Effect of Power Amplifier Nonlinearity on System Performance Metric, Bit-Error-Rate (BER)

by

Farinaz Edalat

Bachelor of Science in Computer Engineering
University of Illinois at Urbana-Champaign, 2001

Submitted to the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2003

Signiture of Author: .....
Department of Electrical Engineering and Computer Science
August 29, 2003

Certified by: ...
.....
Charles G. Sodini
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by: .....
.....
Arthur C. Smith
Professor of Electrical Engineering and Computer Science
Chairman, Committee for Graduate Students

Effect of Power Amplifier Nonlinearity on System Performance Metric, Bit-Error-Rate (BER)

by
Farinaz Edalat

Submitted to the Department of Electrical Engineering and Computer Science
on August 29, 2003 in Partial Fulfillment of the
Requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

In today's high-speed wireless communication applications, the simultaneous demand for bandwidth-efficient modulation schemes such as M-QAM and high power-efficiency of electronic components in the system imposes a contradicting tradeoff. On one hand, with the RF power amplifier (PA) at the transmitter antenna being the major power "hungry" block in many such systems, high power-efficiency for the system is obtained when the PA consumes most of the little supplied dc power (e.g. from a battery) as RF power, which is the power delivered to the load. On the other hand, at high powers, the PA becomes nonlinear, which becomes problematic when the input to the PA has a time-varying envelope. More precisely, when an M-QAM modulated signal (which has varying envelope and phase) is transmitted, the PA nonlinearity characteristics adversely affect the system performance, such as degradation of the Bit Error Rate (BER).

Although linearization and compensation techniques can be used to minimize the above mentioned degradation in system performance, such techniques are mostly not suitable for low-cost and small-size integrated-circuit implementations. As a result, to obtain an optimal solution to this linearity/efficiency tradeoff, the objective of this thesis is to qualitatively and quantitatively characterize the effect of the PA nonlinearity on the system performance metric, BER. To gain an intuitive understanding of the effect of the PA nonlinearity, a system simulation is developed. In addition, to achieve a simple relationship between this circuit nonlinearity and system performance, an appropriate model (AM-AM/AM-PM) to characterize the PA nonlinearity is found. In fact, the results of this thesis in characterizing the effect of the PA nonlinearity on system performance can be useful to PA designers to design cost-effective and efficient PAs by finding a PA operating point that lies in the nonlinear region as much as possible while still meeting the system performance requirements.

Thesis Supervisor: Charles G. Sodini
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

There are many people that I would like to thank without whom the completion of this thesis would not have been possible.

First and foremost, I would like to express my whole-hearted gratitude to my advisor, Professor Charles Sodini, for his invaluable guidance, insight, and encouragement throughout this thesis work. The experience of mixing the system and circuits spices was truly thrilling!

My special appreciation goes to Professor Michael Perrott whose prompt assistance with using the CppSim simulation tool speeded up my work to a great extend and made experience of working with this tool an enjoyable one. I would also like to thank Professor Muriel Medard for her valuable suggestions on some system issues of this project.

My tremendous thanks go to my officemates, Andy Wang and Anh Pham for their technical support and friendship. Andy was a great teacher in sharing his rich and vast theoretical and practical knowledge in communication systems and circuits as well as simulation tools. His ingenious questions and ideas often led me in the right path by making me look at things from a different angle. Anh is who I call the "power amp wiz" to whom I am obliged for much of my understanding of power amplifiers. His countless hours of answering my confusions about circuit theories were precious. I would also like to thank my friend and colleague, Ali Motamedi, for sharing his valuable insights and giving consultations on some power amplifier issues.

My gratefulness is extended to all my other colleagues in the office. I thank Todd Sepke for his assistance with all computing problems. The hot talks on ski adventures and making ski trip plans with John Fiorenza, a truly expert downhill skier, were always inspiring. My thanks go to the rest of the group who inspired me to join the Mallard MTL Hockey team and made the office a fun and comfortable environment to work: Andrew Chen, Albert Chow, Matt Guyton, Lunal Khuon, Jacky Liang, Pablo Acosta-Serafini, and Mark Spaeth.

I would like to thank all my other friends and colleagues for their warmth and laughters that have made MIT a fun place to be. I am especially grateful for the love and encouragement of my closest friends, Bahareh Banijamali and Miriam Bruhn (my best roommate as well). In addition, much appreciation goes to MIT's athletics facilities (Dupont gym and the Z-center): those every evening workout helped me to be more effective and energetic for the rest of the day.

I am very grateful to Marilyn Pierce for making the administrative issues and paper works the most time-efficient.

Finally, I would like to thank my family for always being there for me. I thank my fantastic sister and brother for their love and support. Thank you mom for your unconditional love and wisdom have helped me survive the difficult times. Dad, your brilliant advice and constant supervision have provided me with confidence and strength to always take one step further and challenge myself. Mom and dad this thesis is dedicated to you.

This work is sponsored, in part, by the National Defense Science and Engineering Graduate Fellowship (NDSEG).

Contents

1 Introduction **13**
- 1.1 Thesis Objective . . . 14
- 1.2 Thesis Outline . . . 14

2 Wireless Communications System - Background **17**
- 2.1 Performance Measurement Metrics . . . 18
  - 2.1.1 Bit-Error-Rate (BER) . . . 18
  - 2.1.2 Bandwidth . . . 18
  - 2.1.3 Power efficiency . . . 19
- 2.2 Modulator . . . 19
  - 2.2.1 Quadrature Amplitude Modulation (QAM) . . . 20
  - 2.2.2 Pulse-Shaping (Baseband) Filter . . . 23
- 2.3 Channel . . . 25
- 2.4 Optimum Receiver . . . 26
  - 2.4.1 Demodulator . . . 26
  - 2.4.2 Detector (or Decoder) . . . 29

3 Power Amplifier Characteristics and Nonlinearity Effects - Background **35**
- 3.1 Input-Output Characteristics of PA . . . 35
  - 3.1.1 Input-Output Characteristics - Linear PA . . . 36
  - 3.1.2 Input-Output Characteristics - Nonlinear PA . . . 36
- 3.2 Nonlinearity Effects of PA . . . 41
  - 3.2.1 Single-Tone Test . . . 41
  - 3.2.2 Two-Tone Test . . . 43

4 Simulation **47**
- 4.1 Simulation Tool . . . 47
- 4.2 System Simulation Component Blocks . . . 48
  - 4.2.1 Modulator . . . 49
  - 4.2.2 PA . . . 53
  - 4.2.3 Channel . . . 53
  - 4.2.4 Demodulator . . . 54
  - 4.2.5 Detector . . . 55
  - 4.2.6 System Performance Measurement - BER . . . 55
- 4.3 Generality of the System . . . 57
- 4.4 Issues and Challenges in the Simulation . . . 57
  - 4.4.1 Real or Complex Signals . . . 57

4.4.2 Performance of an Ideal System 58
4.4.3 PA Operating Point 58
4.4.4 SNR 59

5 PA Nonlinearity Model **61**
5.1 PA Nonlinearity Model: AM-AM/AM-PM 61
5.2 Specifications of Experimented PA 63
5.3 Nonlinearity Model of Simulated PA 63

6 Effect of PA Nonlinearity on System Performance Metric - BER **67**
6.1 Literature Background 67
6.2 Visualization of the Effect of PA Nonlinearity on BER 68
6.2.1 Effect of PA Nonlinearity Alone 68
6.2.2 Effect of PA Nonlinearity in a Communications System without AWGN Channel 70
6.2.3 Effect of PA Nonlinearity in a Communications System with AWGN Channel 70
6.3 Quantification of the Effect of PA Nonlinearity on BER 74
6.3.1 Effect of PA Nonlinearity on BER - 16-QAM 74
6.3.2 Effect of PA Nonlinearity on BER: 16-QAM versus 64-QAM 77
6.4 Application: Finding the Optimal PA Operating Point 82

7 Conclusions **85**
7.1 Summary 85
7.2 Future Work 87

List of Figures

2-1 Basic building blocks of a wireless communications system. 17
2-2 A complex implementation of the QAM modulation followed by conversion to a real bandpass signal, $x_{tx}(t)$, to be transmitted over the channel. 21
2-3 An alternative realization of QAM modulation in which the real symbol sequences $\{I_{mr}\}$ and $\{I_{mi}\}$ are acted upon individually. The final real bandpass signal, $x_{tx}(t)$, is obtained by adding the two real bandpass signals on the quadrature carriers. . . 22
2-4 Gray encoding of a 16-QAM signal constellation. 22
2-5 Plot of $B(f)$ when $W < \frac{1}{2T}$. 24
2-6 Plot of $B(f)$ when $W = \frac{1}{2T}$. 24
2-7 Plot of $B(f)$ when $W > \frac{1}{2T}$. 25
2-8 Effect of the roll-factor on the time response and frequency spectrum of the raised-cosine filter. The time and frequency-domain characteristics are depicted for the roll-off factors $\beta = 0.001,\ 0.5,\ 1$. 25
2-9 A model of a band-limited channel: a linear filter with an additive noise. ($\otimes$ represents the convolution operation.) 26
2-10 A plot showing the error probability in detecting the corrupted symbol y when a_1 was sent under the MD rule. Since noise $n > d(a_1, a_2)/2$, it moves the transmitted symbol a_1 closer to the other symbol a_2, which is thus mistakenly detected as the transmitted symbol. 30

3-1 Power Amplifier two-port model. 35
3-2 Input-output voltage characteristics (Source: Pothecary [5]). 37
3-3 Definition of the 1-dB compression point (Source: Razavi [2]). 37
3-4 Power amplifier AM-AM characteristic. 39
3-5 Measured AM-PM for a 1-Watt, 1.9 GHz PA. Circles indicate the 1-dB compression points (Source: Cripps [7]). 40
3-6 Input and output signals in Saleh's PA Model. 40
3-7 Effect of the PA nonlinearity: harmonics and intermodulation distortions. 43
3-8 Effect of the PA nonlinearity: corruption of a signal due to the intermodulation between two interferers (Source: Razavi [2]). 44
3-9 Definition of the third-order intercept point IP_3 based on the growth of output components in a two-tone test (Source: Razavi [2]). 44

4-1 High level schematic description of the simulated system. 48
4-2 Plot showing the response of a square-root raised cosine filter (at the transmitter) to an impulse train (generated by an M-QAM source), followed by the response of a second square-root raised cosine filter (at the receiver) to the output of the first filter. 51

4-3 The effect of roll-off factor on the response of filter to a sequence of randomly generated data. 52
4-4 The modulator consists of an M-QAM source and a square-root raised cosine filter. The M-QAM modulation is a 16-QAM or 64-QAM with Gray encoding mapping and a rectangular constellation. The QAM source generates symbols at a rate of $\frac{1}{T_s} = 1e^6$ $(symbols/sec.)$. The baseband filter is a sqrt raised cosine filter with a roll-off factor $\beta = 1$ and bandwidth $W = 1\ MHz$. Both signal and channel are baseband and band-limited to $W = 1\ MHz$. 53
4-5 Block components of the simulation system. 56

5-1 The AM-AM and AM-PM functions characterizing the PA output amplitude and phase-shift (as a function of the input amplitude) respectively. 62
5-2 Plot of the input-output power characteristics of the experimented PA. 65
5-3 Plot of efficiency versus input voltage amplitude of the experimented PA. 65
5-4 The AM-AM and AM-PM nonlinearity models of the simulation PA block. 66

6-1 A rectangular 16-QAM signal constellation shows the three possible amplitudes: the inner 4 rectangles have the min. amplitude, the outer 4 circles have the max amplitude, and the 8 crosses have the in-between amplitude. 69
6-2 Effect of the PA nonlinearity alone on a 16-QAM constellation. 71
6-3 Effect of the PA nonlinearity in a communications system without channel noise on a 16-QAM constellation. The 16-QAM signal's maximum amplitude is chosen at the saturation point of the PA. 72
6-4 Effect of the PA nonlinearity in a communications system with an AWGN channel on a 16-QAM constellation. 73
6-5 Plot of symbol-error-probability (SER) versus normalized SNR (SNR_{norm}) for a 16-QAM signal going through linear and nonlinear PAs. On the right, plots of the input-power/output-power characteristics and the AM-AM/AM-PM functions of the linear and nonlinear PAs are shown with three operating points marked: 1-dB compression point (circle), a point in the linear region (cross), and a point in the nonlinear region below the 1-dB compression point (cross). The corresponding SNRs of the latter two points are indicated with crosses on the error probability plot. 76
6-6 Comparison of error probabilities of 16-QAM and 64-QAM using performance metrics, Symbol-Error-Probability or BER. 79
6-7 Comparison of 16-QAM and 64-QAM constellations for same minimum-distance and same maximum-power. 80
6-8 Plot of symbol-error-probability versus normalized SNR of 16-qam and 64-qam that have gone through linear and nonlinear PAs. 81
6-9 Finding the optimal PA operating point meeting the system requirement: 16-QAM modulation, maximum $SNR = 8dB$, and $BER \leq 10^{-3}$. 83

List of Tables

5.1 Specifications of the experimented power amplifier used to model the simulation *PA* block (Source: Pham [17]). 63

7.1 Specifications of the system simulation blocks. 86

Chapter 1

Introduction

The existence of wireless technology dates back to 1901, when Guglielmo Marconi successfully transmitted radio signals across the Atlantic Ocean. Since then, there has been a continuous surge in wireless electronics, motivating the wireless communications market to grow accordingly. In addition to the everyday wireless products such as cellular phones and Personal Digital Assistances (PDAs), recently, more luxurious wireless applications such as Wireless Local Area Networks (WLANs), Global Positioning Systems (GPS), and Personal Communications Services (PCS) have been flourishing.

In fact, with the ultimate goal of the wireless industry being elimination of wires for the transmission of information, new design challenges have been emerging in the radio frequency (RF) technology to handle the wireless transmission of voice, data, and video at increasingly higher rates. In particular, to implement fast, efficient and reliable wireless systems, one needs to understand the imposing limitations and find solutions to combat them. The availability of a limited RF spectrum, the finite battery life of portable devices, and the nonlinear nature of some of the circuit elements in the wireless systems are some of the major such limitations.

Today, there is an explosive demand for high-speed wireless electronics. For instance, the development of a WLAN supporting data rates on the order of Giga bits per second that continues transmission of high-speed data from internet to end-use devices, such as cameras, printers, and multimedia video is quite attractive. However, transmission at high data rates occupies a lot of bandwidth. As a result, as the demand for the RF spectrum increases, high-speed data transmission would benefit from bandwidth-efficient modulation schemes such as the Multi-level Quadrature Amplitude Modulation (M-QAM). In addition, in wireless applications, portability demands a highly power-efficient system to increase the battery lifetime, or the available talk-time in cellular systems. However, using an M-QAM modulation in conjunction with driving devices at high power-efficient levels presents a contradicting tradeoff. Such a tradeoff is due to the nonlinear characteristics of the RF power amplifiers (PAs), which are one of the major components in every wireless system. In many applications, PAs are the most power "hungry" blocks, consuming many times the combined power of the rest of the transceiver circuitry. Hence, not only can they severely reduce the battery lifetime, but also degrade the performance of the adjacent circuits by releasing tremendous amounts of heat. Consequently, power-efficiencies of the wireless systems are largely determined by those of their PAs. To obtain high power-efficiency, meaning that a large percentage of the dc power (e.g. battery) be delivered to the load, the PAs must operate at high RF power levels, where they exhibit a nonlinear behavior. On the other

hand, the time-varying envelope of a QAM modulated signal requires a high level of linearity of the transceiver components to achieve an acceptable performance, hence, requiring the operation of the PAs at lower RF power levels in the linear region. Thus, we are faced with a challenging tradeoff between the linearity and power-efficiency of the PAs. Our objective is to optimize this tradeoff.

1.1 Thesis Objective

Traditionally, PA designers solve the linearity/power-efficiency tradeoff problem by over-specifying the PA - i.e. operating it in the linear region to ensure an acceptable performance. However, this leads to low power-efficiency and thus high power-consumption, which cannot be tolerated in portable wireless systems. A higher power-efficiency of the PA can be obtained at the expense of a more nonlinear distortion. In other words, as the input power level enters the nonlinear region of the PA, the power-efficiency increases while the system performance, measured by the Bit-Error-Rate (BER), degrades. As a consequence, for a given PA nonlinearity, we desire to find the maximum input power - in the nonlinear region of the PA - that achieves a BER below the maximum allowable BER. The resulting operating point of the PA achieves the optimum tradeoff between the linearity and power-efficiency. This information is extremely valuable to the system and circuit designers, facilitating the collaboration between the two, and hence, leading to less expensive and more efficient practical design solutions in a shorter amount of time. However, obtaining this information requires knowledge of how the PA nonlinearity parameters are related to the system performance metric, BER, which constitutes the main objective of this thesis.

To obtain an intuitive understanding of how the PA nonlinearity affects the system performance metric, BER, a typical communications system is simulated. In this simulation, we must define a behavioral model for the PA block that captures its nonlinearity characteristics in the most appropriate way to relate to the system performance. After gaining an intuitive understanding of how the PA nonlinearity affects the system performance through the simulation results, we measure the resulting degradation in the system performance using the metric, BER. Finally, from the results, we draw conclusions, characterizing the effects of the PA nonlinearity on the performance both quantitatively and qualitatively.

1.2 Thesis Outline

First and foremost, the objective of this thesis is to present the details of this work, observations and results. Meanwhile, to further understand the value of this thesis work, the relations of the thesis findings to those of the predecessors are discussed in the chapters of the corresponding subjects. In addition, a thorough understanding of this thesis work requires knowledge of some system and circuit theories. Hence, two chapters are devoted to introducing the required system and circuit background.

The above objectives are achieved through seven chapters in the following order. To understand the role of the communications system blocks used in the simulation, we start with providing the necessary background on the system theory in Chapter 2. Next, since the focus of this thesis from the circuits point of view is on the power amplifiers, Chapter 3 is devoted to power amplifiers' characteristics. Chapter 4 introduces the simulated communications system. Because of the direct impact of the method of implementation of the simulation's PA block on the results, Chapter 5 explains how the PA input-output behavior

is modeled. Next, the results of this thesis work on characterization of the effect of the PA nonlinearity on the system performance metric, BER, are presented in Chapter 6. Finally, Chapter 7 concludes with a summary of results and suggestions for future work.

Chapter 2

Wireless Communications System - Background

In general, a digital communications system consists of a transmitter, a channel, and a receiver, with their detailed building blocks as depicted in Fig. 2-1 [1].

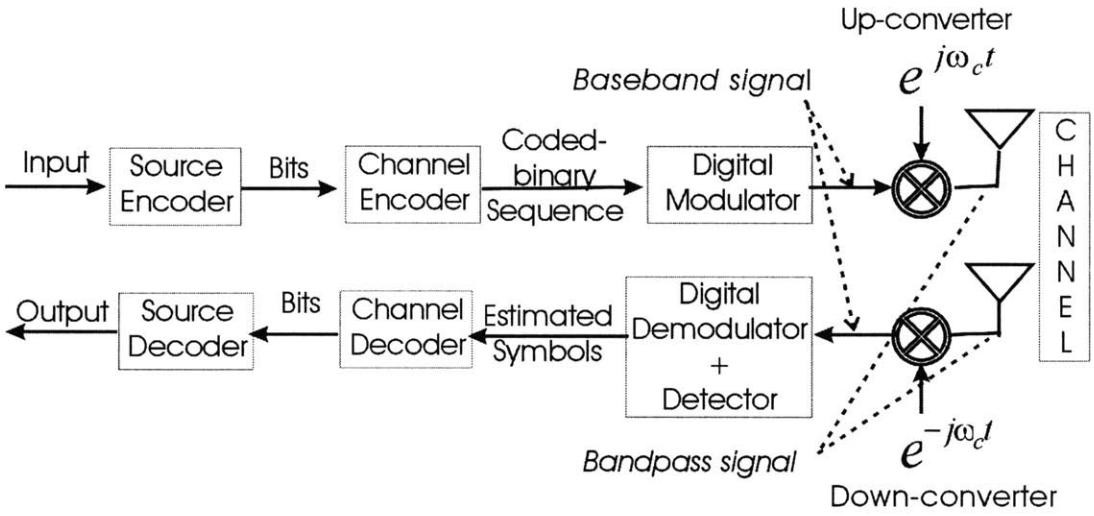

Figure 2-1: Basic building blocks of a wireless communications system.

Initially, the *source encoder* converts the information source - e.g. in the form of voice, video, or digital data - to a stream of bits in an efficient manner by using as few bits as possible. The resulting sequence of binary digits are then coded by the *channel encoder* which introduces redundancy in the binary sequence that can be used at the receiver to combat the distortions due to the circuit nonlinearities, and the noise and interference introduced by the channel. Next, the *digital modulator*, transforms the binary code sequence into signal waveforms. In wireless systems, depending on the application, the signals must be transmitted over a particular frequency band, where the channel resides. Therefore, the modulated signal waveforms are *up-converted* by a carrier to the desired frequency before being sent over the antenna. The *communication channel* is the physical medium through which the signal passes from the transmitter to the receiver. The channel distorts the transmitted signals by various means, the most familiar of which is the additive thermal

noise from the channel and electronics in the system. At the receiver, the corrupted signals received by the antenna pass through a set of blocks that undo the processing performed on the signals at the transmitter to recover the original source information. Hence, the received signal waveforms are first *down-converted* back to the baseband (i.e. frequency band around zero), and subsequently pass through the *demodulator* and *detector* that convert the analog waveforms into a sequence of estimated symbols chosen from the signal set used by the modulator. The *channel decoder*, then, attempts to reconstruct the source message as faithfully as possible.

This chapter provides the communications system background sufficient to understand the particular choice of the building blocks in our simulation system as introduced in Chapter 4. The chapter starts with introducing three major metrics that quantify the performance of wireless systems. Next, a section is dedicated to each fundamental component in a typical communications system: a modulator (consisting of an M-QAM modulation and a pulse-shaping filter), a channel, and an optimum receiver (consisting of a demodulator and a detector).

2.1 Performance Measurement Metrics

In the simplest form, a communications system consists of a modulator (at the transmitter), a channel, and a demodulator (at the receiver). Hence, the overall system performance depends on the design of both the modulator and the demodulator (and detector). In the RF design, there are three parameters that are useful in assessing the performance of a digital modulation/demodulation (modem) technique: the Bit-Error-Rate (BER), bandwidth efficiency, and power efficiency [2]. Subsequently, performances of different types of digital modulations are compared in terms of these parameters.

2.1.1 Bit-Error-Rate (BER)

In analyzing the performance of a system, the quality of the output of the detector in the presence of noise in the channel and transceiver components is measured via a parameter known as the signal-to-noise ratio (SNR). Moreover, for a given transmitted power and noise in the system, such a quality depends on the type of the modem used. In other words, a modem with a higher immunity to noise allows more noise to be tolerated or less power to be transmitted and still achieve a desirable performance. In addition, a measure of how well the corrupted signal is detected is the frequency of occurrence of errors in the decoded sequence - which is the average probability of bit-error. This is measured by the system performance metric, BER defined as the average number of erroneous bits observed at the output of the detector divided by the total number of bits received in a unit of time [2]. In digital modulation, BER is expressed in terms of the SNR.

2.1.2 Bandwidth

Another important characteristic of a modulation scheme is the bandwidth occupied by the modulated signal for a certain bandwidth of the baseband signal. The parameter that quantifies such a characterization is bandwidth efficiency, defined as the data rate (bits per second) over the bandwidth of the modulated signal (Hertz). In band-limited channel applications, such as wireless networks, bandwidth efficiency plays a key criterion in choosing a modem.

2.1.3 Power efficiency

As it will be shown later in this chapter (2.42), the error probability of a modem is a function of the Q-function whose argument is $d_{min}/2\sigma$, where $\sigma^2 = N_0/2$ is the noise variance. More specifically, the higher the minimum-distance, d_{min}, between the symbols in the constellation is, the better the performance is. However, higher d_{min} is achieved at the cost of requiring more signal energy, which is a limited source and is desired to be as low as possible. This presents the conflicting tradeoff between d_{min} and signal energy, E_s. The *asymptotic power efficiency* γ is defined to express "how efficiently a modulation scheme makes use of the available signal energy to generate a given minimum distance" [3] by

$$\sqrt{\gamma \frac{E_s}{N_0/2}} = \frac{d_{min}}{2\sqrt{N_0/2}}$$

Hence,

$$\gamma = \frac{d_{min}^2}{4E_s} \tag{2.1}$$

In other words, a modulation scheme with a higher power efficiency results in a lower error probability with the optimum tradeoff between d_{min} and E_s.

