i_k(t)$ represents the time-varying complex envelope of the current waveform that represents the modulation of the carrier. This complex envelope is simulated using time-domain techniques where the time step of the simulation is proportional to the bandwidth of the complex envelope and not to the carrier frequency. This means much lower time steps and hence much lower simulation time. On the other hand, the carrier is simulated either in the frequency domain using HB or in conventional time-domain transient simulators. In this way, the modulation complex envelope is solved for in the time domain instead of simulating multiple separate tones as in the HB where all frequency components must be solved for simultaneously. This means that the matrix sizes used in the HB solution remain reasonable for simulation. The result of circuit envelope simulations is a time-varying frequency spectrum where the time-varying envelope is not represented by additional tones but rather by the time-varying behavior of the spectrum (Berenji *et al.*, 2006; Yap, 1997).

For example, a two-tone signal can be interpreted as a periodically modulated periodic signal by rearranging equation (1.8) as (Kundert, 2003)

$$v(t) = \sum_{n=-\infty}^{\infty} V_n(t) e^{j2\pi n f_1 t} \quad (1.11)$$

where

$$V_n(t) = \sum_{m=-\infty}^{\infty} V_{nm} e^{j2\pi m f_2 t} \quad (1.12)$$

This form is equivalent to representing the two-tone signal by a conventional Fourier series with Fourier coefficients defined for integer multiples of f_1 , except that the Fourier coefficients themselves are time-varying. These time-varying coefficients are periodic with period $T_2 = 1/f_2$ and can themselves be represented by a Fourier series (Kundert, 2003).

1.4 Organization of the Book

The remainder of this book consists of 9 chapters. In Chapter 2, an introduction to wireless system standards and signal models is presented. Chapter 3 presents the different models of nonlinearity that will be used in other chapters for modeling and simulation of nonlinear distortion. Chapter 4 provides analysis of nonlinear transformation of deterministic signal including single and multi-tone analysis. Closed-form expressions that relate signal distortion to nonlinear system characteristics are presented for various nonlinear models such as power series, the Volterra model and block models. Chapter 5 discusses

the analysis of the response of nonlinear system to random input signals that represent real-world communication signals. This chapter provides the reader with the probabilistic view of nonlinear distortion and presents quantification of distortion using autocorrelation function analysis, nonlinear spectral analysis and the analysis of multichannel nonlinear distortion. Chapter 6 presents the identification of correlated and uncorrelated distortion using the concept of the orthogonalization of the behavioral model. This analysis is used to identify the distortion components responsible for the degradation of wireless system performance. Chapter 7 uses the concepts developed in Chapter 6 and introduces modeling of communication system figures of merit related to nonlinearity such as spectral mask parameters, Signal-to-Noise and Distortion Ratio (SNDR), Noise-to-Power Ratio (NPR), EVM, NF and BER. Chapter 8 provides an introduction to the simulation of communication systems in MATLAB[®] where the basics of simulation of modern wireless communication systems in Simulink[®] are presented. Chapter 9 explains how to use MATLAB[®] to simulate various types of nonlinearity and how to analyze, predict and evaluate the performance of data communication systems under nonlinearity. Chapter 10 explains how to use Simulink[®] to analyze, predict and evaluate the performance of data communication systems under nonlinearity and provides a comprehensive reference for models of nonlinearity in Simulink[®].

1.5 Summary

In this chapter the basics of nonlinearity and nonlinear distortion in wireless communication systems have been presented. The main sources of nonlinearity in wireless systems and their behavior that causes performance degradation have been discussed. It has been shown that nonlinear distortion is one of the most important considerations in the design of wireless systems. It has also been shown that modeling and simulation of nonlinear distortion are important as accompanying activities during the design cycle of any wireless system. This chapter serves as an introduction to the topics that will follow in the subsequent chapters in the sense that all modeling and simulation issues that will be discussed next are justified by the importance of the problem of nonlinear distortion in wireless systems.