2.2 Modulator

In the transmission of digital information over a communication channel, the modulator is the interface device that maps the digital information into analog waveforms that match the channel characteristics [1]. Such an assignment is generally performed by grouping the information sequence I_m into b-bit blocks and mapping each block into one of the $M = 2^b$ deterministic finite energy waveforms $\{x_m(t), m = 1, 2, \ldots, M\}$ in the signal set for transmission over the channel. Depending on whether the transmitted waveform at any time interval depends on the previously transmitted waveforms or not, the modulation method is said to have memory or to be memoryless respectively. In a modulation with memory, dependence between the transmitted waveforms is introduced to match the spectrum shape of the transmitted signal to the channel spectral characteristics. Such spectrum shaping is typically employed in magnetic recording, optical communications, or digital communications over cable systems. In addition, a modulator is classified as either linear or nonlinear depending on whether the principle of superposition applies in the mapping or not. In RF applications, linear memoryless modulation methods are most commonly used and hence, we focus on this type of modulation method.

In wireless systems, digital modulation, in which the carrier is modulated by a digital baseband signal, is more widely used than analog modulation. Similar to analog modulations AM, PM and FM, digital modulations are classified based on whether the mapped analog waveforms differ in amplitude, phase, or frequency - resulting in amplitude-shift-keying (ASK) (more commonly known as pulse-amplitude modulation PAM), phase-shift-keying (PSK), frequency-shift-keying (FSK), or some combination of these signal parameters such as M-level quadrature-amplitude modulation (M-QAM). Comparing PSK and FSK modulations, PSK has a higher minimum distance between signal points in its constellation, and hence has a lower error probability for the same noise and energy per bit. In fact, FSK is widely used in low data rate applications, and therefore for high data rate RF applications, PSK is preferred over FSK. In PSK modulation, the information signal is carried by the

phase of the modulated signal, and thus the transmitted signal waveform has a constant envelope.

On the other hand, in PAM or QAM modulations, the information signal is carried by the amplitude of the modulated signal, resulting in a time-varying envelope. As a result, amplitude-modulated signals are vulnerable to power amplifier nonlinearities, while constant-envelope modulated signals are immune to such nonlinearities. However, despite this disadvantageous property of M-QAM signal sets, their higher bandwidth efficiency makes these signals highly desirable, especially for transmission over band-limited channels. The reason for the higher bandwidth efficiency of M-QAM modulation is explained in the next section.

2.2.1 Quadrature Amplitude Modulation (QAM)

As illustrated in Fig. 2-2, an M-QAM modulation can be realized follows. Initially, the input data bits, arriving at a rate of R bits/second, are mapped, b bits at a time, into a sequence of complex symbols $\{I_m\}$ chosen from an M-QAM symbol set (constellation) of size $M = 2^b$. Hence, the resulting symbol sequence has a *symbol rate* of $R_s = R/b$ symbols/second, and a *symbol interval* of $T_s = 1/R_s = b/R$ seconds [4]. The discrete-time complex sequence $\{I_m\}$, subsequently, passes through a baseband filter with real or complex impulse response $g(t)$, creating a complex baseband signal $x(t)$ as below,

$$x(t) = \sum_{m \in Z} I_m g(t - mT_s) \tag{2.2}$$

In other words, the complex sequence $\{I_m\}$ modulates the amplitudes of a sequence of time shifts $\{g(t - mT_s)\}$ of the baseband pulse $g(t)$. To transmit over a bandpass (real) channel, the baseband signal $x(t)$ is first translated up in frequency by a carrier, $e^{j\omega_c t}$, and then an equivalent spectrum is added in the negative frequencies to make the signal real. These steps are summarized as below,

1. $x(t) \rightarrow x(t)e^{j\omega_c t}$
2. $x(t)e^{j\omega_c t} \rightarrow [x(t)e^{j\omega_c t}] + [x^*(t)e^{-j\omega_c t}] = 2 \cdot \Re\{x(t)e^{j\omega_c t}\} = x_{tx}(t)$

where $\Re$ denotes the real part of the complex-valued quantity in the brackets.

Assuming that $g(t)$ is real, using (2.2), $x_{tx}(t)$ can be expanded into its real and imaginary counterparts, that is

$$x_{tx}(t) = 2 \cdot \Re\{[\sum_m (I_{mr} + jI_{mi})g(t - mT_s)]e^{j\omega_c t}\} \tag{2.3}$$

$$x_{tx}(t) = 2[\sum_m I_{mr} g(t - mT_s)] \cos(\omega_c t) - 2[\sum_m I_{mi} g(t - mT_s)] \sin(\omega_c t) \tag{2.4}$$

The above representation of the transmitted signal can be interpreted as follows. The complex transmitted symbols $\{I_m\}$ can alternatively be realized as real 2-tuples $\{(I_{mr}, I_{mi})\}$, representing their real and imaginary components respectively. The resulting real symbol sequences $\{I_{mr}\}$ and $\{I_{mi}\}$ are then mapped into the real baseband waveforms $\{\sum_m I_{mr} g(t - mT_s)\}$ and $\{\sum_m I_{mi} g(t - mT_s)\}$ respectively (as shown in Fig. 2-3). The resulting real waveforms subsequently modulate the quadrature carriers $\cos \omega_c t$ and $\sin \omega_c t$. The sum of the resulting real bandpass signals gives the final real bandpass signal, $x_{tx}(t)$, to be transmitted over the channel.

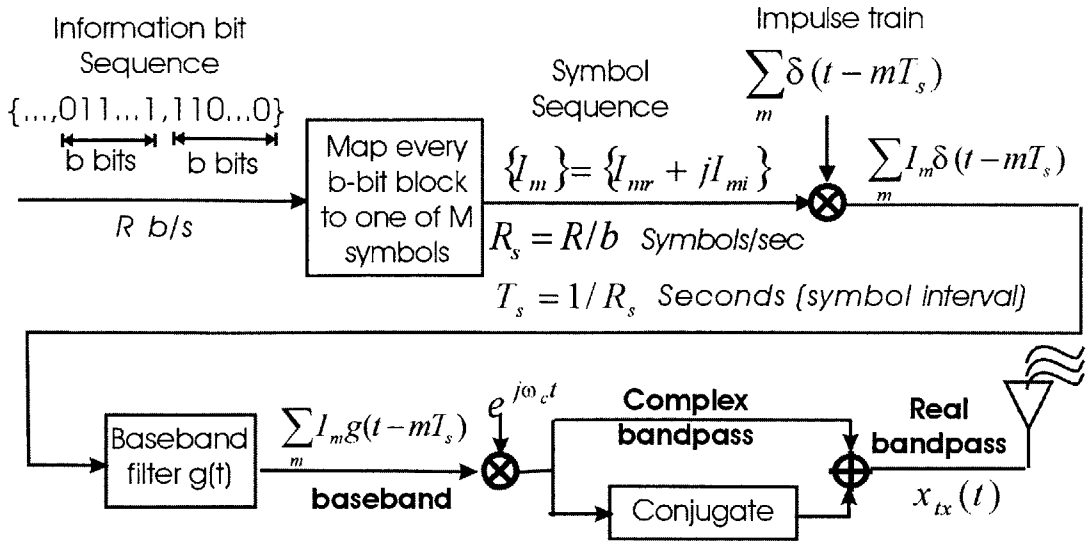

Figure 2-2: A complex implementation of the QAM modulation followed by conversion to a real bandpass signal, $x_{tx}(t)$, to be transmitted over the channel.

We are now ready to assert the reason for the higher bandwidth efficiency that an M-QAM signal set attains compared to that of a PAM signal set. As discussed in the previous paragraph, a complex M-QAM symbol can be generated by simultaneous transmission of two real PAM symbols on quadrature carriers. When $M = 2^{2b}$, an M-QAM signal, $I_m g(t - T_s)$, or carries $2b$ bits of information per T_s seconds, while each corresponding $\sqrt{M}$-PAM signal, $I_{mr}g(t-Ts)$ or $I_{mi}g(t-Ts)$, carries b bits/T_s seconds. Thus, the QAM- and PAM-modulated signals have equal bandwidths $W = 1/T_s$. As defined previously, the bandwidth efficiencies (data rate (bits/sec) per signal bandwidth) of the M-QAM and the corresponding $\sqrt{M}$-PAM signal constellations are

$$Bandwidth\ efficiency\ of\ M-QAM = \frac{\frac{2b\ bits}{T_s\ sec.}}{1/T_s\ Hz} = 2b\ bits/Hz \tag{2.5}$$

$$Bandwidth\ efficiency\ of\ \sqrt{M}-PAM = \frac{\frac{b\ bits}{T_s\ sec.}}{1/T_s\ Hz} = b\ bits/Hz \tag{2.6}$$

Hence, with an M-QAM signal set a higher bandwidth efficiency (by a factor of 2) than a $\sqrt{M} - PAM$ signal set is obtained.

Furthermore, an alternative view of each of the QAM-modulated bandpass waveforms (as indicated in (2.7)) indicates modulation of the carrier in both amplitude and phase - the factor of 2 is ignored for simplicity.

$$x_{tx}(t) = \Re[I_m g(t)e^{j\omega_c t}] = \Re[A_m e^{j\theta_m} g(t)e^{j\omega_c t}] = A_m g(t)\cos(\omega_c t + \theta_m) \tag{2.7}$$

where $A_m = \sqrt{I_{mr}^2 + I_{mi}^2}$ represents the amplitude modulation, and $\theta_m = \tan^{-1}(\frac{I_{mi}}{I_{mr}})$ represents the phase modulation of the carrier.

In an M-QAM technique, the assignment of b information bits to the $M = 2^b$ possible symbols in the signal constellation can be done in various ways. One simple method is *Gray encoding*, as shown in Fig. 2-4, with the property that each symbol differs from its

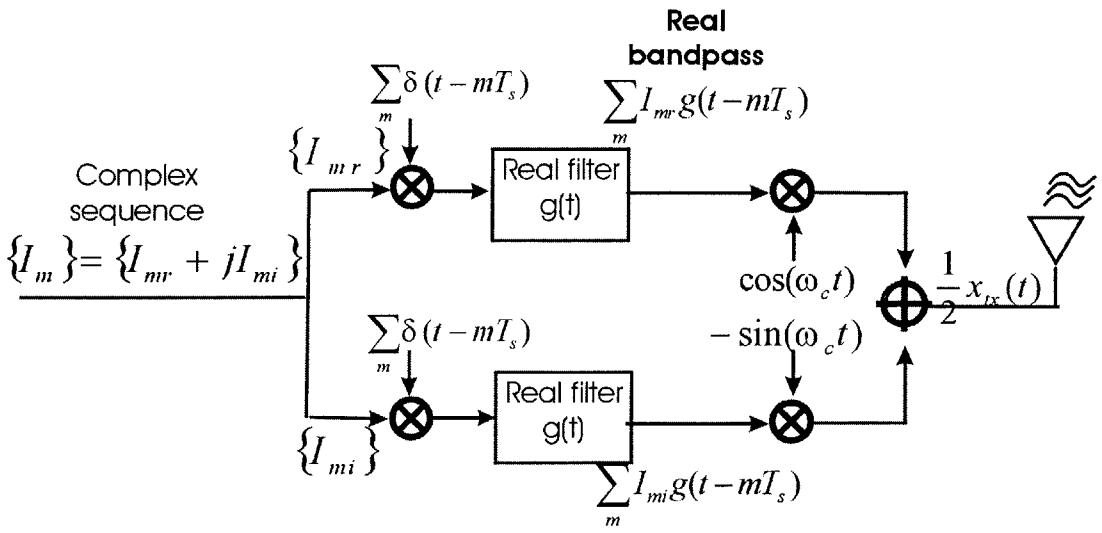

Figure 2-3: An alternative realization of QAM modulation in which the real symbol sequences $\{I_{mr}\}$ and $\{I_{mi}\}$ are acted upon individually. The final real bandpass signal, $x_{tx}(t)$, is obtained by adding the two real bandpass signals on the quadrature carriers.

neighbor symbols by exactly one bit. Such a property is advantageous in detection since the most likely errors caused by circuit nonlinearities and noise in the system involve the erroneous selection of an adjacent symbol to the transmitted symbol. In Gray encoding, a symbol error corresponds to only a single bit error even though the symbol carries b bits of information.

As indicated by (2.2) and Figures 2-2 and 2-3, once the discrete information-bearing sequence of symbols I_m are produced according to a particular mapping, they go though a pulse-shaping (baseband) filter with an impulse response $g(t)$ to transform to analog waveforms to be transmitted over the analog channel. However, to ensure that the signal waveforms are band-limited with minimum inter-symbol interference (ISI), such a filter must have certain characteristics that are presented in the next section.

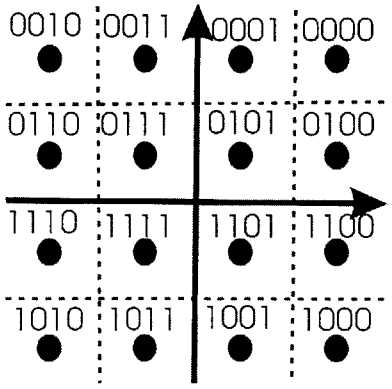

Figure 2-4: Gray encoding of a 16-QAM signal constellation.

2.2.2 Pulse-Shaping (Baseband) Filter

The individual symbols in any baseband waveform, $x(t) = \sum_{m \in Z} I_m \delta(t - mT)$, exist for only a finite duration in time, resulting in the spectrum of such waveforms to extend over all frequencies [5]. As a result, when transmitted over a band-limited channel, some frequency components fall outside the channel bandwidth, thus causing adjacent-channel interference (ACI). In fact, the amount of allowable ACI, commonly expressed by the adjacent-channel power (ACP) or spectral mask, is specified by radio standards. To prevent ACI, these waveforms must be filtered before sending over the channel; this filtering of baseband waveforms is known as *baseband filtering.*

While baseband filtering desirably limits the bandwidth of the transmitted waveforms to suppress spectral leakage into the adjacent channels, it has an adverse effect that cannot be ignored: bandwidth-limiting (in the frequency-domain) results in expansion of the signal waveforms in the time-domain; thus, individual symbols overlap and inter-symbol interference (ISI) occurs. Hence, we are faced with a tradeoff between ACI (also bandwidth efficiency) and ISI: unfiltered signals occupy too much bandwidth (bandwidth-inefficient) while filtered signals cause ISI. Since signals must be band-limited, the objective is to minimize ISI by choosing the filter carefully. In fact, there exist necessary and sufficient conditions, known as the *Nyquist Criterion*, on the frequency characteristics of the band-limited signal $X(f)$ that results in zero ISI [1]. The condition for no ISI is that,

$$x(t = mT) = \begin{cases} 1, & (m=0) \\ 0, & (m \neq 0) \end{cases} \tag{2.8}$$

According to the Nyquist theorem, as proven in [1], the necessary and sufficient condition for $x(t)$ to satisfy the zero-ISI condition (2.8) is that its Fourier transform, $X(f)$, satisfies

$$\sum_{m=-\infty}^{\infty} X(f + \frac{m}{T}) = T. \tag{2.9}$$

Furthermore, the baseband signal $x(t)$ - with $X(f) = 0$ for $|f| > W$ - satisfies the condition in (2.9) only if

$$W \geq \frac{1}{2T}. \tag{2.10}$$

As shown in Fig. 2-5, when $W < \frac{1}{2T}$, since $B(f) = \sum_{m=-\infty}^{\infty} X(f + \frac{m}{T})$ consists of non-overlapping replicas of $X(f)$, separated by $\frac{1}{T}$, there is no choice for $X(f)$ to satisfy $B(f) \equiv T$. Hence, such choice of $X(f)$ results in ISI.

On the other hand, when $W = \frac{1}{2T}$ (where $\frac{1}{T} = 2W$ is the Nyquist rate), there exists only one choice for $X(f)$ that satisfies $B(f) \equiv T$, that is $X(f) = \begin{cases} T, & (|f|<W) \\ 0, & (otherwise) \end{cases}$ (Fig. 2-6). In other words, $X(f)$ has a rectangular spectral shape, corresponding to the *sinc* function in the time-domain as below

$$x(t) = \frac{sin(\pi t/T)}{\pi t/T} \equiv sinc(\frac{\pi t}{T}) \tag{2.11}$$

This suggests that to obtain zero ISI, the minimum bandwidth $X(f)$ can have is $W_{min} = \frac{1}{2T}$, which corresponds to a *sinc* time response. However, the non-causality (therefore non-realizability) and slow convergence-to-zero behavior of the *sinc* function make it an undesirable choice for practical filter implementations.

Finally, when $W > \frac{1}{2T}$, $B(f)$ consists of overlapping replicas of $X(f)$ separated by $\frac{1}{T}$ (Fig. 2-7). To meet $B(f) \equiv T$ for this case, there exist many choices. The pulse spectrum with desirable spectral properties that is widely used in practice has a raised-cosine characteristics: the frequency spectrum has a flat amplitude portion and a roll-off portion that has a sinusoidal form.

$X(f) = T, \quad (0 \leq |f| \leq \frac{1-\beta}{2T})$

$X(f) = \frac{T}{2}\{1 + \cos[\frac{\pi T}{\beta}(|f| - \frac{1-\beta}{2T})]\}, \quad (\frac{1-\beta}{2T} \leq |f| \leq \frac{1+\beta}{2T})$

$X(f) = 0, \quad (|f| > \frac{1+\beta}{2T})$

In the time-domain, $x(t)$ is

$$x(t) = sinc(\pi t/T)\frac{\cos(\pi\beta t/T)}{1-(2\beta t/T)^2} \tag{2.12}$$

where $x(t)$ is normalized such that $x(0) = 1$.

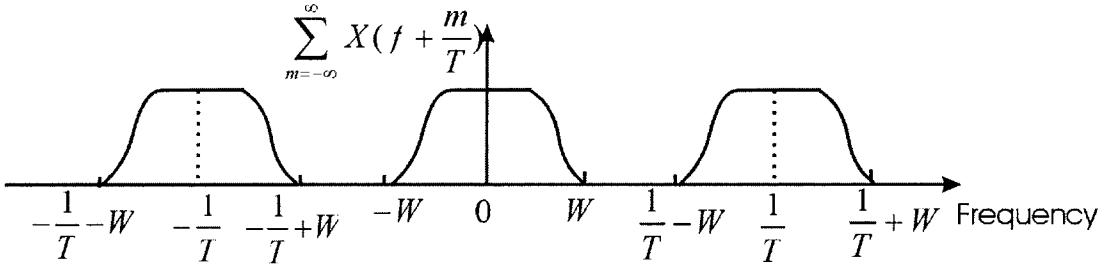

Figure 2-5: Plot of $B(f)$ when $W < \frac{1}{2T}$.

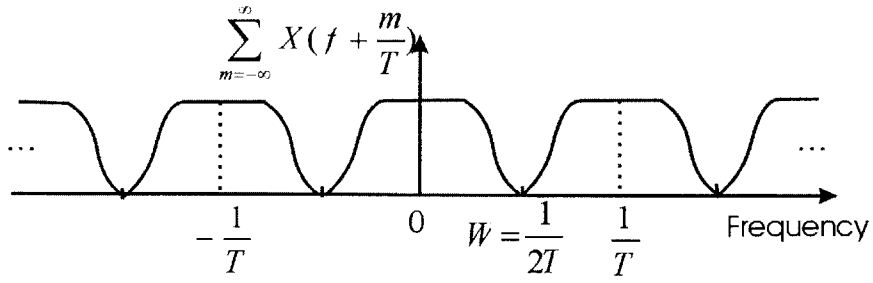

Figure 2-6: Plot of $B(f)$ when $W = \frac{1}{2T}$.

The rate of decay of the raised-cosine spectrum is controlled by the roll-off factor, β, which takes values in the range $0 \leq \beta \leq 1$. When $\beta = 1/2$, the signal bandwidth beyond the Nyquist frequency, $\frac{1}{2T}$, is 50 percent of this frequency, while when $\beta = 1$, such excess bandwidth is 100 percent. Fig. 2-8 illustrates the raised-cosine spectral characteristics and the corresponding pulses in time-domain for $\beta = 0$, $\frac{1}{2}$, *and* 1. The comparison between the various curves indicates that in the time-domain, higher β is more desirable - resulting in less ISI - since the decay rate is faster. On the other hand, since less spreading in the time-domain results in more expansion in the frequency-domain, higher β gives more spectral

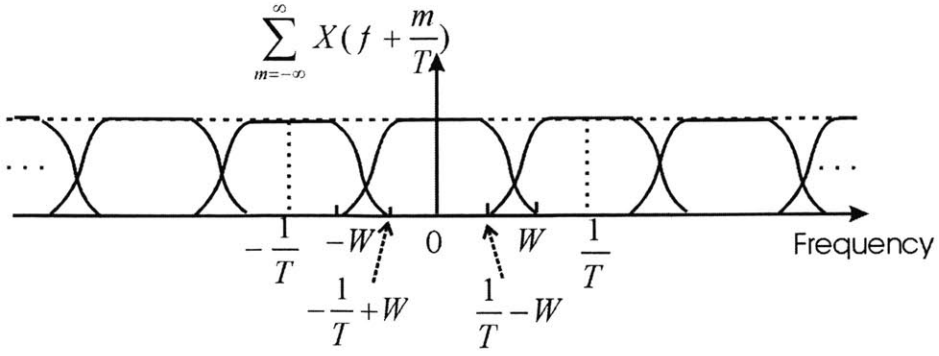

Figure 2-7: Plot of $B(f)$ when $W > \frac{1}{2T}$.

bandwidth - resulting in more ACI. Thus, a particular value of roll-off factor presents a tradeoff between ISI and ACI (or bandwidth efficiency).

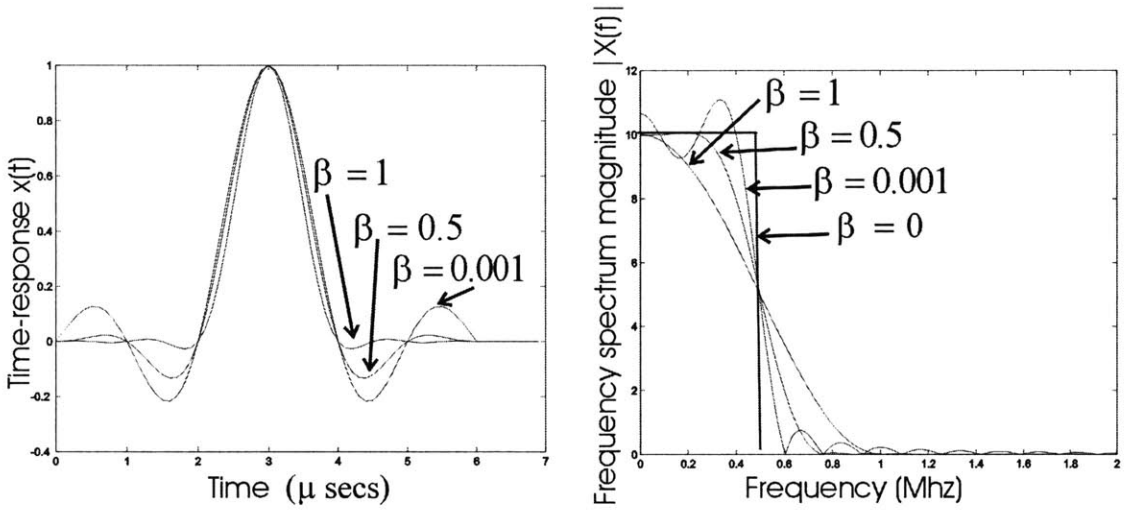

Figure 2-8: Effect of the roll-factor on the time response and frequency spectrum of the raised-cosine filter. The time and frequency-domain characteristics are depicted for the roll-off factors $\beta = 0.001,\ 0.5,\ 1$.

2.3 Channel

A band-limited channel is generally modeled mathematically as a linear filter with an additive noise, as depicted in Fig. 2-9.

The linear filter is characterized with an equivalent low-pass frequency response, $C(f)$, of bandwidth W (i.e. $C(f) = 0$ for $|f| > W$). The filter's frequency response can be expressed as

$$C(f) = |C(f)|e^{j\theta(f)} \tag{2.13}$$

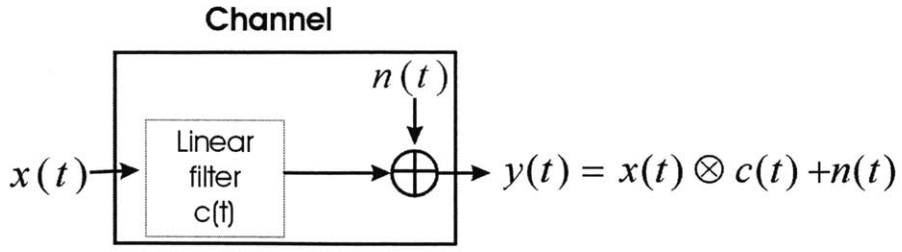

Figure 2-9: A model of a band-limited channel: a linear filter with an additive noise. ($\otimes$ represents the convolution operation.)

An ideal channel is one with a constant amplitude response $|C(f)|$ for all $|f| \leq W$, and a phase response $\theta(f)$ that is a linear function of the frequency. On the other hand, a non-ideal channel has a non-constant $|C(f)|$ and/or a nonlinear $\theta(f)$. Thus, the non-ideal channel results in distortion of the transmitted signal in amplitude and/or phase.

2.4 Optimum Receiver

When the signal arrives at the receiving antenna, it needs to be digitized by a demodulator and then decoded into the symbols in the original symbol constellation. In digital communications, an optimum receiver is typically one that minimizes the probability of error. As reasoned in the next two sections, to design such a receiver, the demodulator must consist of a matched filter and an optimum sampler to give the maximum signal-to-noise ratio at its output, and the detector (or decoder) must be optimal.

2.4.1 Demodulator

Reversing the action of the modulator, the demodulator is an interface device that maps the received analog waveforms into digital information. Indeed, the method chosen to extract the digital data from the received analog waveforms has great impact on the overall performance of the system. More precisely, as it will become clear in this section, the error rate of the system is primarily affected by the signal-to-noise ratio (SNR) at the output of the demodulator. Hence, the objective is to design a demodulator that gives the maximum possible SNR at the optimum sampling instant.

At the receiving antenna, the signal has been corrupted by noise that is random in nature. Therefore, averaging over one symbol period would average out the noise components that vary significantly in a symbol period. Such averaging can be performed by using a filter with appropriate characteristics. We will show that for a signal that is corrupted by an additive white Gaussian noise process(AWGN), there exists an optimum filter that performs the averaging operation and as a result gives the maximum SNR at the optimum sampling point; such a filter is known as the *matched filter* since the shape of its impulse-response is matched to the shape of the transmitted signal waveform [2].

Suppose that $r(t) = x(t) + n(t)$ is the received signal, where the transmitted signal $x(t)$ (with symbol period T) is corrupted by an AWGN process, $n(t)$. Let $h(t)$ be the impulse response of the matched filter which is causal (that is $h(t) = 0$ for $t < 0$) since the signal waveform is causal. At the output of the filter we have

$$y(t) = \int_0^t r(\tau)h(t-\tau)d\tau = \int_0^t x(\tau)h(t-\tau)d\tau + \int_0^t n(\tau)h(t-\tau)d\tau \tag{2.14}$$

The resulting signal is then sampled at its optimum point, $t = T$ - where the signal has its peak value as a consequence of the averaging operation of the matched filter. Thus, we obtain

$$y(T) = \int_0^T x(\tau)h(T-\tau)d\tau + \int_0^T n(\tau)h(T-\tau)d\tau = y_s(T) + y_n(T) \tag{2.15}$$

To compute $SNR_{out} = \frac{y_s^2(T)}{E[y_n^2(T)]}$, we need to find the average signal power, $y_s^2(T)$, and the average noise power, $E[y_n^2(T)]$, at the output of the sampler [1]. The average power in the signal waveform can be expressed as

$$y_s^2(T) = [\int_0^T x(\tau)h(T-\tau)d\tau]^2 \tag{2.16}$$

The average noise power (with noise variance $\sigma_n^2 = \frac{N_0}{2}$) is obtained by

$E[y_n^2(T)] = \int_0^T \int_0^T E[n(\tau)n(t)]h(T-\tau)h(T-\tau)dtd\tau$

$E[y_n^2(T)] = \frac{N_0}{2} \int_0^T \int_0^T \delta(t-\tau)h(T-\tau)h(T-\tau)dtd\tau$

$E[y_n^2(T)] = \frac{N_0}{2} \int_0^T h^2(T-t)dt$

Thus, SNR_{out} can be evaluated as,

$$SNR_{out} = \frac{[\int_0^T x(\tau)h(T-\tau)d\tau]^2}{\frac{N_0}{2}\int_0^T h^2(T-t)dt} = \frac{[\int_0^T h(\tau)x(T-\tau)d\tau]^2}{\frac{N_0}{2}\int_0^T h^2(T-t)dt} \tag{2.17}$$

As mentioned earlier in this section, to obtain an optimum demodulator, the filter's transfer characteristics must be chosen such that SNR_{out} is maximized. Since the noise power density is fixed and uncontrollable, SNR_{out} is maximized when its numerator is at maximum. Let $g_1(t) = h(t)$ and $g_2(t) = x(T-t)$, then by Cauchy-Schwartz inequality [1] we have

$$\int_{-\infty}^{\infty}(g_1(t)g_2(t)dt)^2 \leq \int_{-\infty}^{\infty} g_1^2(t)dt \int_{-\infty}^{\infty} g_2^2(t)dt \tag{2.18}$$

In (2.18), the equality is met when $g_1(t) = \mathrm{C} \cdot g_2(t)$ (where C is an arbitrary constant). In other words, SNR_{out} is at maximum when the filter response $h(t)$ is matched to the transmitted signal waveform $x(t)$, that is when

$$h(t) = C \cdot x(T-t) \tag{2.19}$$

Substituting (2.19) in (2.17), we obtain

$$max(SNR_{out}) = \frac{\int_0^T x^2(T-\tau)d\tau C^2 \int_0^T x^2(T-\tau)d\tau}{\frac{N_0}{2}C^2\int_0^T x^2(T-\tau)d\tau} = \frac{2}{N_0}\int_0^T x^2(t)dt = \frac{2E_s}{N_0} \tag{2.20}$$

where E_s is the energy under $x(t)$.

Similarly, the above result for SNR_{out} can be obtained by working in the frequency-domain. The Fourier transform of the matched filter can be expressed as,

$$h(t) = x(T-t) \leftrightarrow H(f) = \int_0^T x(T-t)e^{-j2\pi ft}dt = (\int_0^T x(\tau)e^{j2\pi f\tau}d\tau)e^{-j2\pi fT} \quad (2.21)$$

The last expression can be recognized as

$$H(f) = X^*(f)e^{-j2\pi fT} \quad (2.22)$$

where $X^*(f)$ is the conjugate of $X(f)$ and $e^{-j2\pi fT}$ represents the sampling delay. In fact, (2.22) indicates that the spectrums of the matched filter and signal waveform have the same magnitude response, $|H(f)| = |X(f)|$. The signal component at the output of matched filter, in time- and frequency-domain, is

$$y_s(t) = x(t) \otimes h(t) \leftrightarrow Y_s(f) = X(f) \times H(f) \quad (2.23)$$

where $\otimes$ represents the convolution operation, and $\times$ is a multiplication. Substitution of (2.22) in (2.23) yields

$$Y_s(f) = |X(f)|^2 e^{-j2\pi fT} \leftrightarrow y_s(t) = \int_{-\infty}^{\infty} |X(f)|^2 e^{-j2\pi fT} e^{j2\pi ft}dt \quad (2.24)$$

Next, $y_s(t)$ is sampled at its optimum sampling point $t = T$. Using (2.24) and Parseval's theorem [1] (that the energy under a signal in time-domain is equal to the energy contained in its frequency spectrum), we obtain the signal energy, E_s.

$$y_s(T) = \int_{-\infty}^{\infty} |X(f)|^2 dt \equiv \int_0^T x^2(t)dt = E_s \quad (2.25)$$

The total signal power is thus,

$$P_s = y_s^2(T) = E_s^2 \quad (2.26)$$

Meanwhile, the noise component at the output of matched filter has the power spectral density $\Phi_n(f) = \frac{1}{2}|H(f)|^2 N_0$. So, the total noise power P_n will be

$$P_n = \int_{-\infty}^{\infty} \Phi_n(f)df \quad (2.27)$$

$$P_n = \frac{N_0}{2}\int_{-\infty}^{\infty} |H(f)|^2 df = \frac{N_0}{2}\int_{-\infty}^{\infty} |X(f)|^2 df = \frac{N_0}{2}E_s \quad (2.28)$$

Consequently, using the total signal and noise powers (in Equations 2.26 and 2.28), the SNR at the output of demodulator can be evaluated as

$$SNR_{out} = \frac{P_s}{P_n} = \frac{E_s^2}{\frac{N_0}{2}E_s} = \frac{2E_s}{N_0} \quad (2.29)$$

Therefore, by using the frequency interpretation of the matched filter, we obtained the same maximum SNR_{out} as in (2.20). Moreover, an important observation that can be made from SNR expression in (2.29) is its dependence on the energy (or power) of the transmitted signal, regardless of its detailed shape and characteristics.

2.4.2 Detector (or Decoder)

After the received signal (over a signal period) is digitalized by the demodulator, it needs to be decoded to one of the signals in the original signal constellation, denoted by A.

Decision Rules

As discussed earlier, we are interested in an optimum detector, one that is based on the minimum-probability-of-error (MPE) decision rule [6]:

> Given a received signal y, choose a signal $\hat{a} \in A$ that minimizes the probability of decision error.

In fact, the probability that a decision $\hat{a}$ is correct is the *a posteriori probability* $p(\hat{a}|y)$. Hence, the MPE rule is equivalent to the maximum-a-posteriori-probability rule (MAP), stated as:

> A signal $\hat{a} \in A$ is chosen such that $p(\hat{a}|y) = max\{p(a_j|y),\ among\ all\ a_j \in A\}$.

By Bayes' Law, we have

$$p(a_j|y) = \frac{p(y|a_j)p(a_j)}{P(y)} \tag{2.30}$$

where $p(y|a_j)$ is called *the likelihood probability*, $p(y)$ is the probability that y is received (which is the same regardless of the transmitted signal), and $p(a_j)$ is the a priori probability that a_j was transmitted. In fact, when all signals are transmitted with equal probability, the MAP rule becomes equivalent to the maximum likelihood rule (ML), expressed as:

> Choose $\hat{a} \in A$ such that $p(y|\hat{a} = max\{p(y|a_j),\ among\ all\ a_j \in A\}$.

Finally, under an AWGN channel, the received signal sequence is $\{y = a_j + n\}$ where the noise $\{n\}$ is a sequence of independent, identically distributed (iid) Gaussian random variables with mean zero, variance $\sigma^2 = N_0/2$ and a probability density function (pdf),

$$p_N(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-n^2}{2\sigma^2}} \tag{2.31}$$

Using the noise pdf (2.31), the likelihood function can be evaluated as

$$p(y|a_j) = p_N(y - a_j) = p_N(y - a_j) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-a_j)^2}{2\sigma^2}} \tag{2.32}$$

From expression 2.32, it is apparent that $p(y|a_j)$ is maximized when $(y - a_j)^2$ is minimized. In addition, since y and a_j denote discrete signals, or points in the signal-space,

$$min(y - a_j)^2 \equiv min|y - a_j|^2 \equiv min|y - a_j| \tag{2.33}$$

where $|y - a_j|$ denotes the distance between the two points in the signal constellation. Hence, the ML rule is equivalent to the minimum-distance (MD) rule:

> The signal $\hat{a} \in A$ is chosen such that $|y - \hat{a}| = min\{|y - a_j|\ among\ all\ a_j \in A\}$.

Consequently, for equiprobable transmitted signals going under an AWGN channel, the minimum-distance (MD) rule is equivalent to the minimum-probability-of-error (MPE) detection rule. In concise, an optimum detector chooses a signal in the signal set, A, that is closest (in distance) to the received signal, y.

The Union Bound and the Union Bound Estimate (UBE)

Pairwise error probability:

Consider a signal set that consists of two signals a_1 and a_2; a_1 is sent, and the corrupted received signal is $y = a_1 + n$ (where n is a white Gaussian random variable with pdf as given in (2.31). Using the MD rule, an error in detection occurs if the received signal y is detected closer to a_2 than a_1. Suppose $Pr\{a_1 \rightarrow a_2\}$ denotes such pairwise error probability, and $d(a_1, a_2)$ denotes the distance between the two signals. As indicated in Fig. 2-10, $Pr\{error|a_1 \; is \; sent\}$ is the marked area under the Gaussian pdf $p(y|a_1)$ (with mean a_1 and variance σ^2). In other words,

$$Pr\{error|a_1\} = \int_{a_1+d(a_1,a_2)/2}^{\infty} p(y|a_1)dy = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{a_1+d(a_1,a_2)/2}^{\infty} e^{\frac{-(y-a_1)^2}{2\sigma^2}} dy \tag{2.34}$$

Substituting $n = y - a_1$ in (2.34) we obtain,

$$Pr\{error|a_1\} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{d(a_1,a_2)/2}^{\infty} e^{\frac{-n^2}{2\sigma^2}} dn = Q(\frac{d(a_1, a_2)}{2\sigma}) \tag{2.35}$$

where $Q(x)$ - the area from $x \rightarrow \infty$ under the pdf of a Gaussian random variable with mean 0 and variance $\sigma^2 = 1$ - is given by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{\frac{-x^2}{2}} dx \tag{2.36}$$

As the pairwise error probability in (2.35) indicates, the error probability of choosing signal a_2 for a_1 depends on the distance $d(a_1, a_2)$ between a_1 and a_2 and the noise standard deviation σ.

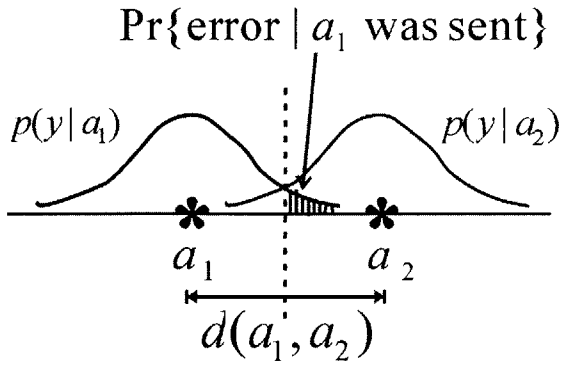

Figure 2-10: A plot showing the error probability in detecting the corrupted symbol y when a_1 was sent under the MD rule. Since noise $n > d(a_1, a_2)/2$, it moves the transmitted symbol a_1 closer to the other symbol a_2, which is thus mistakenly detected as the transmitted symbol.

Non-binary error probability:

However, for non-binary signal set A, in general, one cannot obtain a closed-form expression for the error probability of an optimum (MD) detector. Instead, we can use an upper bound,

called the *the union bound*, in terms of pairwise error probabilities. Furthermore, there exists a dominant term in the union bound, known as the *union bound estimate* (UBE), that provides an excellent approximation in most cases. This section derives the union bound and the UBE for the error probability.

Based on the fundamental union bound of probability theory: $Pr\{A \cup B\} \leq Pr\{A\} + Pr\{B\}$ (for any two events A and B), the union bound on error probability states that:

> $Pr\{MD\ detection\ error|a_j\ is\ sent\}$ - i.e. the probability that the received signal y is closer to some other signal $a_{j'} \in A$ than to a_j - is upper-bounded by the sum of the pairwise error probabilities, $Pr\{a_j \rightarrow a_{j'}\}$, to all other signals $a_{j'} \neq a_j \in A$. In other words,
>
> $$Pr\{error|a_j\} \leq \sum_{a_{j'} \in A, a_{j'} \neq a_j} Pr\{a_j \rightarrow a_{j'}\} = \sum_{a_{j'} \in A, a_{j'} \neq a_j} Q(\frac{d(a_j, a_{j'})}{2\sigma}). \quad (2.37)$$
>
> As observed earlier, since the pairwise error probabilities depend primarily on the corresponding pairwise distances, the terms in the union bound (2.37) can be grouped by their distances. Thus, suppose that D is the set of all distances $d(a_j, a_{j'})$ in A, and $K_d(a_j)$ is the number of signals in A at distance d from a_j; the above bound can then be re-expressed as
>
> $$Pr\{error|a_j\} \leq \sum_{d \in D} K_d(a_j) Q(\frac{d}{2\sigma}) \quad (2.38)$$

In addition, an important property of the $Q(x)$ function is that it decays exponentially with x^2 (as proved by Forney [6]) according to

$$Q(x) \approx e^{\frac{-x^2}{2}} \quad (2.39)$$

Using this property in the union bound (2.38), we obtain

$$Pr\{error|a_j\} \leq \sum_{d \in D} K_d(a_j) e^{(\frac{-(\frac{d}{2\sigma})^2}{2})} \quad (2.40)$$

> Since the individual terms in (2.40) decay exponentially with larger distances, the terms with minimum-distance, $d_{min} = min\{all\ distances \in D\}$, dominate the sum. In fact, these $K_{min}(a_j)$ terms constitute the UBE of $Pr\{error|a_j\}$, represented as
>
> $$Pr\{error|a_j\} \approx K_{min}(a_j) Q(\frac{d_{min}}{2\sigma}) \quad (2.41)$$
>
> In other words, the UBE is based on the idea that a signal a_j can be at most misdetected for one of its nearest neighbors at distance d_{min} from a_j. As a result, such estimate is valid if the next nearest neighbors are at sufficiently greater distance and there are not too many of them.

Using the UBE for the error probabilities of individual signals in the signal set A, we obtain the total error probability $Pr(E)$ by averaging over all such error probabilities

$$Pr\{error\} = \overline{P(error|a_j)} \approx K_{min} Q(\frac{d_{min}}{2\sigma}) \tag{2.42}$$

where the overbar denotes averaging over all the signals $a_j \in A$, and $K_{min} = \overline{K_{min}(a_j)}$. Consequently, as Forney [6] discusses, the dependence of the error probability with optimum (MD) detection on the ratio $d_{min}/2\sigma$ (to the accuracy of the UBE) suggests that in constellation designs the objective is to maximize d^2_{min} for a given average energy per signal E_s of the constellation, or equivalently to minimize E_s for a given d^2_{min}, or one can design a constellation for which the figure-of-merit $d^2_{min}/4E_s$ is optimized. In addition, a secondary objective is to minimize the error coefficient K_{min}.

Error Probability of M-PAM and $\sqrt{M} \times \sqrt{M}$-QAM (M-QAM) Constellations

For an M-QAM signal with $M = 2^k$ and k even, since the M-QAM signal is equivalent to two $\sqrt{M}$-PAM signals on quadrature-phase carriers, the two PAM signals can be separated at the demodulator. Hence, an M-QAM signal is detected correctly provided that the corresponding two $\sqrt{M}$-PAM signals are detected correctly, i.e.

$$Pr_{(M-QAM)}\{correct\} = (Pr_{(\sqrt{M}-PAM)}\{correct\})^2 = (1 - Pr_{(\sqrt{M}-PAM)}\{error\})^2 \tag{2.43}$$

where $Pr_{(\sqrt{M}-PAM)}\{error\}$ denotes the error probability of a $\sqrt{M}$-PAM signal set. Next, we derive the error probability for an M-PAM signal set:

Let $A = \{\pm d/2, \pm 3d/2, \ldots, \pm(M-1)d/2\}$ be an M-PAM signal set with equiprobable signals and $d = d_{min}$. To find the estimated total error probability under an MD detection going through an AWGN channel (as given in 2.42), we need to evaluate the average number of nearest neighbors as below,

$$K_{min} = \frac{2}{M}(1) + \frac{M-2}{M}(2) = 2(\frac{M-1}{M}) \tag{2.44}$$

In addition, as shown by Gallager [4], the average energy per symbol of an M-PAM signal set is

$$E_s = \frac{d^2(M^2-1)}{12} \tag{2.45}$$

Substituting (2.44) for K_{min} and (2.45) for d_{min} in the error probability expression (2.42), we obtain

$$Pr_{(M-PAM)}\{error\} = K_{min} Q(\sqrt{\frac{d^2}{4\sigma^2}}) = 2(\frac{M-1}{M}) Q(\sqrt{\frac{3E_s}{\sigma^2(M^2-1)}}) \tag{2.46}$$

In fact, the error probability (2.46) for an M-PAM signal constellation is exact [1].

To obtain the error probability for a $\sqrt{M} \times \sqrt{M}$-QAM given in (2.43), we first use (2.46) to obtain the error probability for a $\sqrt{M}$-PAM (2.46) with average symbol energy half of the average energy E_s in M-PAM, that is

$$Pr_{(\sqrt{M}-PAM)}\{error\} \approx 2(\frac{\sqrt{M}-1}{\sqrt{M}})Q(\sqrt{\frac{3E_s}{2\sigma^2(M-1)}}) \tag{2.47}$$

As discussed at the beginning of this chapter, the error probability is most commonly expressed in terms of the signal-to-noise ratio defined as

$$SNR = \frac{average\ signal\ energy}{average\ noise\ energy} = \frac{E_{avg}}{N_0} \tag{2.48}$$

where $N_0/2 = \sigma^2$ is the noise variance per real dimension and E_{avg} is the average energy per 2 dimensions. Moreover, in band-limited system designs, the figure-of-merit for the signal set is

$$SNR_{norm} = \frac{SNR}{2^\rho - 1} \tag{2.49}$$

which is the SNR normalized for spectral efficiency ρ (= number of bits per two-dimensions (2D)). For an M-PAM signal constellation with $M = 2^b$ (b = bits), $\rho = 2b$ bits/2D; thus,

$$SNR_{norm} = \frac{SNR}{2^{2b} - 1} = \frac{SNR}{M^2 - 1} \tag{2.50}$$

Substituting SNR (2.48) (where $E_{avg} = 2E_s$) and E_s (2.45) in (2.50), we obtain

$$SNR_{norm}(M - PAM) = \frac{E_s}{\sigma^2(M^2 - 1)} \tag{2.51}$$

Furthermore, for a $\sqrt{M}$-PAM with $E_s/2$ average energy per symbol, we have

$$SNR_{norm}(\sqrt{M} - PAM) = \frac{E_s}{2\sigma^2(M - 1)} \tag{2.52}$$

As a result, using (2.47) and (2.52), we obtain the average error probability for an M-QAM (2.43) in terms of SNR_{norm} as below,

$$Pr_{(M-QAM)}\{error\} = (1 - Pr_{(\sqrt{M}-PAM)}\{error\})^2 \tag{2.53}$$

where,

$$Pr_{(\sqrt{M}-PAM)}\{error\} \approx 2(1 - \frac{1}{\sqrt{M}})Q(\sqrt{3SNR_{norm}}) \tag{2.54}$$

Chapter 3

Power Amplifier Characteristics and Nonlinearity Effects - Background

This chapter provides some background on power amplifiers (PAs) that is sufficient to better appreciate the goal of this thesis work and understand the results from the circuits point of view. The chapter starts with an introduction of the input-output characteristics of a power amplifier - i.e. how a PA responds to an input signal at different levels, and how a nonlinear response differs from a linear (ideal) response. Next, the adverse effects of the nonlinearity of a PA are introduced by examining its response to single- and two-tone sinusoids, which is the traditional approach in analyzing the nonlinear effects of a PA.

3.1 Input-Output Characteristics of PA

Consider a two-port network model of a PA, as shown in Figure 3-1, with an input voltage $x(t)$, an output voltage $y(t)$, and a transfer function $H(\omega)$ [5].

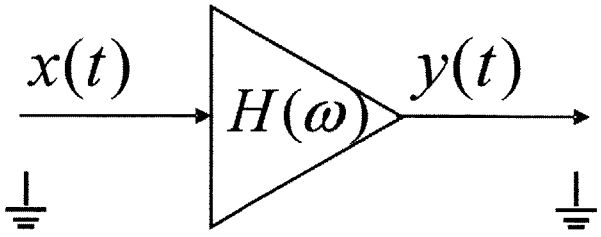

Figure 3-1: Power Amplifier two-port model.

For simplicity, we limit our analysis to a time-invariant, memoryless system in which the output at any point in time is a function of the input at that moment, independent of the previous values. To study the PA nonlinearity, first we need a model of the PA transfer characteristics. In this section, we present general models that are most suitable for our study. We start with introducing the characteristics of a linear PA, and subsequently show how the nonlinear characteristics deviate from this ideal response.

3.1.1 Input-Output Characteristics - Linear PA

In an ideal (linear) PA, the output voltage is a constant, a_1, times the input voltage, i.e.

$$y(t) = a_1 \cdot x(t) \tag{3.1}$$

Since the input and output voltages are vectors (having both amplitude and phase), the constant, a_1, is a complex scaler with an amplitude, $|a_1|$, and a phase, $\angle a_1$. More explicitly, (3.1) can be re-expressed as,

$$y(t) = |a_1||x(t)| \cdot e^{j(\angle x(t) + \angle a_1)} \tag{3.2}$$

From (3.2), the output voltage amplitude, $|y(t)|$, and phase, $\angle y(t)$, are given by,

$$|y(t)| = |a_1||x(t)| \tag{3.3}$$
$$\angle y(t) = \angle x(t) + \angle a_1 \tag{3.4}$$

Hence, the output voltage amplitude is a linear amplification of the input voltage amplitude - i.e. the gain, $|a_1|$, is constant across all input amplitudes. In addition, the output voltage phase is the input voltage phase plus a constant phase-shift, $\angle a_1$ - which is the time delay between the input and output signals at the given frequency.

Similarly, in terms of the frequency response, $H(\omega)$, an ideal PA has constant frequency characteristics - that is, constant magnitude (gain) and linear phase (constant delay) over the bandwidth of the input signal. In addition, such a PA is memoryless.

3.1.2 Input-Output Characteristics - Nonlinear PA

In practice, the fundamental devices used in a PA, such as the transistors and capacitors, have nonlinearity characteristics that make the output voltage a function of higher order terms of the input voltage, resulting in a *nonlinear* input-output characteristic. Also, typically a nonlinear PA has a frequency-dependent and input amplitude-dependent gain, a nonlinear phase, and memory.

Power Series Model

In the most general way, a nonlinear PA has an input-output voltage characteristics that is linear for small input voltages and becomes nonlinear at higher input levels. Such a PA can be modeled mathematically by a Taylor series (power series) such that

$$y(t) = a_1 x(t) + a_2 x^2(t) + a_3 x^3(t) + \ldots + a_n x^n(t) \tag{3.5}$$

Figure 3-2 shows a typical voltage (amplitude) transfer function of a nonlinear PA that can be modeled by (3.5), with the linear response drawn for comparison. The PA constants, $a_{1\ldots n}$, determine the exact shape of the input-output characteristics, where a_1 is the linear gain of the PA and higher-order coefficients characterize the strengths of the corresponding nonlinearities. More specifically, the third-order term, a_3, causes gain compression at both ends while the second-order term, a_2, causes gain expansion at the positive end and gain compression at the negative end [5]. In fact, both terms are generally present to a greater or lesser extent, resulting in the distortion of the output signal at all input signal levels.

Also, many analysis truncate the power series to include second- and third-order effects only. Although third-order nonlinearity is an important contributor to the compression and saturation nonlinearity effects, the fifth- and seventh-order terms become significant as the 1-dB compression point is approached and can dominate at higher drive levels. The *1-dB compression point* is defined as the input power (or voltage) level at which the nonlinear output power is 1-dB below the linear output power (Figure 3-3) [2]. Hence, the restriction of the power series model to the third-order nonlinearity suffices when studying the PA at operation levels below the 1-dB compression point, as it is the case in this thesis.

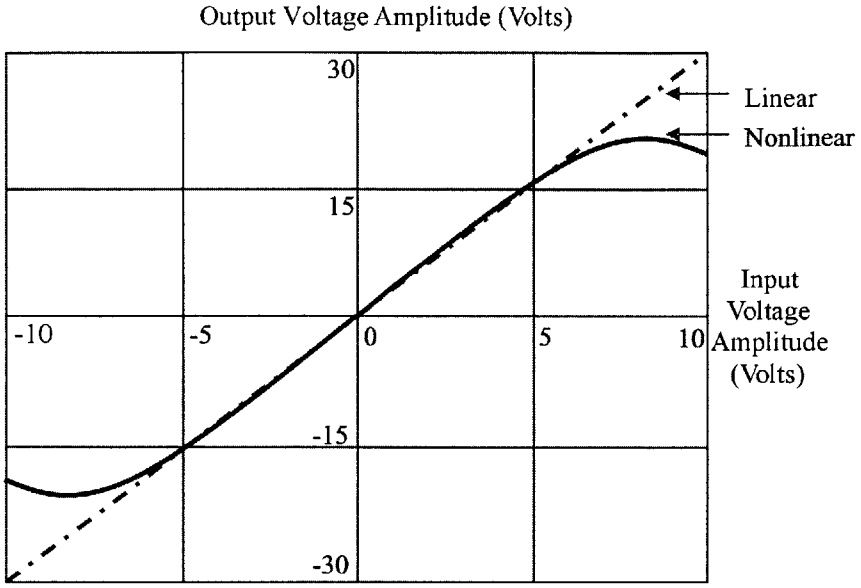

Figure 3-2: Input-output voltage characteristics (Source: Pothecary [5]).

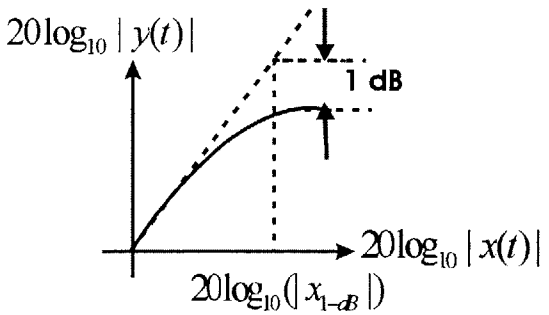

Figure 3-3: Definition of the 1-dB compression point (Source: Razavi [2]).

However, although the above power series is used in many references as a generalized formulation for the nonlinear behavior, it has some limitations [7]. For instance, there is no phase component in the linear and nonlinear terms. As a result, the power series characterization of the PA is limited to modeling the nonlinearity behavior in a small operating zone in the immediate vicinity of the dc operating point. However, over a larger operating zone,

the output voltage indicates both amplitude and phase variations from the input voltage. Hence, a more suitable power series model is one in which the input voltage $x(t)$, output voltage $y(t)$ and coefficients $a_{1...n}$ are complex. In addition, when the PA has memory, an even stronger formulation of the power series, called the *Volterra series,* is used [3]. Still, the Volterra/power-series formulations are only useful for analyzing the weakly nonlinear properties of the PA, when the PA is backed-off from the 1-dB compression point.

Power amplifiers operating beyond the 1-dB compression point require a different treatment since the nonlinearities become strong and arise through the cutoff and clipping behavior of the transistor. Cripps [7] derives a comprehensive model for a nonlinear PA that characterizes both the weak and strong nonlinearities. However, such a model describes the nonlinearity behavior using device physics parameters. Hence, since we desire a behavioral model of the PA that is general and device independent, this model cannot be used for our study.

AM-AM/AM-PM Model

As mentioned above, voltages are vector quantities: they have both amplitude and phase. Therefore, for some studies, it might be more meaningful to treat the amplitude and phase nonlinearities separately. Such treatment resembles the method used for the frequency response with the difference that for the latter case the amplitude and phase responses are functions of the frequency rather than the input amplitude. Based on such an approach, there exists an amplitude-phase model that have been frequently used in the literature to study frequency-independent, bandpass and memoryless nonlinear power amplifiers. In this model, the portion of the output wave falling in the same spectral zone as the band-limited input wave is described in terms of the envelope of the input wave, rather than its instantaneous value. Such a model is specified by two functions: the AM-AM conversion and the AM-PM conversion (where AM denotes amplitude modulation and PM denotes phase modulation).

Consider the instantaneous amplitude-phase model of a nonlinear PA, expressed as,

$$y(t) = g(|x(t)|)e^{j[\angle x(t) + f(|x(t)|)]} \tag{3.6}$$

In the above model, (3.6), the output amplitude is a nonlinear function (AM-AM function $g(.)$) of the input amplitude, and the output phase is the input phase plus a phase-shift that is a nonlinear function (AM-PM function $f(.)$) of the input amplitude. Both the AM-AM and AM-PM functions depend on the type of the amplifier (e.g. solid-state (SSPA) or traveling-wave-tube (TWT)), the operating frequency, bandwidth or data-rate of the modulation scheme, and the class of operation in which the amplifier is used. Nevertheless, the AM-AM characteristic has the same approximate shape, as illustrated in Figure 3-4, across various amplifier types and operations. Depending on the application, the AM-AM function is modeled in various ways. These models include the truncated power series [8], Bessel series [9], Fourier series [10], empirical describing functions such as the 2-parameter formula in [11] or the 4-parameter formula in [12], and specific functions that are semi-physical based [13].

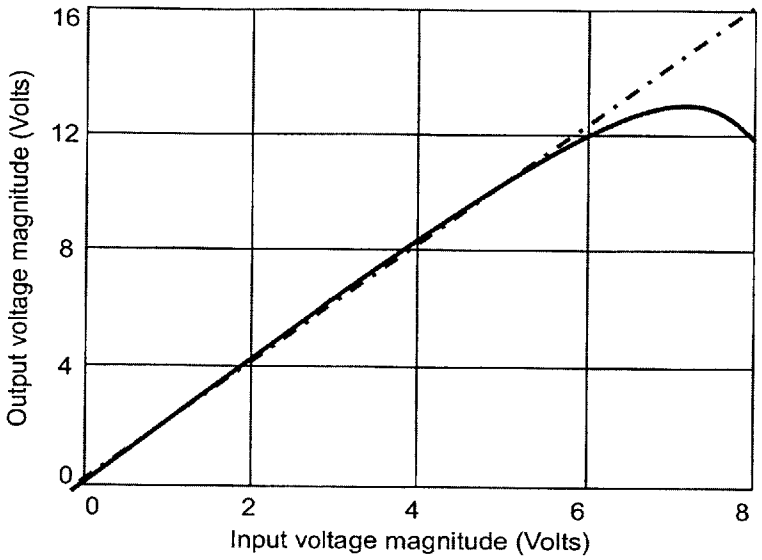

Figure 3-4: Power amplifier AM-AM characteristic.

In contrast, the shape of the AM-PM function varies significantly from one amplifier to another. In fact, there is a limited understanding of what exactly causes the peculiar behavior of the AM-PM nonlinearity. Figure 3-5 illustrates the AM-PM plots from the data measured from class A, AB (AB_1) and "deep" AB (AB_2) amplifiers [7]. As shown in this Figure, the class A amplifier exhibits the best AM-PM behavior (as expected), where the phase-shift is almost constant and rises when the input is at about the 1-dB compression point, P_{1dB}. Under class AB_1 operation, such a phase-drift starts at a lower power level, in the linear region of the PA. Finally, in the deeper class AB (AB_2), the AM-PM shows an almost linear phase change up to the P_{1dB}, at which it reverses rapidly. These AM-PM plots illustrate that in the more nonlinear classes of operation, the AM-PM nonlinearity behavior starts in the linear region below the 1-dB compression point.

Though the curves shown in Figure 3-5 are typical, they do vary appreciably depending on the device technology. Therefore, depending on the shape of the particular AM-PM curve, the AM-PM nonlinearity is modeled by the power series, empirical describing functions [11] and [12], or a formula involving device physics parameters [13]. However, although several representations of the AM-AM and AM-PM functions are proposed in the literature, they are generally complex in form and require the knowledge of many parameters to fit the data well. Saleh [11] and Ghorbani [14] propose compact expressions for the two functions, which despite their simplicity, the authors claim that they fit the measurement data accurately and even better than previously reported formulas that involved more parameters.

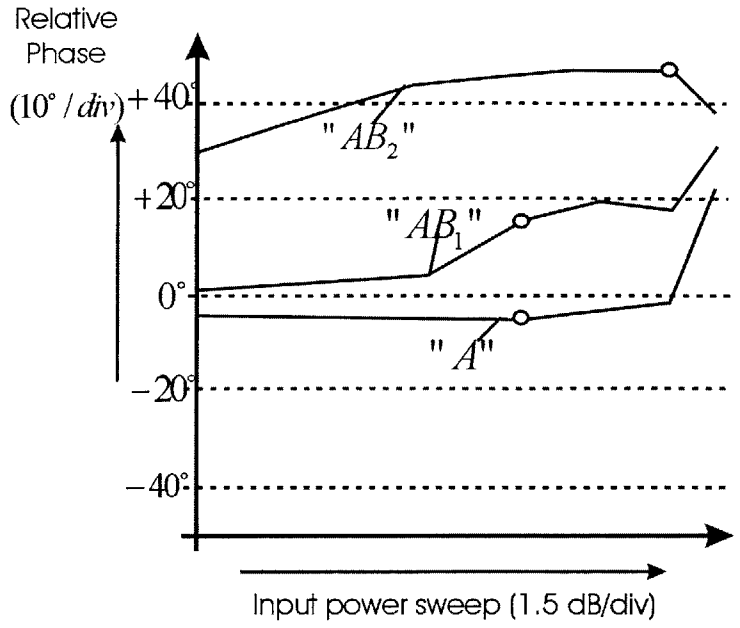

Figure 3-5: Measured AM-PM for a 1-Watt, 1.9 GHz PA. Circles indicate the 1-dB compression points (Source: Cripps [7]).

In "*Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers*", Saleh [11] presents amplitude-phase nonlinear models of TWT amplifiers that are based on simple 2-parameter formulas. Suppose that in such a model, the input and output signals of a nonlinear TWT amplifier are expressed as in Figure 3-6.

$$x(t) = r(t)\cos(\omega_c t + \phi(t)) \longrightarrow \boxed{\text{PA}} \longrightarrow y(t) = g[r(t)]\cos(\omega_c t + \phi(t) + f[r(t)])$$

Figure 3-6: Input and output signals in Saleh's PA Model.

In this model, $g[r(t)]$ - representing the AM-AM function - is an odd function of the input signal envelope, $r(t)$, with a linear leading term, and $f[r(t)]$ - representing the AM-PM function - is an even function of $r(t)$ with a quadratic leading term. Therefore, the 2-parameter formulas proposed for these two functions have the following forms,

$$g[r(t)] = \frac{\alpha_g r}{1 + \beta_g r^2} \tag{3.7}$$

$$f[r(t)] = \frac{\alpha_f r^2}{1 + \beta_f r^2} \tag{3.8}$$

As it can be seen from the above formulas, at very large input voltage levels, $r(t)$, $g[r(t)]$ is proportional to $\frac{1}{r}$ and, $f[r(t)]$ approaches a constant.

Similarly, to model the nonlinearities in SSPAs, in "*The Effect of Solid State Power*

Amplifiers (SSPAs) Nonlinearities on MPSK and M-QAM Signal Transmission," Ghorbani and Sheikhan [14] propose simple 4-parameter formulas for the AM-AM and AM-PM functions. With the input and output signals to a SSPA amplifier expressed as before (in Saleh's model), the 4-parameter formulas for the AM-AM and AM-PM functions are

$$g[r(t)] = \frac{x_1 r^{x_2}}{1 + x_3 r^{x_2}} + x_4 r \tag{3.9}$$

$$f[r(t)] = \frac{y_1 r^{y_2}}{1 + y_3 r^{y_2}} + y_4 r \tag{3.10}$$

Next, the optimal values for the parameters in each of the above formulas are obtained by employing least-squares curve-fitting techniques over the input-voltage amplitude/output-voltage amplitude (AM-AM) and the input-voltage amplitude/output-voltage phase (AM-PM) measurements. In both papers, the resulting optimal formulas provided an excellent fit to the measured data.

3.2 Nonlinearity Effects of PA

To examine the nonlinearity effects of a PA on a modulated signal, let us simplify the model for PA to one exhibiting third-order nonlinearities only, neglecting the higher orders. Also, although it is possible to obtain an analytical result that includes the higher-order nonlinearities, the result may not be convenient to use since these higher-order levels are typically not given. In fact, it has been shown that such a simplification approximates the nonlinearity reasonably well in most cases, especially at the operating points below the 1-dB compression point. Hence, when $x(t)$ is the input voltage to the PA, the output voltage is modeled by

$$y(t) \approx \alpha_0 + \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \tag{3.11}$$

Traditionally, to analyze the nonlinear effects in a PA, a single sinusoid (single-tone test) and two sinusoids (two-tone test) are applied at the input of the PA and the resulting components in the output are examined in each case. The single- and two-tone tests are introduced next.

3.2.1 Single-Tone Test

Suppose that the modulated input signal to the PA is characterized as

$$x(t) = A(t)\cos(\omega_c t + \phi(t)) \tag{3.12}$$

where ω_c is the carrier frequency, $A(t)$ is the modulated amplitude, and $\phi(t)$ is the modulated phase of the carrier. Next, we evaluate the response of the PA to two types of modulated input signals: one with a constant envelope and another with a variable envelope.

Input Signal with Constant Envelope

Suppose that the envelope $A(t)$ is a constant, A_c. The third-order nonlinearity terms in the output can be determined as follows,

$$
\begin{aligned}
y(t) &= \alpha_3 x^3(t) + \dots \\
&= \alpha_3 A_c^3 \cos^3(\omega_c t + \phi(t)) + \dots \\
&= \frac{3\alpha_3 A_c^3}{4} \cos(\omega_c t + \phi(t)) + \frac{\alpha_3 A_c^3}{4} \cos(3\omega_c t + 3\phi(t)) + \dots
\end{aligned} \tag{3.13}
$$

The second term in (3.13) is the modulated signal around $\omega = 3\omega_c$. However, since the bandwidth of the original signal, $A_c \cos(\omega_c t + \phi(t))$, is much less than ω_c, the bandwidth of $\cos(3\omega_c t + 3\phi(t))$ is also quite small. Hence, the signal spectrum around $3\omega_c$ does not interfere with the signal spectrum around the fundamental frequency, ω_c, and as a result, the shape of the spectrum in the vicinity of ω_c remains unchanged. Consequently, the PA nonlinearity does not distort a signal with a constant envelope.

Input Signal with Varying Envelope

On the other hand, when the input signal envelope, $A(t)$, is variable, the input can be represented by the baseband in-phase, $x_I(t)$, and quadrature-phase, $x_Q(t)$, components as below,

$$
\begin{aligned}
x(t) &= A(t) \cos(\omega_c t + \phi(t)) + \dots \\
&= A(t) \cos(\phi(t)) \cos(\omega_c t) - A(t) \sin(\phi(t)) \sin(\omega_c t) + \dots \\
&= x_I(t) \cos(\omega_c t) - x_Q(t) \sin(\omega_c t) + \dots
\end{aligned} \tag{3.14}
$$

The third-order nonlinearity terms in the output are then given by,

$$
\begin{aligned}
y(t) &= \alpha_3 [x_I(t) \cos(\omega_c t) - x_Q(t) \sin(\omega_c t)]^3 + \dots \\
&= \alpha_3 x_I^3(t) \cos^3(\omega_c t) - \alpha_3 x_Q^3(t) \sin^3(\omega_c t) + \dots \\
&= \alpha_3 x_I^3(t) \frac{3\cos(\omega_c t) + \cos(3\omega_c t)}{4} - \alpha_3 x_Q^3(t) \frac{3\sin(\omega_c t) - \sin(3\omega_c t)}{4} + \dots
\end{aligned} \tag{3.15}
$$

From (3.15), it is apparent that the output contains the spectra of $x_I^3(t)$ and $x_Q^3(t)$ centered around the fundamental frequency. However, the two signals, $x_I^3(t)$ and $x_Q^3(t)$, have broader spectra than $x_I(t)$ and $x_Q(t)$ since the fourier transform of $x_I^3(t)$ ($F[x_I^3(t)]$) is the convolution of the fourier transform of $x_I(t)$ ($F[x_I(t)] = X_I(f)$) three times, i.e. $F[x_I^3(t)] = X_I(f) \otimes X_I(f) \otimes X_I(f)$.

Hence, when the input signal with a time-varying envelope is band-limited and the PA exhibits significant nonlinearity, the shapes of the spectra of $x_I(t)$ and $x_Q(t)$ in the vicinity of ω_c are not preserved, and furthermore, the spectra are not limited to the original bandwidths. Consequently, the effect of the PA nonlinearity in changing the shape of signal spectrum in the band of interest results in Bit-Error-Rate (BER) degradation at the detector. Moreover, the spectrum broadening (known as spectral regrowth or out-of-band emission) effect of the PA nonlinearity reduces bandwidth efficiency and introduces

Adjacent Channel Interference (ACI) that is commonly measured by the Adjacent Channel Power Ratio (ACPR). In this thesis, the PA nonlinearity effect on the BER degradation is investigated.

From the above discussions, it can be inferred that a PSK- or FSK-modulated signal (which has a constant envelope) has no linearity requirement, and thus highly nonlinear, power-efficient PAs can be used without degrading the performance. On the other hand, for an envelope-varying modulated signal (such as an ASK or M-QAM), operation in the PA's nonlinear region introduces distortion.

3.2.2 Two-Tone Test

In addition to the aforementioned nonlinearity effect on a single-tone signal, when multiple carriers at different frequencies are applied at the input of a nonlinear PA, energy is transmitted to frequencies that are sums and differences of the original frequencies. These frequencies are not harmonics of the input frequencies. The signals at such frequencies are called *intermodulation (IM) products* which in practice severely limit the performance of power amplifiers. Let us examine the IM products that arise when two equal-amplitude sinusoids of different frequencies are applied to a PA. When a two-tone sinusoidal signal,

$$x(t) = A\cos(\omega_1 t) + A\cos(\omega_2 t) \tag{3.16}$$

is applied to the input of a nonlinear PA, spectral components (as illustrated in Fig. 3-7) are generated by up to third-order nonlinearities in four frequency bands, namely the dc, fundamental, second- and third-harmonic bands.

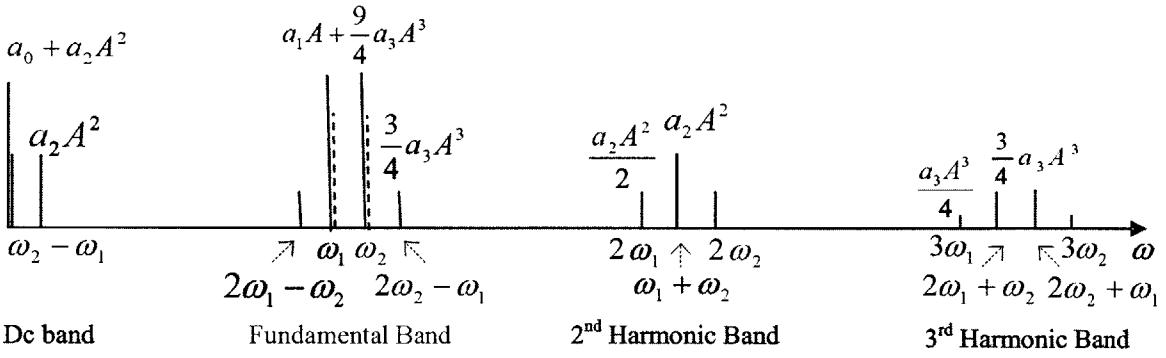

Figure 3-7: Effect of the PA nonlinearity: harmonics and intermodulation distortions.

In other words, the fundamental signals at ω_1 and ω_2 are

$$(\alpha_1 A + \frac{9}{4}\alpha_3 A^3)(\cos(\omega_1 t) + \cos(\omega_2 t)) \tag{3.17}$$

The second-harmonic components at $2\omega_1$ and $2\omega_2$, and the third-harmonic components at $3\omega_1$ and $3\omega_2$ are

$$\frac{\alpha_2 A^2}{2}(\cos(2\omega_1 t) + \cos(2\omega_2 t)) + \frac{\alpha_3 A^3}{4}(\cos(3\omega_1 t) + \cos(3\omega_2 t)) \tag{3.18}$$

Finally, the IM products are,

$$\alpha_2 A^2 \cos(\omega_1 \pm \omega_2) + \frac{3\alpha_3 A^3}{4}(\cos(2\omega_1 \pm \omega_2)t + \cos(2\omega_2 \pm \omega_1)t) \tag{3.19}$$

As Fig. 3-7 indicates, while some of the unwanted spectral components, such as the harmonics and dc components, can be easily filtered out, the third-order IM components, which are close to the fundamental signals, cannot be filtered out. The adverse effect of this behavior becomes clear when we examine a situation where a weak signal accompanied by two strong interferers is corrupted by the intermodulation between the two interferers. As depicted in Fig. 3-8, the presence of the interferers' higher-power third-order IM component in the channel of interest severely degrades the performance of the desired signal.

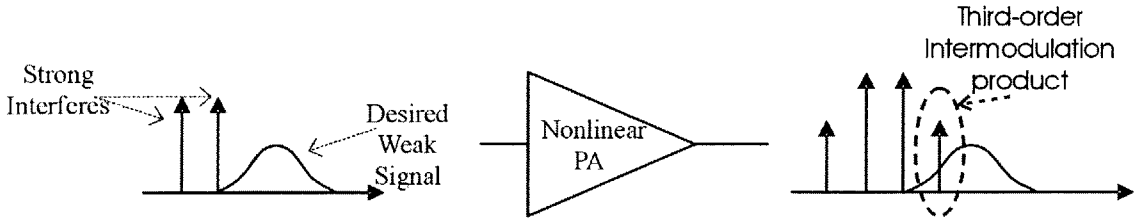

Figure 3-8: Effect of the PA nonlinearity: corruption of a signal due to the intermodulation between two interferers (Source: Razavi [2]).

In fact, corruption of signals due to the third-order IM of two nearby interferers is so common and so critical that a performance metric has been defined to characterize such a behavior. Called *third-order intercept point* IP_3, this parameter is measured by a two-tone test in which the amplitude of the two carriers, A, is chosen to be sufficiently small so that higher-order nonlinear terms are negligible. In other words, IP_3 characterizes the third-order nonlinearities only. As Fig. 3-9(a) indicates, as the input level, A, increases, the fundamental components increase proportionally with A, while the third-order IM products increase proportionally with A^3. Hence, on a logarithmic scale, the IM product levels increase at a rate three times that of the fundamentals. Fig. 3-9(b) indicates this phenomenon, and shows the IP_3 (with input coordinate IIP_3 and output coordinate OIP_3), which is defined as the point at which the fundamental line and IM product line intersect.

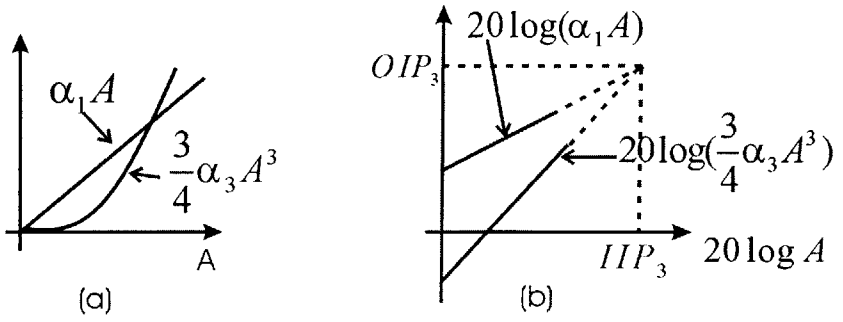

Figure 3-9: Definition of the third-order intercept point IP_3 based on the growth of output components in a two-tone test (Source: Razavi [2]).

Consequently, from the above discussions of single- and two-tone tests, two important observations are made. First, at the output of a PA modeled by third-order nonlinearities, some of the third-order components reside at the same frequencies as the linear components; these distort the linear output components through gain compression (AM-AM) and phase distortion (AM-PM). Second, the third-order components that fall outside of the linear spectrum but close to it cause spectral regrowth (or bandwidth inefficiency).

Chapter 4

Simulation

In finding the relationship between circuit nonlinearities and the system performance, a pure analytical approach is undesirable since the mathematical complexity hinders the derivation of any intuition. Intuition, however, is a powerful tool that can lead us to innovative design solutions. But how can this intuition be obtained? Generally, intuition can be derived from examination of the results that reveal the characteristics of the key elements and the causes and effects. In fact, system simulation is an effective way to obtain some results that indicate circuit nonlinearity characteristics and its effect on system performance. Hence, we devised a semi-anlytical/simulated approach for our investigation.

This chapter introduces the simulation tool, the associated building blocks, the underlying assumptions and simplifications, and the problems and challenges that arose. The underlying concepts of the building blocks of the simulation system were explained in Chapter 2, to which the readers are encouraged to refer for further understanding.

4.1 Simulation Tool

Most of the simulation is carried out in the CppSim Behavioral Simulation tool, developed by Michael H. Perrott [15]. CppSim is a simulation tool that allows the designer to quickly put together a system in a graphical manner using a library of system primitives with corresponding code descriptions in C++. In fact, the use of C++ as the code language facilitates quick computation at low-level and offers high-level structural constructs such as classes. Hence, in addition to providing the power and speed of running C++ codes, this tool offers the designers maximum freedom in developing simulation codes for their systems.

Though the need to write the simulation codes in C++ might represent a drawback to using this tool, CppSim package resolves this issue by providing two supplementary tools. First, it provides a set of C++ classes that allows a fast and convenient implementation of the system primitives. For instance, common system blocks such as filters, VCO's and signal generators are easily realized using these classes. Second, the CppSim package provides a SPICE-compatible netlist (produced by a graphical schematic capture program) to C++ conversion utility that automatically generates code from a graphical description using any standard schematic editor package, such as Cadence.

In CppSim, the blocks are described based on the input/state/output relationships, similar to the Matlab Simulink but in an object-oriented manner. In the object-oriented approach, the blocks are treated as objects that update their outputs one sample at a time based on the inputs that are specified one sample at a time. The behavioral specifications

of the block - the way the inputs influence the outputs - are set at the beginning of a simulation run. The block behavior can be a function of the state information besides the block inputs; the state information is preserved inside its corresponding block so that the simulator need not keep track of it. In summary, CppSim calculates the overall system behavior by computing the output of each block one at a time for each sample point in the simulation. This approach carries the advantage of a straightforward description of the blocks compared to Simulink, that requires a state-space description of the blocks which is usually cumbersome.

For the reasons mentioned above, using CppSim provides a fast and convenient way to simulate large systems on a behavioral level with no need for hardware implementation. In addition, if an implementation of the hardware is needed, the simulation can be integrated within all mainstream IC CAD tools that support SPICE netlisting. In this research, the objective is to characterize the power amplifier nonlinearity and its effect on the system performance by simulating a system under simplified assumptions, which is introduced next.

4.2 System Simulation Component Blocks

As described in Chapter 2, the most fundamental building blocks in every communications system are the modulator, channel, demodulator, and the detector. To this system, we add another block that is the most critical element in our research; it is one that models the behavior of a linear or nonlinear power amplifier. Consequently, we obtain the following block diagram for our system, as illustrated in Fig. 4-1. As shown in this figure, most of the simulation is performed in CppSim, and Matlab is used in extracting the signal samples from the simulated system to do additional system performance measurements.

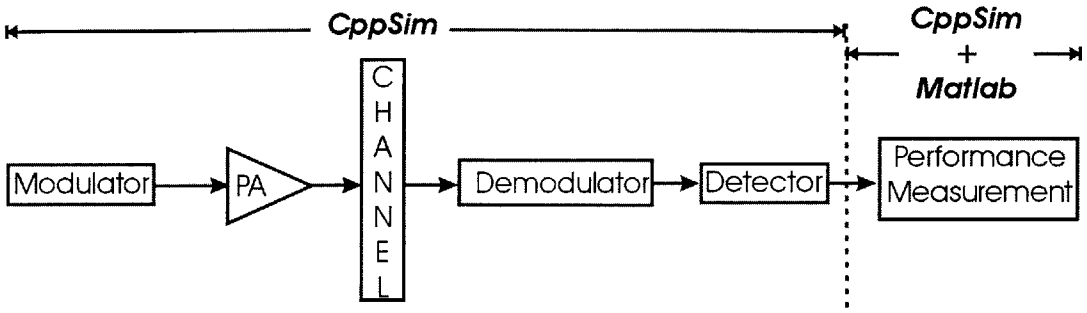

Figure 4-1: High level schematic description of the simulated system.

In this section, the characteristics of each of the blocks in Fig. 4-1 are given, including the particular choice, the reason for such selection, and implementation details.

4.2.1 Modulator

M-level Quadrature Amplitude Modulation

16-QAM or 64-QAM, Gray encoding:

As reasoned in Chapter 2, for band-limited communications systems, M-QAM techniques are more appropriate due to their higher spectral efficiency. Hence, we use a 16-QAM or 64-QAM in our simulation, which are common levels used in most wireless applications such as Wireless Giga-bit LAN (WiGLAN). In both M-QAM systems, the bit-to-symbol mapping is performed based on the Gray encoding technique, as described in Chapter 2. Fig. 2-4 (in Chapter 2) and [16] illustrate the Gray encoding used for the 16-QAM and 64-QAM constellations in our system simulation.

Rectangular constellation:

Next, what constellation type should we choose? There are many different possible constellation types that can arise from the M symbols by assuming various combinations of amplitude and phase. As indicated in Chapter 2, the error rate performance of a QAM system is determined primarily by the minimum distance between the symbols. The farther the symbols are from each other, the lower the probability error is. Thus, constellations with the same minimum distance between the symbols achieve the same performance. In addition, the transmitted power is a limiting factor in system designs, often limited by the Federal Communications Commission (FCC) regulations to a certain threshold inside and outside the band of interest; furthermore, in mobile applications, low-power signals are required. Hence, for a given minimum distance, an optimum constellation is one that has the least power. With the above discussion, to meet certain requirements on power and minimum distance, we are limited to a few constellation choices. The constellation we chose for our study is a rectangular constellation. As explained in [1], although such constellations are not the optimum for M-QAM with $M \geq 16$, they offer some advantages. First, for a given minimum distance, the power consumed by a rectangular constellation is only slightly higher than that for the best QAM. Second, these constellations are easily generated as two PAM signals imposed on phase-quadrature carriers, and third they are easily demodulated. In fact, for the above benefits, rectangular M-QAM modulations are the most frequently used in practice.

Equivalent Baseband Signal and Channel

In Chapter 2, the QAM-modulated signal waveform, as given by Equation 2.7, is transmitted as a bandpass signal obtained by an up-conversion to the carrier frequency, ω_c. In RF systems, this frequency transformation is required to transmit the signal over the bandpass channel. At the receiver antenna, the signal is down-converted back to the baseband and subsequently, demodulated. In other words, the functions of up- and down-converters on a modulated signal are primarily frequency translations. Therefore, with no loss of generality and for mathematical and simulation convenience, it is desirable to reduce all bandpass signals and channels to *equivalent low-pass (baseband) signals and channels.* Such a simplified assumption is made in our simulation, and the entire system is simulated at the baseband.

The QAM bandpass signal in (2.7) can be re-expressed as,

$$x_{tx}(t) = Re[x_l(t)e^{j\omega_c t}] \tag{4.1}$$

where the low-pass signal $x_l(t)$, representing the complex envelope of the real signal $x_{tx}(t)$, is the equivalent baseband signal. In fact, the effect of mixing the modulated signal by a carrier (to translate to a higher frequency) on the simulation results is accounted for by the PA model since the transfer characteristics of the PA (and hence, its nonlinearity) are directly affected by the frequency of operation.

In addition, we assume that our channel is band-limited, as is the case in RF applications in accordance to the FCC regulations on the particular spectrum allocated. In addition, generally in RF systems the channel bandwidth is much smaller compared to the carrier frequency - such systems are referred to as narrow-band systems. In fact, the characteristics of a PA operating over a narrow band of frequencies remain approximately constant over that range. The above reasons justify the validity of the results obtained from performing our simulation at the baseband.

Pulse-shaping Filter

Characteristics of the square-root raised cosine filter:

In the simulation, the input to the pulse-shaping filter is a train of impulses located at integer multiples of the symbol period, T_s (Equation 4.2). The amplitude of the impulses are the complex symbol sequence, I_m, produced by the M-QAM source.

$$s(t) = \sum_{m \in Z} I_m \delta(t - mT_s) \tag{4.2}$$

Suppose that $g(t)$ is the impulse response of the baseband filter. The output of the filter is then,

$$x(t) = \sum_{m \in Z} I_m g(t - mT_s) \tag{4.3}$$

This suggests that the spectral characteristics of the filter determine the spectrum of the transmitted signal. As discussed in detail in Chapter 2, a band-limited signal spectrum that has no ISI and is commonly used in practice is one with the raised cosine characteristics. Hence, to achieve the overall desired raised cosine characteristics, a square-root raised cosine filter is used at both the transmitter (as the baseband filter) and the receiver (as the matched filter), both with a roll-off factor of $\beta = 1$. The response of these two filters are depicted in Fig. 4-2. This figure indicates that the output of the second square-root raised cosine filter is equivalent to the output of a normal raised cosine filter (which is normalized to a peak value of unity). As it is explained later in this section, this value of roll-off factor is chosen to give the minimum ISI since for the purpose of this research we are less concerned with the ACI (which is highest when $\beta = 1$).

With $\beta = 1$, the raised cosine filter has the impulse response

$$h_{rc}(t) = sinc(\frac{\pi t}{T_s}) \frac{\cos(\frac{\pi \beta t}{T_s})}{1 - (\frac{2\beta t}{T_s})^2} \tag{4.4}$$

where $sinc(\frac{\pi t}{T_s}) = \frac{\sin(\frac{\pi t}{T_s})}{\frac{\pi t}{T_s}}$. In the frequency domain, this filter has the spectrum

$$H_{rc}(f) = \frac{T_s}{2}[1 + \cos(\pi T_s |f|)] = T_s \cos^2(\frac{\pi T_s |f|}{2}), \quad 0 \leq |f| \leq \frac{1}{T_s} \tag{4.5}$$

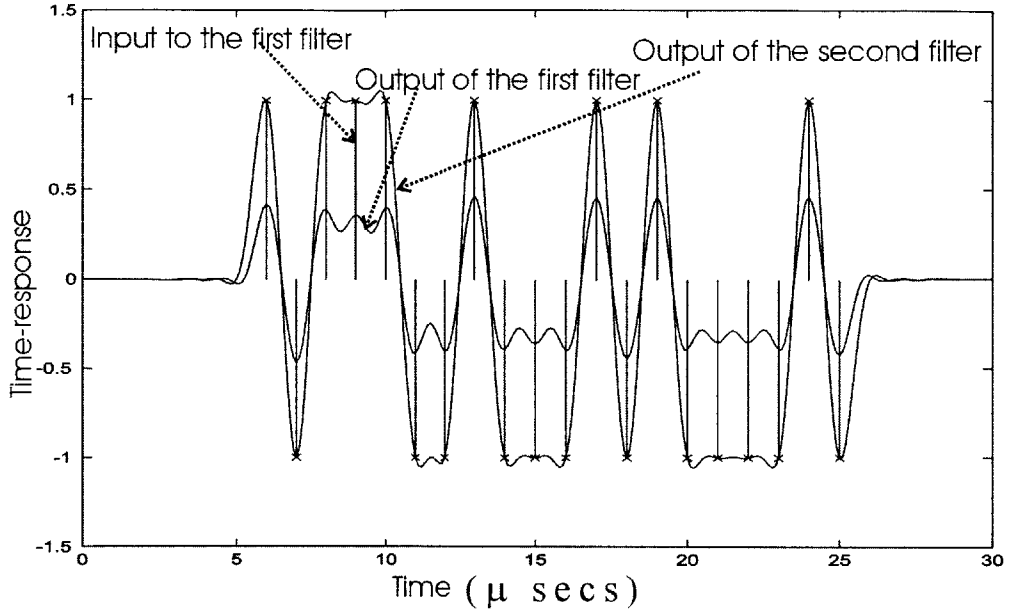

Figure 4-2: Plot showing the response of a square-root raised cosine filter (at the transmitter) to an impulse train (generated by an M-QAM source), followed by the response of a second square-root raised cosine filter (at the receiver) to the output of the first filter.

Hence, this raised cosine filter has a bandwidth of $W = \frac{1}{T_s}$.

Subsequently, the square-root raised cosine filters at the transmitter and receiver are characterized by

$$G(f) = \sqrt{|H_{rc}(f)|}e^{-j2\pi f t_0} \tag{4.6}$$

where t_0 is a nominal delay to make the filters causal - to ensure their physical realizability. From the expression in (4.6), it is clear that the square-root raised cosine filters are also band-limited to $W = \frac{1}{T_s}$.

Choosing a roll-off factor of $\beta = 1$:

At the output of the matched filter, the signal waveform is sampled at its optimum point, $t = mT_s$. Therefore, to minimize ISI, the square-root raised cosine filters should decay as rapidly as possible (in the time-domain) to minimize the contribution from preceding symbols at the sampling points. Higher β results in a higher decay rate in time-domain and hence, decreased levels of ISI. However, since rapid delay in the time-domain corresponds to a slow roll-off in the frequency-domain, the desirable decreased levels of ISI come at the cost of increased levels of ACI in the frequency-domain. On the other hand, as β goes to zero, the filter spectrum approaches a rectangular spectrum with sharp edges, or equivalently a *sinc* function in the time-domain that extends over all times. The effect of roll-off factor on the time-response of the filter to an impulse train (with random amplitudes) is depicted in Fig. 4-3. Since in this simulation we are ignoring ACI, the objective is to choose a square-root raised cosine filter that gives the minimum ISI; hence, a roll-off factor of $\beta = 1$ is chosen.

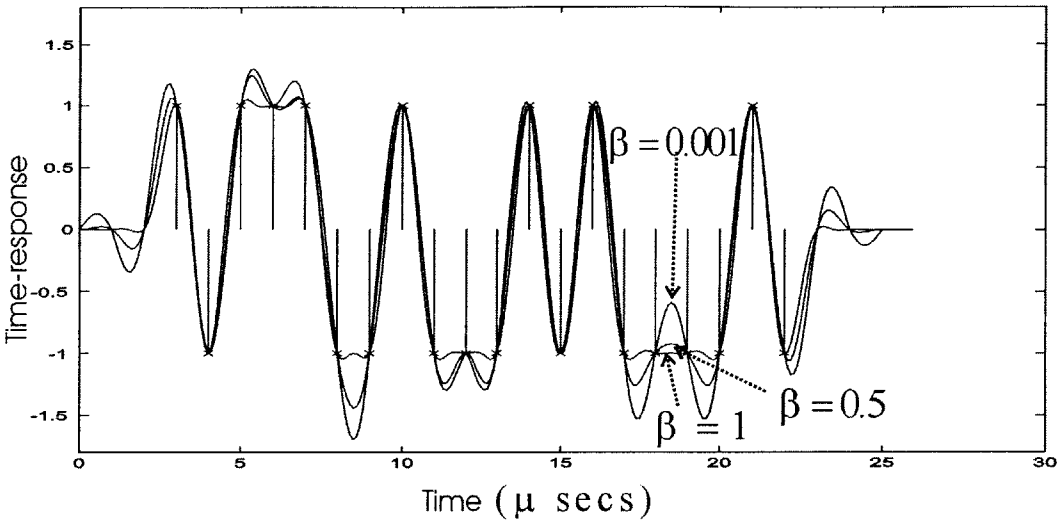

Figure 4-3: The effect of roll-off factor on the response of filter to a sequence of randomly generated data.

Implementation of the square-root raised cosine filter:

The square-root raised cosine filter in the simulation is implemented by using the built-in matlab function "rcosflt" to create the filter tap coefficients, and subsequently, load those coefficients in the "Filter" class predefined in CppSim.

More specifically, the filter coefficients, h, are created in matlab by

$$h = rcosflt(1, 1e^6, 10e^6, 'sqrt/Fs', 1). \tag{4.7}$$

According to the declaration given in the matlab 6, $h = rcosflt(x, F_d, F_s, type, \beta)$ filters the input signal x using a square-root (sqrt) raised cosine finite-length impulse response (FIR) filter. F_d and F_s, in Hz, are the sampling frequencies of the input x and output h signals respectively (F_s must be an integer multiple of F_d). Hence, this filter, first up-samples the input by zero-padding, and then filters the resulting signal. Using *type = 'sqrt'*, a square-root raised cosine filter is designed with filter coefficients normalized such that the convolution of two such filters is approximately equal to a 'normal' raised cosine filter (the approximation is due to the finite length of the filters). A 'normal' raised cosine filter is one with filter coefficients normalized to give a peak coefficient of one. In addition, *type = 'Fs'* is used when the input signal need not be up-sampled. In this case, the input has a sampling frequency of F_s with $\frac{F_s}{F_d}$ samples per symbol. Using *type = 'Fs'*, the number of tap coefficients (length of the impulse response) of the filter is

$$length(h) = 1 + \frac{F_s}{F_d} * 2 * delay \tag{4.8}$$

where *delay = 3* by default, when it is not specified. Finally, $\beta \in [0,\ 1]$ is the roll-off factor of the filter. Hence, by feeding an impulse ($x = 1$) into the filter, as in (4.7), we obtain the filter's tap coefficients in h. The tap coefficients are then saved in a file "rcosflt.dat" to be used by the "Filter" class in CppSim.

More System Parameters at the Transmitter

For the chosen simulation symbol rate of $\frac{1}{T_s} = 1e^6$ $(symbols/sec.)$, the baseband filter and hence the transmitted signal waveform are band-limited to $W = 1\ MHz$. With an M-QAM source ($M = 2^b$, $b\ bits/symbol$), the system data rate is

$$R = \frac{b\ (bits/symbol)}{T_s\ (seconds/symbol)} = \frac{log_2(M)}{T_s}\ (bits/sec.) \tag{4.9}$$

Hence, for the 16-QAM and 64-QAM sources, the data rates are $R_{16-qam} = \frac{4}{T_s} = 4$ $(Mega\ bits/sec.)$ and $R_{64-qam} = \frac{6}{T_s} = 6$ $(Mega\ bits/sec.)$ respectively.

Summary of the Modulator

Fig. 4-4 summarizes the constituent blocks of the simulation modulator and their specifications.

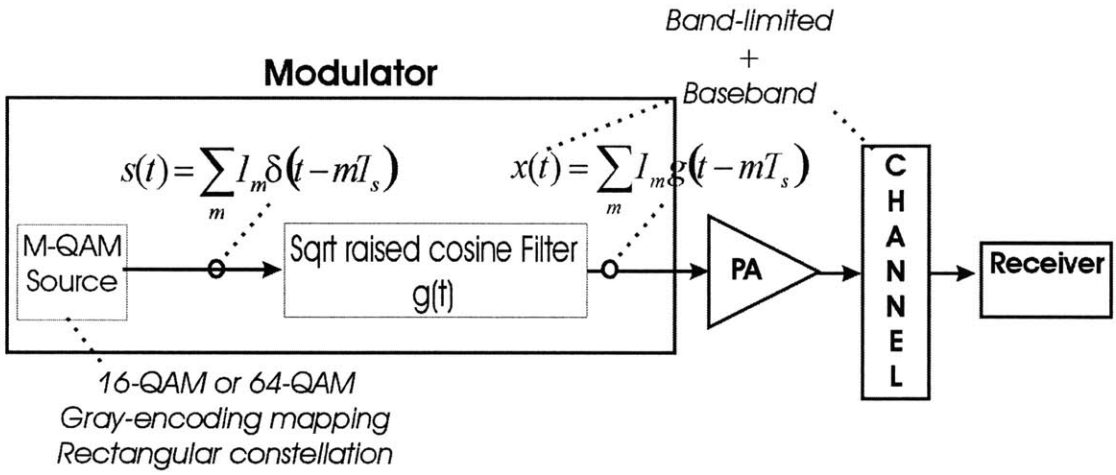

Figure 4-4: The modulator consists of an M-QAM source and a square-root raised cosine filter. The M-QAM modulation is a 16-QAM or 64-QAM with Gray encoding mapping and a rectangular constellation. The QAM source generates symbols at a rate of $\frac{1}{T_s} = 1e^6$ $(symbols/sec.)$. The baseband filter is a sqrt raised cosine filter with a roll-off factor $\beta = 1$ and bandwidth $W = 1\ MHz$. Both signal and channel are baseband and band-limited to $W = 1\ MHz$.

4.2.2 PA

Next, the band-limited signal waveform at the output of the square-root raised cosine filter goes through the power amplifier. As explained in chapter 5, to work with realistic results, the PA's input-output characteristic is modeled by employing curve-fitting techniques to the measured data of an experimented PA designed by Anh Pham. Chapter 5 presents the specifications of this PA, the behavioral model of the simulation PA and its implementation.

4.2.3 Channel

To study the effect of circuit nonlinearity on the system performance, we start with an ideal, band-limited channel that is modeled with a linear filter, $C(f)$, and an additive white

noise. As explained in Chapter 2, the filter of an ideal channel is characterized by: $C(f) = 1$ over the channel bandwidth W, and $C(f) = 0$ outside the channel spectrum. In addition, without loss of generality and for simplicity, we choose the channel bandwidth to be equal to that of the transmitted signal waveform, i.e. $W = \frac{1}{T_s}$.

With the choice of an ideal linear filter for the channel together with the transmitted signal band-limited to the channel bandwidth, the signal at the output of the channel filter is identical to that of its input. Equivalently, we can ignore the filter in the channel model and simplify the model to an additive white noise channel, in which the transmitted signal $x(t)$ is corrupted by an additive random noise process [1]. Razavi [2] defines noise as "any random interference unrelated to the signal of interest." With such a definition, deterministic distortions such as the harmonics and intermodulation products generated by the PA nonlinearity are not classified as noise. Physically, the additive noise process rises from electronic components and amplifiers in the receiver and from interference in the transmission of the signal (as in wireless systems). Called *thermal noise*, such a noise is statistically characterized as a Gaussian process. The resulting channel is called the *additive white Gaussian noise (AWGN) channel.* In fact, because of the simplicity, mathematical tractability, and applicability to a broad class of physical communication channels, such a model is frequently used in communications system analysis and design. For above reasons, the AWGN channel is the channel of our choice in the simulation system.

4.2.4 Demodulator

To implement a demodulator that is optimum, in the sense that it gives the maximum signal-to-noise ratio (SNR) at its output, it must consist of a matched filter and an optimum sampler.

Matched Filter

A matched filter is a filter with an impulse response matched to the transmitted signal waveform (as defined in Chapter 2). Suppose that the transmitted signals are $\{x_m(t) = I_m g(t - mT_s)\}$, where $\{I_m\}$ is a sequence of complex symbols generated by the M-QAM source, and $g(t)$ is the impulse response of the square-root raised cosine filter at the transmitter. Hence, with the above definition of matched filter, we have

$$h(t) = \begin{cases} C \cdot x_m(T_s - t), & for\ (m-1)T_s \leq t \leq mT_s \\ 0 & otherwise \end{cases} \tag{4.10}$$

With a change of variables, $t - (m-1)T_s \rightarrow t$, we obtain

$$h(t) = C \cdot x_m(mT_s - t), \quad for\ 0 \leq t \leq T_s. \tag{4.11}$$

Thus,

$$h(t) = C \cdot x_m(mT_s - t) = C \cdot I_m g(mT_s - t - mT_s) = C' \cdot g(-t) \tag{4.12}$$

Where C is an arbitrary constant, and $C' = C \cdot I_m$, the value of which does not affect the output SNR since this multiplicative factor drops out from the numerator and denominator of the SNR (as indicated by (2.20) in Chapter 2). Thus, we can ignore the constant term and express the matched filter as $h(t) = g(-t)$. In addition, since $g(t)$ is an even function $g(t) = g(-t)$, we obtain $h(t) = g(t)$. In other words, we choose a square-root raised cosine filter for both the baseband filter at the transmitter and the matched filter at the receiver.

Sampler

The signal waveforms at the output of matched filter are, subsequently, sampled by the *sampler* block at their optimum points to give a sequence of discrete symbols that are corrupted versions of the QAM-source generated symbols. There exist two convenient methods to find the optimum sampling point.

One way is to correlate the output of the matched filter, $y[n]$, with various delayed versions of the transmitted signal (at the output of the baseband filter), $x[n-d]$, and choose the delay d that gives the maximum correlation. Correlation is performed by sample-by-sample multiplication of the two signals and adding the resulting terms. The optimum sampling point is then $t = d + 1$.

Another method is to look at the *eye-diagram* of the signal at the output of the matched filter and find the optimum sampling instant from that. In fact, an eye-diagram (resembling the human eye) is commonly used to display the qualitative effects of ISI. It is obtained by observing the received signal through an oscilloscope, with the time axis normalized at the symbol rate, $1/T_s$. In the system simulation, the eye-diagram is plotted by using CppSim's matlab function "eyediagram" (or similarly matlab6's "eyediagram" function). It is constructed by slicing the demodulated signal (in the absence of noise) in T_s-second segments, and superimposing the slices in the interval (0, T_s) [3]. The ISI reduces the eye-opening and distorts the zero-crossings, therefore leaving less margins for the noise to cause errors [1]. As a result, the point with maximum eye-opening - i.e. with minimum ISI - is the optimal sampling point which provides the most protection against noise.

4.2.5 Detector

To study the effect of the PA nonlinearity on the system performance, the detector is chosen as one that is optimal for a rectangular M-QAM constellation in a linear system (with a linear PA). Such an optimum detector is based on the minimum-distance (MD) detection rule since under the generation of equiprobable symbols (by the QAM source) going through an AWGN channel, the MD rule is equivalent to the minimum-probability-of-error (MPE) rule (as proved in Chapter 2). In other words, the detector makes decisions on the received symbols such that the resulting probability error in detection is minimized.

4.2.6 System Performance Measurement - BER

As indicated in Figure 4-5, the complex symbols generated by the M-QAM source are decomposed into their real and imaginary components (which are independent random variables with identical statistics). The resulting two real symbol sequences pass through identical simulation blocks (such as a *real-baseband filter* and an *imag-baseband filter*). They go through the same PA to obtain the PA output amplitude and phase. The PA complex output is again decomposed into its real and imaginary components, passing through identical (but different) simulation blocks until the two real signals join at the *system performance measurement* block.

To measure the error probability of the simulated system (with linear or nonlinear PA), the *real-compare* and *imag-compare* blocks compare the signal waveforms at the output of the *real-demodulator* and *imag-demodulator* with the corresponding real and imaginary waveforms at the input of the transmitter baseband filters. However, due to the delay that exists from the transmitter input to the receiver output (exerted by the two filters), the

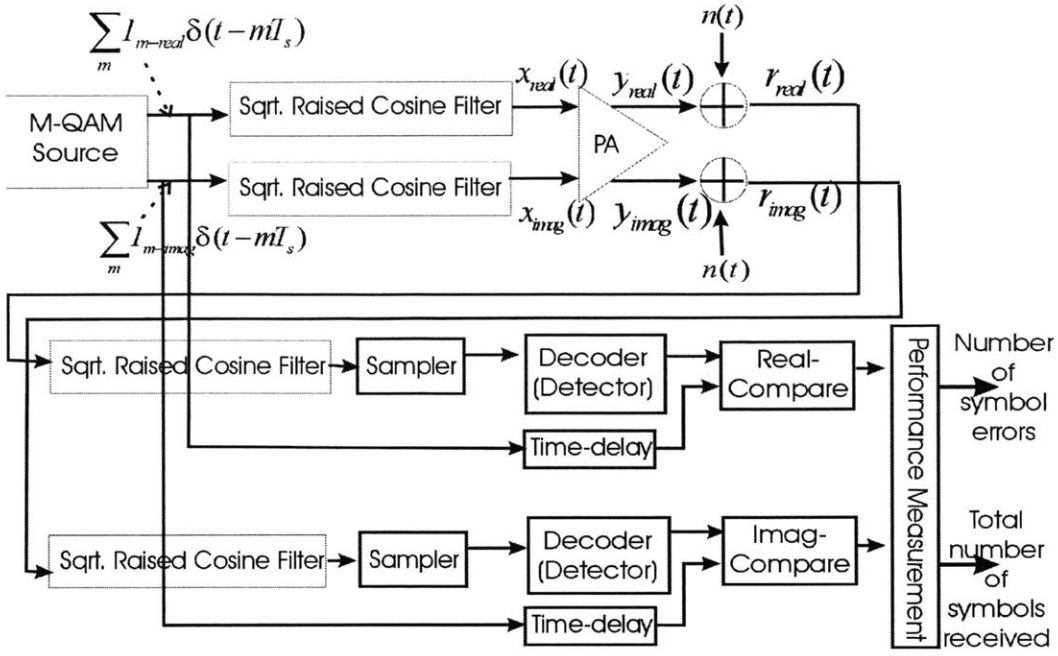

Figure 4-5: Block components of the simulation system.

signal at the transmitter must be lined up with the signal at the receiver to perform the comparison. Hence, the *time-delay* block shifts the signal at transmitter by the correct delay (obtained by correlation or eye-diagram method), which is 60 samples, contributed by the two filters each with 30 tap coefficients. The *compare* block outputs a "0" if the two samples are equal, and outputs a "1" otherwise.

The resulting real and imaginary waveforms then go through the *SER*, which outputs the number of error symbols and the total number of symbols received, which are loaded into matlab. A matlab script then calculates the BERs and plots the BER versus SNR curve. In addition, to compare the performance of the nonlinear PA to that of a linear PA, the matlab script evaluates the performance error of an ideal AWGN channel (with the white Gaussian noise having zero mean and variance σ^2) using the exact analytical expressions for the M-QAM ((2.53) given in Chapter 2) and plots the corresponding curve in the same BER versus SNR figure. Indeed, we have verified that the performance (BER) of an ideal system with no PA but with an AWGN channel (from theoretical expression) matches very closely the simulated performance (BER) of a system with a linear PA and an AWGN channel. Chapter 6 provides the simulated results followed by the PA nonlinearity/BER relationship they indicate.

4.3 Generality of the System

The level of generality of this system is as follows. The blocks that are modulation-type specific are the *M-QAM source* and the *M-QAM MD decoder*. For the *QAM source*, we can choose the level, M (number of symbols in the rectangular constellation), and the minimum-distance d_{min}, between the symbols. In addition, the *PA* block model is based on an experimented nonlinear PA, as specified in Chapter 5. The remaining blocks are general. Besides d_{min}, there are three other parameters that are specified for a particular simulation run: symbol period (T_s), gain of the corresponding linear PA, noise variance (σ^2), and the number of simulation runs.

To use this simulation tool in the course of a PA pre-design stage, a PA designer can modify the system simulation by

- replacing the QAM source and detector (based on specific system rate, symbol period T_s (hence signal bandwidth $1/T_s$), and d_{min} between symbols)
- changing the roll-off factor β of the square-root raised cosine filter blocks
- changing the filters' number of tap coefficients which have a direct influence on the transmission delay from input signal to the output signal
- changing the noise variance, σ^2, of the AWGN channel
- modeling the PA's approximate nonlinearity

In other words, the above parameters can be thought of as "knobs" that one can use to tailor this system simulation to one of interest.

4.4 Issues and Challenges in the Simulation

In creation of the system simulation described above, we faced many subtle issues, the solutions to which revealed more in-depth understanding of the system parameters in the simulation. The goal of this section is to present the questions that arose and their proposed solutions.

4.4.1 Real or Complex Signals

Since the QAM symbols are complex, should we work with complex signals throughout the simulation?

The answer is *No* simply because CppSim does not support complex data types. Hence, instead we work with the real and imaginary components of the complex signals. Fortunately, this does not present a disadvantage since in practice, the input to the PA is also real - for instance a carrier whose amplitude and phase are determined by the amplitude and phase of complex QAM symbols. The PA, then uses both the real and imaginary components of the input signal (or equivalently its magnitude and phase) in its behavioral model to find the real and imaginary components of the output signal. Hence, all system blocks are duplicated for the parallelized real and imaginary paths, except for the *PA* block and the last block, *performance measurement*, that computes the number of symbols in error.

4.4.2 Performance of an Ideal System

What is the ideal system to which we compare the performance under the influence of PA nonlinearity?

To understand the influence of PA nonlinearity on system performance (for instance by quantifying the additional degradation in performance due to this nonlinearity), we need an appropriate baseline to compare to. One suitable choice is the same system except with a linear PA.

The receiver is designed as one that is optimal for a system with a linear PA as follows. Since a linear PA amplifies all symbols in the QAM by the same gain and rotates all by the same phase, the decision threshold levels must be amplified by the linear gain and rotated by the constant phase accordingly. However, as explained next, since the constant phase shift can be recovered and hence removed from all symbols, without loss of generality, the linear PA is modeled with zero phase shift. A general power series model of a linear PA is

$$y(t) = a_0 + a_1 x(t) \tag{4.13}$$

where x(t) is the input voltage (with amplitude and phase) to the PA and y(t) is the output voltage. Hence, the coefficients a_0 and a_1 are complex. Equivalently, we can write (4.13) as

$$\begin{aligned} y(t) - a_0 &= a_1 x(t) \\ y(t) - a_0 &= |a_1| e^{j(\angle a_1)} x(t) \\ (y(t) - a_0) e^{-j\angle a_1)} &= |a_1| x(t) \end{aligned} \tag{4.14}$$

As shown above, since the constant offset, a_0, and the constant phase shift, $\angle a_1$, can be removed, for simplicity the linear PA is modeled as $y(t) = |a_1|x(t)$. This results in a detector with decision threshold levels multiplied by the gain of the linear PA. As a result, under an AWGN channel, we obtain the same performance as one without a PA. For this reason, we consider a system with linear PA as an ideal system with an ideal performance.

4.4.3 PA Operating Point

How should we define the PA operating point? and how can it be changed?

The input signal to the PA has the form $x(t) = \sum_{m \in Z} I_m g(t - mT)$ with random amplitude and phase according to the random complex sequence $\{I_m\}$ chosen from the M symbols by the M-QAM source. Hence, the maximum possible power of the PA input is determined by the symbols in M-QAM with maximum power (or maximum amplitude). For instance, in a 16-QAM, there are 3 distinctive amplitudes that are possessed by the symbols. Since the PA amplitude and phase nonlinearity depend on input signal amplitude, for every QAM, there are 3 possible input signal amplitudes that occur randomly. In addition, since the input is an analog waveform, its amplitude in fact sweeps all amplitudes in the vicinity of these 3 discrete amplitudes, the value of which are also affected by the filter. Since the input with peak power causes the most nonlinearity, the PA should be designed for this peak-power level to avoid distortion even if they might rarely occur [5]. However, for linearly modulated signals such as M-QAM, the peaks occur randomly. Hence, we define the operating point of the PA as the maximum amplitude of the input signal to the PA.

To obtain operating point of the PA , the simulated signal amplitudes at the input of the PA are first sampled at their optimum points. The PA operating point is the maximum of the resulting sampled amplitudes. With this procedure, we can find the operating point of the PA for a given simulated minimum-distance d_{min} between symbols in the rectangular QAM source. In fact, to get valid results, we must choose minimum-distances that result in operating the PA below its maximum nonlinearity at the saturation point.

4.4.4 SNR

Where and how should we measure the SNR? What is the PA operating point corresponding to a given SNR? How should we choose the noise variance?

In a communications system, SNR is typically defined as the average SNR at the output of the matched filter by

$$SNR = \frac{P_{signal}}{P_{noise}} \tag{4.15}$$

where P_{signal} is the average signal power at the output of the channel, $P_{noise} = N_0 W$ is the average noise power over the channel bandwidth W. In the discrete domain with E_s average energy per symbol and N_0 average noise energy per symbol, SNR is expressed as

$$SNR = \frac{E_s}{N_0} \tag{4.16}$$

The above two SNR equations are equivalent by the relation $P_{signal} = E_s W$ between the power in the analog signal waveform and the power in the corresponding discrete symbol (with symbol duration $1/W$). Consequently, (as showed in Chapter 2) for an M-QAM signal with minimum-distance, d_{min}, and average energy per symbol given by

$$E_s = \frac{d_{min}^2 (M-1)}{6} \quad , \tag{4.17}$$

the SNR, with noise variance $\sigma^2 = N_0/2$, is

$$SNR = \frac{d_{min}^2 (M-1)}{12\sigma^2} \tag{4.18}$$

In the system simulation, SNR is found by measuring the signal energy at the channel output which is equivalent to the signal energy at the PA output (since the channel is ideal). However, since the signal constellation at the PA output is distorted, E_s cannot be evaluated analytically. As stated earlier, for a given M-QAM with a specific d_{min}, the PA operating point is defined as the peak power in the constellation. However, in SNR calculation, the *average* power at the PA output (P_s) matters, which is measured by taking the average of all simulated symbol powers in the PA output constellation. Next, the average output energy is evaluated from $E_s = 2 \cdot R \cdot P_s$, where $R = 50\ ohms$ is the input impedance of the transmitter antenna.

Furthermore, as Forney [6] explains, in comparison of performance (BER) of various M-QAM modulation schemes, SNR is typically normalized for the spectral efficiency, ρ, by

$$SNR_{norm} = \frac{SNR}{2^\rho - 1} \tag{4.19}$$

where $2^{\rho} = 2^{b} = 2^{log_2 M} = M$. Hence, the plots of performance versus normal SNR, provided in Chapter 6, refer to the above normalized definition of SNR (in dB).

Finally, the noise variance, σ^2, is fixed for all simulated M-QAM modulated signals (with different d_{mins}) since in practical the systems noise variance does not change. The value of the noise variance in simulation is chosen based on the desired SNR for a given signal power.

Chapter 5

PA Nonlinearity Model

To devise a simulation tool that can be used to determine the effect of PA nonlinearity on system performance metric, BER, we need to model the input-output characteristics of the simulation *PA* block in an appropriate way. Indeed, finding a simple and intuitive relation depends critically on the parameters chosen to characterize the PA nonlinearity. As mentioned in Chapter 3, there exist several general PA nonlinearity characterizations that we can choose from, including the third-order intercept point (IP3) and 1-dB compression point, the power series model, and the AM-AM/AM-PM model. In this work, we found that the AM-AM/AM-PM model characterizes the PA nonlinearity in the best way that can be related to BER.

In this chapter, the first section justifies our choice of the AM-AM/AM-PM PA nonlinearity model for our analysis of the circuit-nonlinearity/system-performance relationship. Next, to test our theory and results with real data, we modeled the PA using measured and simulated data of an experimented class-A power amplifier, the specifications of which are outlined in Section 2. Finally, combining the results of the first two sections, the last section presents the behavioral characteristics of the simulation *PA* used for our investigation.

5.1 PA Nonlinearity Model: AM-AM/AM-PM

When a time-varying envelope signal goes through a nonlinear PA, it experiences nonlinearity distortions in both amplitude and phase. A remarkable observation is that both amplitude and phase of the PA output signal are distorted as a function of the PA input amplitude only. More specifically, as shown in Fig. 5-1, the output signal, $y(t)$, gets amplified nonlinearly as a function of the input signal amplitude, $|x(t)|$, starting with approximately linear amplification at very low input levels and becoming more nonlinear with increasing gain compression as the input level rises. Hence, the resulting nonlinear functions, $g(|x(t)|)$ and $f(|x(t)|)$, are named AM-AM and AM-PM respectively.

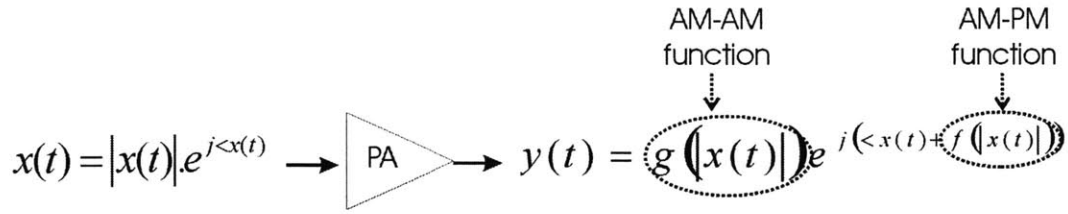

Figure 5-1: The AM-AM and AM-PM functions characterizing the PA output amplitude and phase-shift (as a function of the input amplitude) respectively.

As emphasized in Chapter 4, one of the primary goals of the system simulation is to obtain an *intuitive* understanding of the effect of circuit nonlinearity on system performance. To achieve this goal, we must characterize the PA nonlinearity such that its effect on system performance can be best visualized. As a result, we start with examining the qualitative effects of the various system simulation blocks on the signal from the input source down to the performance measurements. Indeed, as it will be shown in this chapter, such level of understanding of the signal operations would naturally lead us to the AM-AM/AM-PM nonlinearity model for the PA.

In the system simulation, the *M-QAM source* generates a sequence of symbols, $\{I_m\}$, that are complex (i.e. $I_m = I_{mr} + jI_{mi}$). Thus, they can be best visualized and distinguished when plotted in the polar coordinates. In such a coordinate system, each complex signal is uniquely specified by its amplitude, $A_m = \sqrt{I_{mr}^2 + I_{mi}^2}$, and phase, $\theta_m = \tan^{-1}(\frac{I_{mi}}{I_{mr}})$, as

$$I_m = A_m e^{j\theta_m} \tag{5.1}$$

In fact, such a method of visualization of an M-QAM signal set, so called *signal-space* or *signal constellation*, has proven valuable in understanding various effects in digital modems. For instance, the method based on which the optimum (minimum-distance MD) detector makes decisions (under an AWGN channel) can be visualized best from the signal constellation as follows. Using the MD decision rule, the detector defines decision threshold levels on the source signal constellation, which are the decision region boundaries. Hence, the decision region in which the received signal (at the output of sampler) is detected determines the estimated transmitted signal. Eventually, the system performance error probability is measured by evaluating the number of errors made in such a detection (by comparing the transmitted and estimated signals) over transmission of a large number of signals.

The above mentioned visualization method of the system performance (BER) evaluation from the signal constellation suggests that we can, similarly, examine the effect of the PA nonlinearity on system performance by comparing signal constellations at the input and output of the PA. So what PA nonlinearity model would serve our purpose best? Each signal (or symbol) in the signal constellation has both amplitude and phase, suggesting carrying of information bits by both. Hence, both amplitude and phase of the output signal are crucial in the resulting BER - for instance distortion of the signal amplitude/phase can move the transmitted signal to another decision region from where it was originally sent, resulting in a signal error. As a result, the PA nonlinearity model that separately characterizes the individual effects of nonlinearity on amplitude and phase of the output signal would capture the nonlinearity effects more accurately than a power series model.

In the power series model, the complex output signal is found from a single power series function ((3.5) in Chapter 3) of the complex input signal, from which the output amplitude and phase are extracted.

For above reasons, the AM-AM/AM-PM model characterizes the PA nonlinearity most appropriately to visualize its effect on the BER. Besides the appropriateness of the model, the AM-AM/AM-PM nonlinearity characterization has the additional benefits of being measurable, tractable by mathematics, and most conveniently, being familiar and well-understood among PA designers.

5.2 Specifications of Experimented PA

To test our simulation results with real data, we used the transfer characteristics data of an experimented class-A power amplifier - designed by Anh Pham and fabricated in IBM 6HP SiGe BiCMOS process - to specify the nonlinearity characteristics of the simulation *PA* block. The specifications of this PA is indicated in Table 5.1 [17]. In particular, such a PA, operating at 5.8 GHz frequency, has a maximum output power of 20 dBm (= 100 mW that is set by the FCC requirement), and a nominal gain of 17.5 dB. The PA input-output power characteristics and efficiency plots are shown in Figures 5-2 and 5-3 respectively. According to Pham, such a PA is suitable to use for high-speed wireless applications such as the WiGLAN operating at 5.8 GHz frequency over a 150-MHz channel, and it satisfies the ISM (International-Scientific-Medical) band regulations on the maximum output power and harmonics power levels.

Maximum Output Power	$P_{out-max}$	20	dBm
Power Dynamic Range	$\triangle P_0$	$0 \sim 20$	dBm
Center Frequency	f_0	5.8	GHz
Bandwidth	$\triangle f$	150	MHz
Gain	G	≥ 16	dB
Gain Variation	$\triangle G$	≤ 0.5	dB
Second-order Harmonic		≤ -40	dBc
Third-order Harmonic		≤ -50	dBc
Supply Voltage	V_{cc}	1.6	V
SiGe 6HP BiCMOS Process	f_T	~ 48	GHz

Table 5.1: Specifications of the experimented power amplifier used to model the simulation *PA* block (Source: Pham [17]).

Furthermore, it should be emphasized that our simulation tool can be used for any power amplifier defined with an approximate nonlinearity characterization in order to perform a similar investigation on the circuit-nonlinearity/system performance relationship.

5.3 Nonlinearity Model of Simulated PA

Since the results of this simulation demonstrating the effect of PA nonlinearity on the system performance (as presented in Chapters 6) are based on the nonlinearities of the PA specified above, it is important to know the exact nonlinearity characteristics of this PA

and the way it is implemented. Fig. 5-4 illustrates the plots of the AM-AM and AM-PM characteristics, showing the real data (circles) together with the curve of the simulation model fitted to this data. A linear least-squares curve-fitting technique, such as polynomial fitting using matlab6's "polyfit" built-in function, is used with a polynomial order of 7 for the AM-AM function and a polynomial order of 5 for the AM-PM characteristic. These polynomial orders are chosen as such because they fitted the data well without increasing the simulation time. In addition to the polynomial model $y = \sum_{i=0}^{n} a_i x^i$, matlab's library offers other curve-fitting models such as $y = ax^b$ or $y = -a + bx^c$ (where a, b, c are the parameters), and rational expressions (ratio of polynomials). As matlab mentions, the polynomial and rational models are used when a simple model is needed - one that gives flexibility with data even with a complicated structure. Indeed, since this simulation tool is to be used at the pre-design stages of the PA when the exact nonlinearities are unknown, a specific function to characterize the exact nonlinearity (such as those mentioned in Chapter 3 (3.10)) is not suitable. Hence, modeling the AM-AM/AM-PM characterizations with polynomials is general and appropriate for our purpose.

Suppose that $x(t)$ is the input signal (voltage) to the simulation *PA* block. Let $a_1, \ldots, a_7$ and $b_1, \ldots, b_5$ be the real polynomial coefficients of the AM-AM and AM-PM functions respectively, obtained from the polynomial curve-fitting method mentioned above. These modeling (nonlinearity) parameters are then loaded in the behavioral implementation of the *PA* block to find the instantaneous output signal, $y(t)$. Hence, first from the input signal amplitude, $|x(t)| = \sqrt{x(t)_{real}^2 + x(t)_{imag}^2}$, the AM-AM function, $g(.)$, and AM-PM function, $f(.)$, are found, that is

$$g(|x(t)|) = a_0 + a_1|x(t)| + a_2|x(t)|^2 + a_3|x(t)|^3 + a_4|x(t)|^4 + a_5|x(t)|^5 + a_6|x(t)|^6 + a_7|x(t)|^7 \quad (5.2)$$

$$f(|x(t)|) = b_0 + b_1|x(t)| + b_2|x(t)|^2 + b_3|x(t)|^3 + b_4|x(t)|^4 + b_5|x(t)|^5 \quad (5.3)$$

Next, using the input signal phase, $\angle x(t) = \tan^{-1}(\frac{x(t)_{imag}}{x(t)_{real}})$, together with these nonlinearity functions of the input signal amplitude, the real and imaginary components of the PA output signal (voltage) are computed by

$$y(t)_{real} = g(|x(t)|)\cos\left(\angle x(t) + f(|x(t)|)\right) \quad (5.4)$$

$$y(t)_{imag} = g(|x(t)|)\sin\left(\angle x(t) + f(|x(t)|)\right) \quad (5.5)$$

resulting in the desired output signal expressed in the familiar format,

$$y(t) = y(t)_{real} + jy(t)_{imag} = g(|x(t)|)e^{j[\angle x(t) + f(|x(t)|)]} \quad (5.6)$$

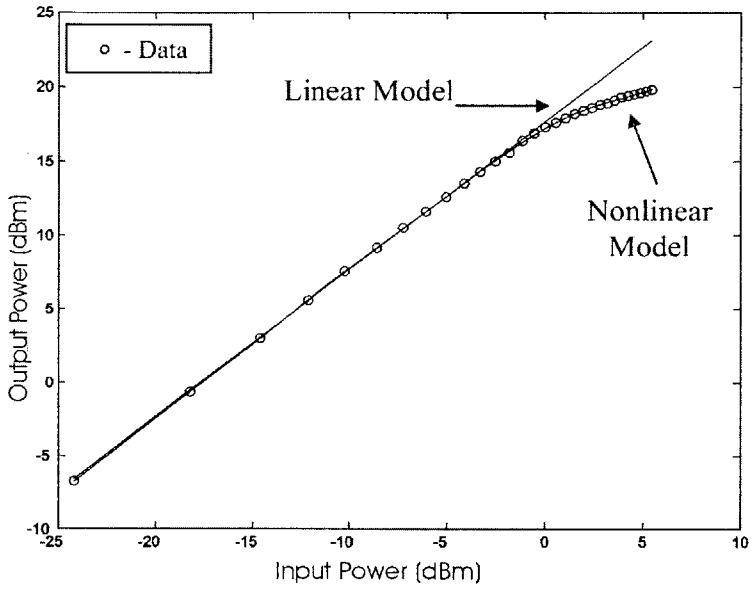

Figure 5-2: Plot of the input-output power characteristics of the experimented PA.

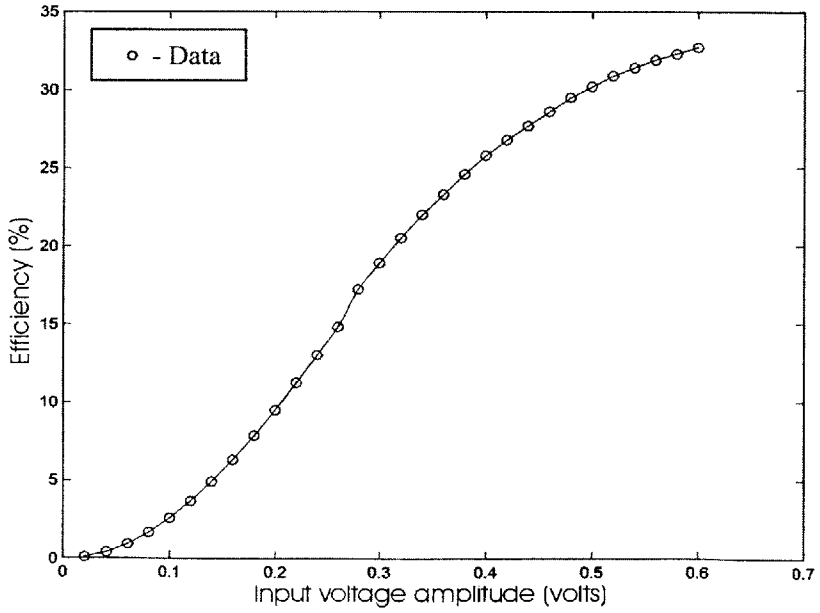

Figure 5-3: Plot of efficiency versus input voltage amplitude of the experimented PA.

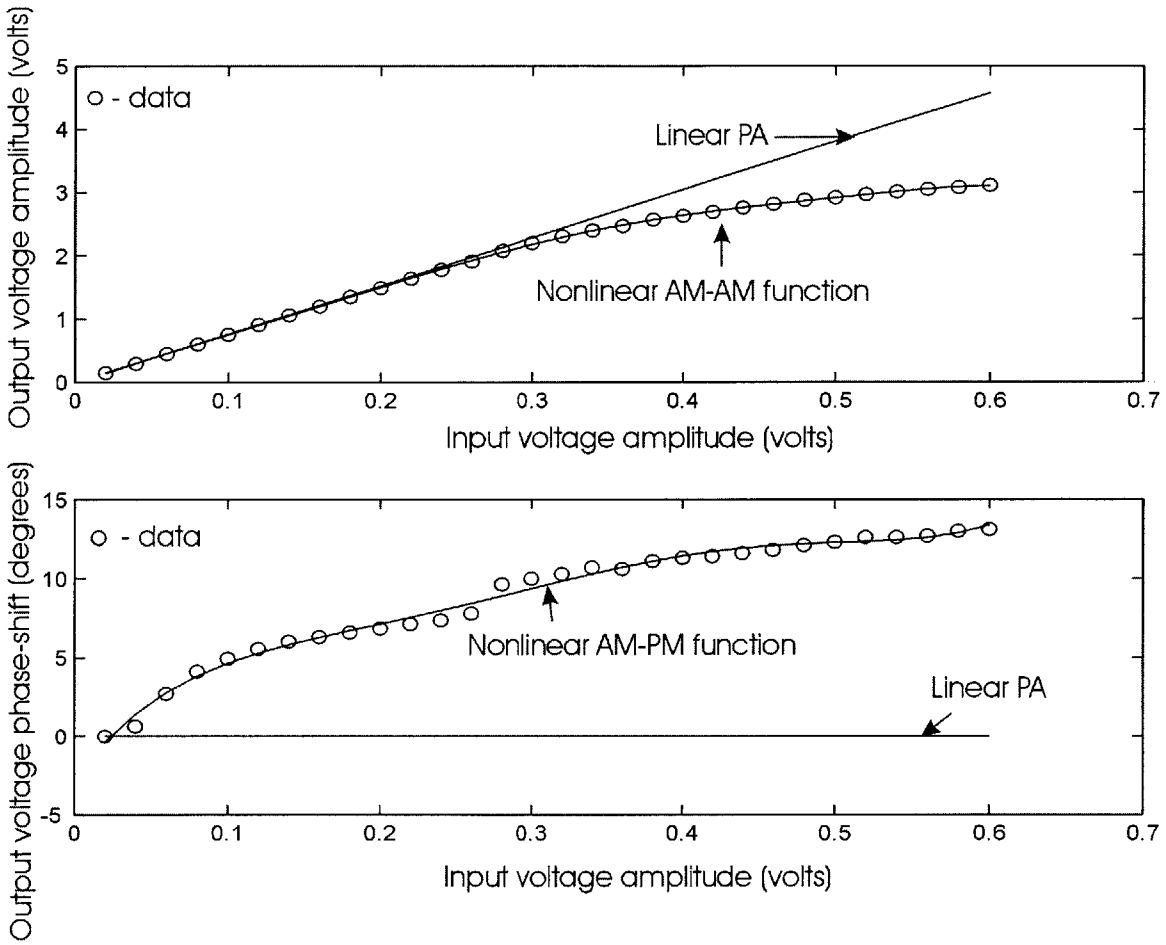

Figure 5-4: The AM-AM and AM-PM nonlinearity models of the simulation PA block.

Chapter 6

Effect of PA Nonlinearity on System Performance Metric - BER

One of the main goals of this thesis is to find a relationship between BER and appropriate PA nonlinearity parameters. In Chapter 5, we showed that the AM-AM/AM-PM model characterizes the PA nonlinearity most appropriately for our purpose. There exist a few papers that propose closed-form analytical expressions for BER in terms of the nonlinearity parameters of a particular PA for a given modulation scheme. The proposed expressions, however, are power amplifier, modulation-type specific and complex in form, providing no intuition. Therefore, to obtain a relationship between BER and the circuit nonlinearity parameters that is general, simple and intuitive, we take a semi-analytical/simulation approach. In this approach, we simulate a digital communications system in which a modulated signal goes through a nonlinear PA, an AWGN channel, and a demodulator and detector (as introduced in Chapter 4). Theoretical formulas, written in matlab scripts, evaluate the BER for the ideal (linear PA) case, and the result is compared with the simulated BER (with the nonlinear PA) to observe the additional errors that arise from distortions due to the PA nonlinearity. The combined simulation-analytical process provides an intuitive and quantitative understanding of the effect of the PA nonlinearity on BER.

To indicate the contribution of this thesis work in characterizing the effect of PA nonlinearity on BER, this chapter starts with an overview of some of the major previous works done in this subject. Sections 2 and 3, then, present the results of this thesis by first providing an intuitive understanding of the effect of PA nonlinearity on BER, and subsequently, quantifying such an effect. Finally, the chapter concludes with an attractive application that shows how the established relationship between BER and the PA nonlinearity can be used in practice.

6.1 Literature Background

As mentioned in Chapter 3, in "*The Effect of Solid State Power Amplifiers (SSPAs) Nonlinearities on MPSK and M-QAM Signal Transmission,*" Ghorbani and Sheikhan [14] propose 4-parameter formulas for the AM-AM and AM-PM functions. Using this nonlinearity model, they derive closed-form expressions for the symbol-error-probability for MPSK and 16-QAM signals going through SSPAs in the presence of AWGN and coherent receivers for terrestrial links. These expressions are obtained by closely following the procedures in [18] and [19] for MPSK and 16-QAM signals over satellite links, but applying them to terrestrial

links instead. In fact, as we desire, these error-probability expressions are in terms of the AM-AM and AM-PM functions and the noise variance. However, the complexity of such expressions prevent the extraction of any relation between circuit nonlinearity and system performance. For this reason, we avoid a pure mathematical approach to define such a relationship.

In addition, Giubilei [20] in the correspondence, "*Comments on Semi-analytic BER Evaluation by Simulation for Noisy Nonlinear Bandpass Channel,*" suggests a different semi-analytical approach to evaluate the performance of communication channels with bandpass nonlinearity. The semi-analytical approach is taken based on the assumption that performance evaluation entirely by simulation is computer-time expensive, and by all-analytical the real complexity of the channel is not realized. In the model used to characterize the PA nonlinearity, the total signal at the output of the PA is decomposed into a signal component and an equivalent noise component, which is a function of the channel noise variance and input signal envelope. In this approach, simulation is used to determine the effects of nonlinear distortion and equivalent noise statistics on the system performance. BER is, then, evaluated analytically by averaging the probability of error over the simulated signal samples received. However, once again the complex form of the expression modeling the equivalent nonlinearity makes this approach inappropriate for our investigation.

6.2 Visualization of the Effect of PA Nonlinearity on BER

As discussed in Chapter 5, the *signal constellation* of a signal waveform provides an intuitive understanding of various effects in a digital communications system. In fact, signal constellation is commonly used in understanding the effect of channel noise upon detection of the received waveform. In this thesis, we use signal constellation to visualize how the PA nonlinearity - as another source of distortion - affects the system performance metric, BER.

With the aid of the simulation tool, we extract and examine the signal constellation of an M-QAM signal in three different scenarios, each of which provides a unique insight into our understanding of this circuit-nonlinearity/system-performance relationship. First, to understand the direct effect of the PA nonlinearity on signal constellation, we look at the signal constellations at the input and output of a nonlinear PA. Second, in a communications system with a nonlinear PA but without the channel noise, we compare the signal constellation at the input of the PA with that at the output of the matched filter. Finally, we examine the latter case with the noise channel included in the system. This section presents the results of the above examinations.

Consider the rectangular signal constellation of a 16-QAM signal as shown in Fig. 6-1. We note that among the 16 different symbols in the 16-QAM signal set, there are only three different amplitudes. The 4 inner-most symbols (displayed as rectangles) at $\frac{\pi}{4}$, $\frac{3\pi}{4}$, $\frac{5\pi}{4}$, and $\frac{7\pi}{4}$ angles have the minimum amplitude. The 4 outer-most symbols (circles) at the same angles as the inner symbols have the maximum amplitude. Finally, the 8 remaining symbols (crosses) have an amplitude half-way between the minimum and maximum amplitudes. Suppose a 16-QAM signal with such a constellation is fed into a linear and nonlinear PA.

6.2.1 Effect of PA Nonlinearity Alone

In a *linear PA*, with linear AM-AM and AM-PM functions, the output amplitude is a constant times the input amplitude and the output phase is the input phase plus a constant

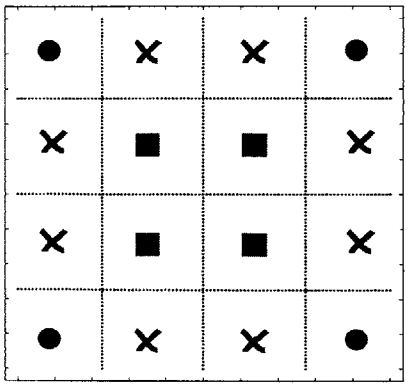

Figure 6-1: A rectangular 16-QAM signal constellation shows the three possible amplitudes: the inner 4 rectangles have the min. amplitude, the outer 4 circles have the max amplitude, and the 8 crosses have the in-between amplitude.

phase-shift. Hence, at the output of a linear PA, all symbols of various amplitudes in a signal constellation are amplified and phase-shifted by the same amount. This results in a signal constellation that is a rotated, uniform amplification version of the input PA constellation with the overall shape (i.e. the location of symbols relative to each other) preserved. In addition, to simplify examination of the signal constellations at the outputs of linear and nonlinear PAs, we ignore the constant phase-shift that is common to all symbols. This is a fair assumption since this constant phase-shift can be tracked and thus removed from the output phase. Fig. 6-2 shows signal constellations at the input and output of a linear PA.

On the other hand, in a *nonlinear PA*, the AM-AM and AM-PM functions are nonlinear, resulting in an amplitude-dependent amplification and rotation of the symbols. More specifically, symbols of equal amplitude are amplified and rotated by same amounts while symbols of different amplitudes go under different amounts of amplification and rotation based on the nonlinearity characteristics. Therefore, the net effect of the PA nonlinearity is determined by examining the individual effects of the AM-AM and AM-PM nonlinearities on the constellation shape. The reader is encouraged to refer to Fig. 5-4 of Chapter 5 to review the particular AM-AM and AM-PM nonlinearity characteristics considered here.

As shown in Fig. 6-2, under the *AM-AM nonlinearity*, symbols with low amplitude levels (that reside in the linear region of the PA) go under an amplification equal to that of the corresponding linear PA. However, symbols with amplitude levels in the nonlinear region of the PA get less amplified than the amount if they had been amplified linearly. As a result, in the output constellation of a nonlinear PA, symbols of minimum amplitude closely coincide with the corresponding symbols at the output of a linear PA. On the other hand, symbols of higher amplitudes go through a gain compression relative to their linear counterparts. The amount of this gain compression increases with increasing amplitude level. Consequently, depending on where in the nonlinear region of PA the maximum amplitude level of the input 16-QAM signal resides, the gain compression nonlinearity effect of the AM-AM function can adversely distort the constellation shape. The most distortion occurs when the maximum amplitude level resides at the maximum nonlinearity (saturation level) of the PA characteristics.

Similarly, when the PA nonlinearity is due solely to the *AM-PM nonlinearity*, symbols of minimum amplitude go under no phase shift relative to their linear counterparts (Fig.

6-2). As the amplitude level increases, the particular behavior of the AM-PM nonlinearity in this case (5-4) suggests an increasing phase-shift correspondingly. As a result, symbols of maximum amplitude get rotated the most, and all by the same amount. In concise, the AM-PM nonlinearity distorts the constellation shape by exerting different amounts of phase-shifts based on the symbols' amplitude levels.

Consequently, the combined AM-AM and AM-PM nonlinearity effects result in the net effect of the PA nonlinearity on the 16-QAM constellation as shown in Fig. 6-2. Interestingly, the two nonlinearities act independently of one another, with one distorting symbol amplitudes and the other distorting symbol phases.

6.2.2 Effect of PA Nonlinearity in a Communications System without AWGN Channel

Let us now consider the PA in a communications system with no channel noise as in Fig. 6-3. To measure the system performance under distortions due to the PA nonlinearity, we compare signal constellations at the output of the receiver sampler (after the matched filter) that has gone through linear and nonlinear PAs. To observe the effect of the nonlinearity, we examine a 16-QAM input signal with maximum amplitude at the saturation point of the PA. As the distorted output constellation in Fig. 6-3 indicates, under this high level of nonlinearity, the 4 outer-most symbols are always shifted to a neighboring decision region. This results in the misdetection of 4 out of every 16 (equiprobable) symbols, giving rise to a symbol-error-rate (or bit-error-rate) of 1/4.

In addition, we observe that at the location of the received symbol that had gone through a linear PA, a group of closely-spaced symbols instead of a single symbol are received. This phenomenon is due to the ISI that arises upon violation of the Nyquist criterion by the nonlinear PA between the transmitter pulse-shaping filter and the receiver matched-filter. As discussed in Chapter 2, the Nyquist criterion must be met to achieve zero ISI.

6.2.3 Effect of PA Nonlinearity in a Communications System with AWGN Channel

At last, we complete the communications system by adding the AWGN channel, as shown in Fig. 6-4. Once again, we compare the constellations at the output of the matched filter when they have gone through a linear and a nonlinear PA respectively.

With a *linear PA*, the effect of the AWGN channel alone on the signal constellation is observed; it distorts the location of all transmitted symbols to random locations determined by the noise's Gaussian distribution. Hence, a circular cloud around the location of each received symbol under no noise is observed, with its wideness determined by the noise variance. Therefore, depending on the noise variance and the minimum-distance between the neighboring symbols, some symbols might be misdetected for their neighboring symbols, thus, degrading the BER. Notice, however, that the overall shape of the signal constellation is still preserved.

On the other hand with a *nonlinear PA*, the received symbol clouds (due to the AWGN) are moved together to different locations. These locations are determined by the sole exercise of the PA nonlinearity on the signal constellation, as mentioned earlier.

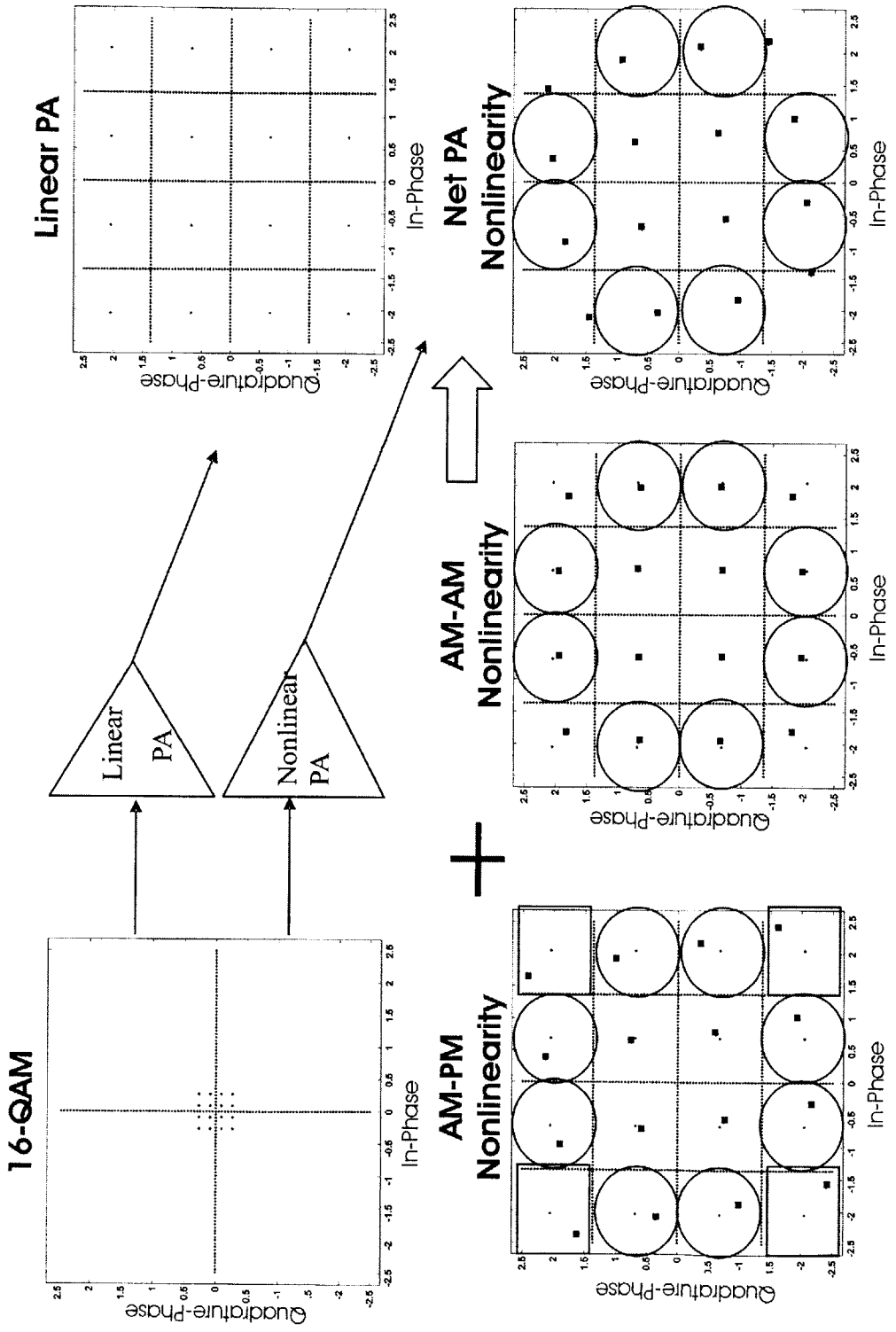

Figure 6-2: Effect of the PA nonlinearity alone on a 16-QAM constellation.

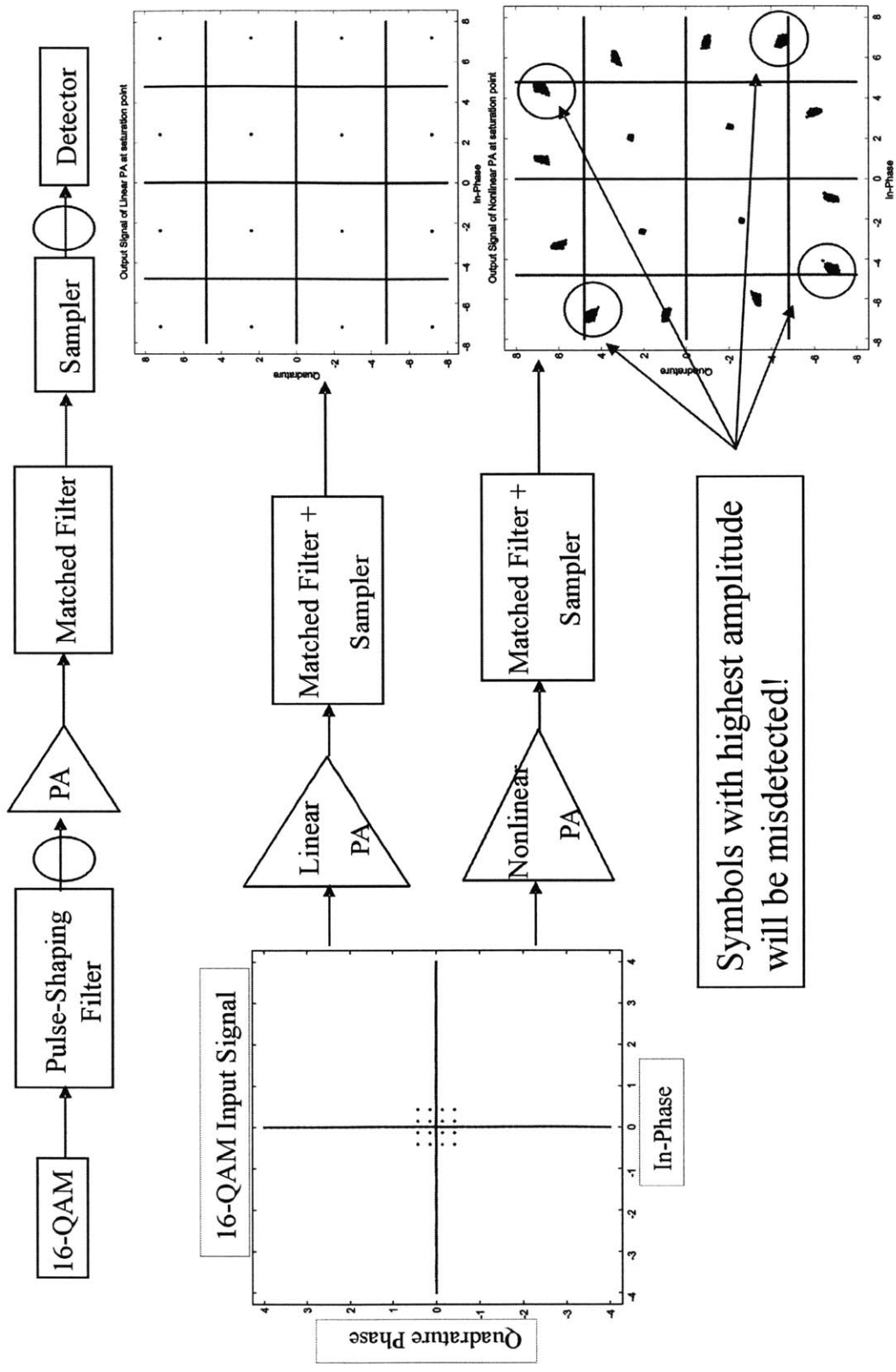

Figure 6-3: Effect of the PA nonlinearity in a communications system without channel noise on a 16-QAM constellation. The 16-QAM signal's maximum amplitude is chosen at the saturation point of the PA.

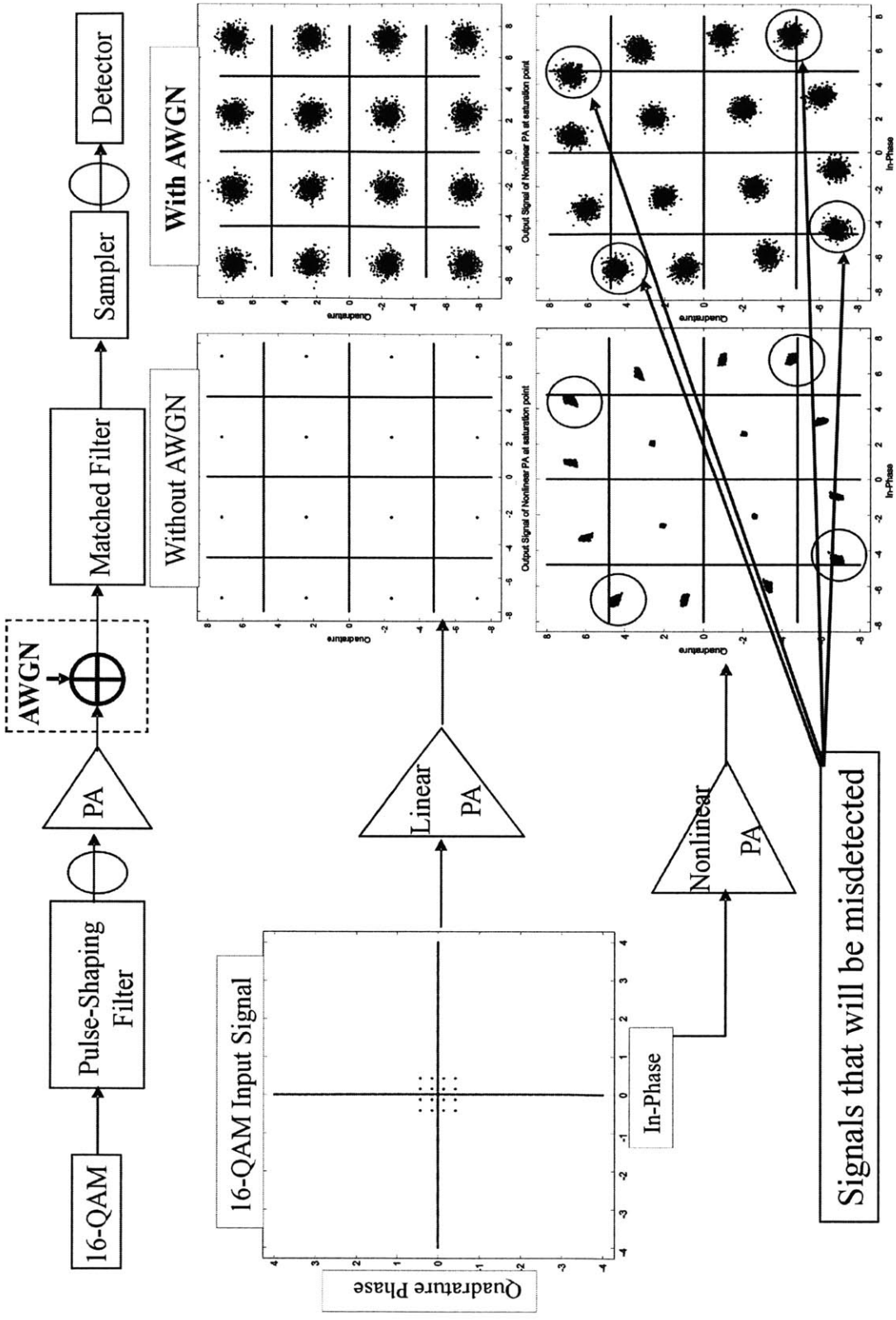

Figure 6-4: Effect of the PA nonlinearity in a communications system with an AWGN channel on a 16-QAM constellation.

6.3 Quantification of the Effect of PA Nonlinearity on BER

In the previous section, the amount of constellation shape distortion exerted by the PA nonlinearity gave us an intuitive understanding of how the PA nonlinearity degrades system performance, BER. In this section, we quantify the relationship between PA nonlinearity and BER by measuring the amount of performance degradation (through BER) at a known PA nonlinearity. Hence, we plot the symbol-error-probability (or symbol-error rate, SER) versus SNR for a fixed noise variance (of the AWGN) when a 16-QAM signal goes through a linear PA and a nonlinear PA respectively (Fig. 6-5).

6.3.1 Effect of PA Nonlinearity on BER - 16-QAM

SER $\cong$ BER:

In our system simulation, the SER and BER measurement metrics are equivalent. The reason is due to the use of Gray encoding in assigning the b information bits to the $M = 2^b$ possible symbols in the M-QAM signal set. As mentioned in Chapter 2, under such a mapping, each symbol differs in exactly one binary digit from its adjacent symbols. As a result, since in our examined signal constellations, the distortions due to the PA nonlinearity and AWGN move a symbol at worst to the first adjacent decision region, a symbol error corresponds to one bit error.

From SNR to the PA operating point:

As mentioned above, to simulate a BER versus SNR curve, the noise variance of the AWGN channel is fixed. Thus, we vary the SNR by varying the transmitted signal power through changing the maximum amplitude of the 16-QAM signal set generated at the source. The average signal power in SNR calculation is found by averaging the simulated signal power at the output of the matched filter (in the absence of channel noise).

In addition, for each SNR value - corresponding to a 16-QAM with a unique maximum-amplitude level - we can measure the maximum input signal power to the PA. Hence, we can associate each SNR value (in the BER vs. SNR plot) with the corresponding PA input power and amplitude level, which indicate the amount of PA nonlinearity on the PA power, AM-AM and AM-PM transfer characteristics. Consequently, from the translation of the SNR values to the corresponding PA input powers or amplitudes, the amount of performance degradation (BER) as a function of the PA operating point can be determined.

Characterization of the effect of PA nonlinearity on BER:

To determine the effect of the PA nonlinearity alone on the BER, we compare the simulated BER vs. SNR curves obtained when a linear PA and a nonlinear PA are employed in the system simulation under the same noise variance of the AWGN channel. From such a comparison, two remarkable observations are found (Fig. 6-5). First, the two curves coincide very closely at low SNRs, and second, their deviation from each other increases as the SNR level rises. More precisely, the curve due to the nonlinear PA always lies above that of the linear PA. At low SNR levels (where the nonlinear PA is operating in the linear region), performance degradation is due primarily to the channel noise and hence the two curves have same BER values. As the SNR level increases, the PA operating point moves to the nonlinear region, in which case both the channel noise and PA nonlinearity contribute to the BER degradation. Hence, the BER of the nonlinear PA curve will be higher than that

of the linear PA curve. Eventually, at very high SNR levels (when the channel noise slightly affects the BER), the significant amount of the PA nonlinearity becomes the dominant factor in performance degradation. This PA nonlinearity effect is clearly indicated by the large gap (as indicated by an arrow in (6-5)) between the two curves at high SNR levels.

In addition, from a different perspective, we can observe an interesting relationship between BER and the minimum-distance, d_{min}, between the adjacent symbols in a signal constellation. In Chapter 2, from the error probability expression (2.42), we found that the error rate is directly correlated to the d_{min} and the average symbol energy E_s. More specifically, if two different signal constellations have the same E_s for the same d_{min}, the error performances of the two constellations will be the same. In the presence of the PA nonlinearity, however, we observe a conflicting tradeoff between the d_{min} and BER. On one hand, a higher d_{min} results in a better BER. On the other hand, the higher d_{min} corresponds to operating the PA in a more nonlinear region, which results in a worse BER. In fact, the BER curve for a nonlinear PA (as shown in Fig. 6-5) indicates that as d_{min} increases (corresponding to higher SNR levels), the effect of d_{min} in lowering BER wins the adverse effect of PA nonlinearity in increasing BER. In other words, under the influence of the PA nonlinearity, the BER vs. SNR curve is still concave (resulting in lower BERs at higher SNRs) although the rate of the decay is significantly slower, depending on the severity of the PA nonlinearity.

Consequently, besides understanding how a given amount of AM-AM and AM-PM nonlinearity of a PA degrades the performance, we have indicated a method that quantifies the resulting performance degradation (measured by BER) as well. Hence, the above two sections indicated methods to qualitatively and quantitatively relate a given PA nonlinearity to the corresponding BER.

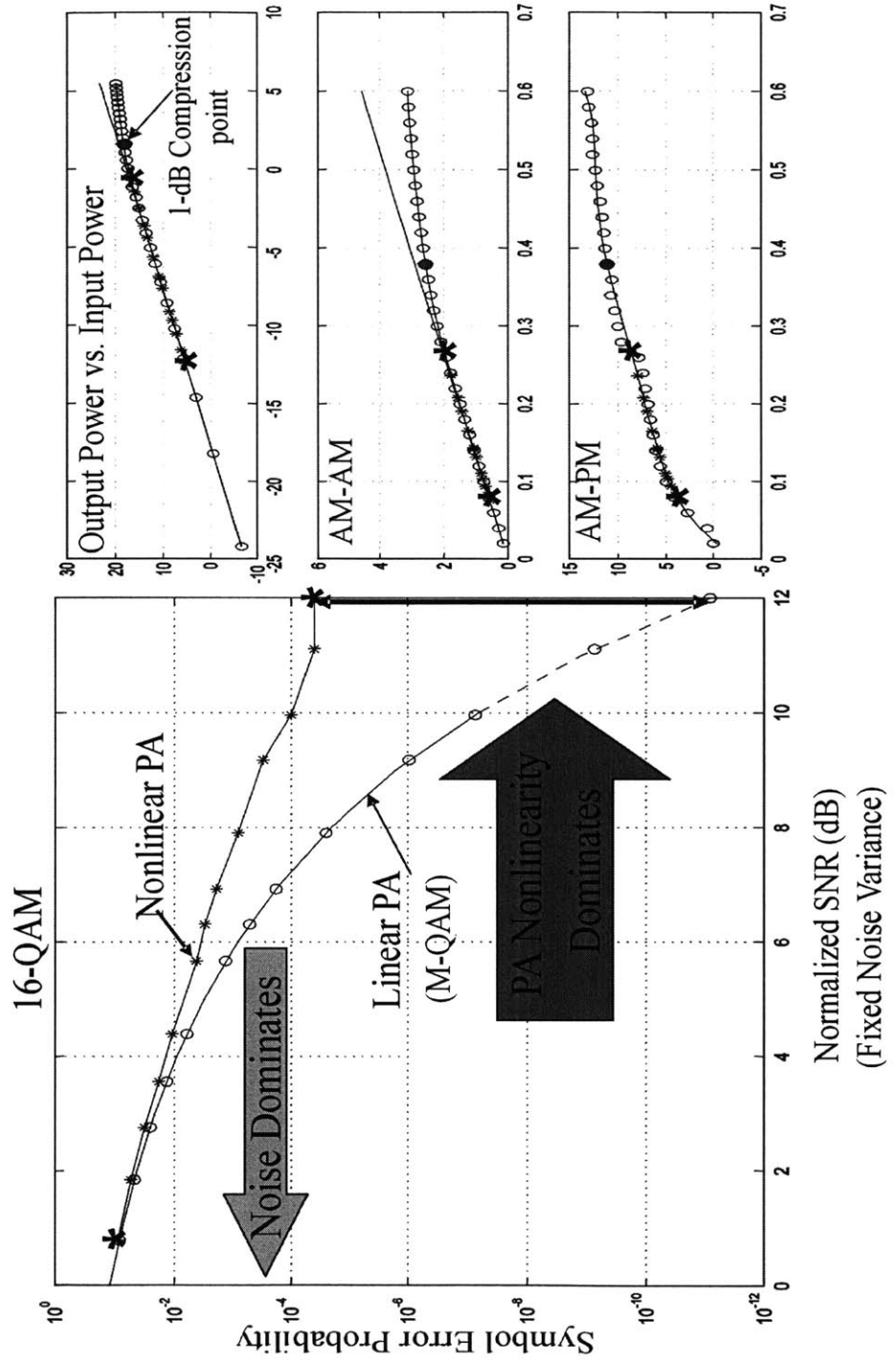

Figure 6-5: Plot of symbol-error-probability (SER) versus normalized SNR (SNR_{norm}) for a 16-QAM signal going through linear and nonlinear PAs. On the right, plots of the input-power/output-power characteristics and the AM-AM/AM-PM functions of the linear and nonlinear PAs are shown with three operating points marked: 1-dB compression point (circle), a point in the linear region (cross), and a point in the nonlinear region below the 1-dB compression point (cross). The corresponding SNRs of the latter two points are indicated with crosses on the error probability plot.

6.3.2 Effect of PA Nonlinearity on BER: 16-QAM versus 64-QAM

So far, the analysis was done by focusing on a 16-QAM modulated signal. In this section, we examine how the type of the modulation scheme might affect our results. For the purpose of such an examination, we compare 16-QAM and 64-QAM signal sets. As defined in Chapter 2, for an M-QAM signal set, where b information bits are transmitted per symbol $M = 2^b$, the spectral efficiency (ρ) is b bits per two dimensions (or b bits per symbol). Hence, a 64-QAM with spectral efficiency of $\rho_{64} = 6$ *bits/symbol* is more spectrally efficient than a 16-QAM signal set with $\rho_{16} = 4$ *bits/symbol*. This immediately reveals the advantage of using a higher level QAM in transmission of high data rates over narrow-band channels. Hence, for a fair comparison of the performance degradation of the two modulation schemes, their difference in spectral efficiency must be accounted for. Next, we discuss how this is achieved.

Performance comparison when the two modulations have different or same symbol rates:

What are the data rates - R number of transmitted bits per second - of the two modulation schemes? The data rate depends on the symbol duration, T_s seconds/symbol, and the number of bits per symbol, b, by

$$R = \frac{b}{T_s} \; bits/second \tag{6.1}$$

Hence, for the same symbol durations T_s (thus the same bandwidths $W = \frac{1}{T_s}$), a 16-QAM has a data rate of $R_{16} = \frac{4}{T_s}$ while a 64-QAM has a higher data rate ($R_{64} = \frac{6}{T_s}$) by factor of 1.5. On the other hand, for the same data rate of transmission (i.e. when $R_{64} = \frac{6}{T_{64}} = \frac{4}{T_{16}} = R_{16}$), a 64-QAM has a symbol duration 1.5 times higher than that of a 16-QAM, i.e. $T_{64} = 1.5 \cdot T_{16}$.

Subsequently, under the same data rates (thus different bandwidths), we must compare the performance of the 16-QAM and 64-QAM signals by examining their symbol-error-probabilities. In this case, since $T_{64} > T_{16}$, for the same total number of bits transmitted per unit time, less symbols are transmitted with 64-QAM. For this reason, for the same number of observed symbol errors, 64-QAM results in a higher symbol-error-probability. As illustrated in Fig. 6-6, this results in the symbol-error-probability curve of the 64-QAM to lie above that of the 16-QAM.

On the other hand, when the data rates are different (hence same bandwidths), performance comparison must be made by measuring the BERs since the two modulations give the same symbol-error-probabilities. In fact, this is the case in our simulation, in which symbol durations are assumed in the same for both modulations ($T_{16} = T_{64}$). As a result, for the same total number of symbols S transmitted per unit time, we have transmitted $4 \cdot S$ and $6 \cdot S$ total bits under the 16-QAM and 64-QAM respectively. In addition, under employment of the Gray encoding for both modulation schemes, a symbol error amounts to one bit error in both cases. This suggest that for the same number of symbol errors (hence same number of bit errors),

$$BER_{(16-QAM)} = \frac{1}{4} \cdot P_s(E) \quad while \quad BER_{(64-QAM)} = \frac{1}{6} \cdot P_s(E) \tag{6.2}$$

In other words, under transmission at different bit rates (data rates) but same symbol rates, the two modulations give rise to the same symbol-error-probability ($P_s(E)$); however, the

64-QAM achieves a lower level of BER, resulting in the $BER_{(64-QAM)}$ curve to be below the $BER_{(16-QAM)}$ curve. Fig. 6-6(b) clearly illustrates this difference.

Performance comparison under same minimum-distance or same maximum-power:

In addition, as illustrated in Fig. 6-7, two M-QAM signal sets can be compared in two different ways. One with the *same minimum-distance* d_{min} between the symbols, and another with the *same maximum-power* (or the same maximum-amplitude) in the signal set. Consider the plot of bit-error-probability (BER) vs. SNR-per-bit for the 16-QAM and 64-QAM signals (Fig. 6-7). In this case, the difference in spectral efficiency is accounted for by comparing the performance (BER) with respect to SNR per bit (defined as SNR per symbol divided by the number of bits per symbol). We observe that under the same minimum-distance, the two modulations achieve the same BER since (as proved in Chapter 2) the error probability is a function of d_{min} of the signal set. However, under such a comparison, as it is apparent from the signal constellations, a 64-QAM requires more signal power (hence more SNR) to achieve the same BER.

On the other hand, when the two modulations have equal maximum-power level, 16-QAM symbols are more distant from each other. Hence, the inverse relationship between d_{min} and BER results in a lower BER value for the 16-QAM than the 64-QAM at the same SNR value.

To find the effect of modulation scheme on performance degradation, we must operate the PA at the same nonlinearity level under both schemes, which corresponds to the same maximum power for the modulations. So, in our comparison, the 16-QAM and 64-QAM signals have the same maximum-power level.

Characterization of the effect of modulation type on BER with PA nonlinearity:

As discussed earlier in this section, since in our simulation the two modulations have the same symbol duration ($T_{16} = T_{64}$), their performances must be compared by measuring their BERs (in this case, $P_s(E) \neq BER$). Consider the plot of BER vs. normalized SNR of a 16-QAM and 64-QAM, as shown in Fig. 6-8. As mentioned before (Fig. 6-6(b)), in the ideal (linear PA) case, the higher data rate of the 64-QAM results in its BER curve to be lower than that of the 16-QAM at all SNR levels.

Next, we examine the performance when a nonlinear PA is employed. First, we note that for each modulation, their linear and nonlinear BER curves display the same behavior observed previously: that is, they are equal at low SNR levels with increasing deviation as the SNR increases. Second, an interesting observation is made when we compare such a behavior between the two QAMs. We observe that at the same SNR levels, the 64-QAM nonlinear curve deviates from its linear curve more than that of the 16-QAM. This results in the nonlinear 64-QAM curve to quickly go above the nonlinear 16-QAM curve. The reason behind this behavior is the smaller minimum-distance between symbols in the 64-QAM constellation which results in more error (hence larger gap between its linear and nonlinear BER curves) for the same amount of distortion in amplitude and phase due to the AM-AM and AM-PM PA nonlinearities at a given operating point (corresponding to the SNR level).

Figure 6-6: Comparison of error probabilities of 16-QAM and 64-QAM using performance metrics, Symbol-Error-Probability or BER.

Figure 6-7: Comparison of 16-QAM and 64-QAM constellations for same minimum-distance and same maximum-power.

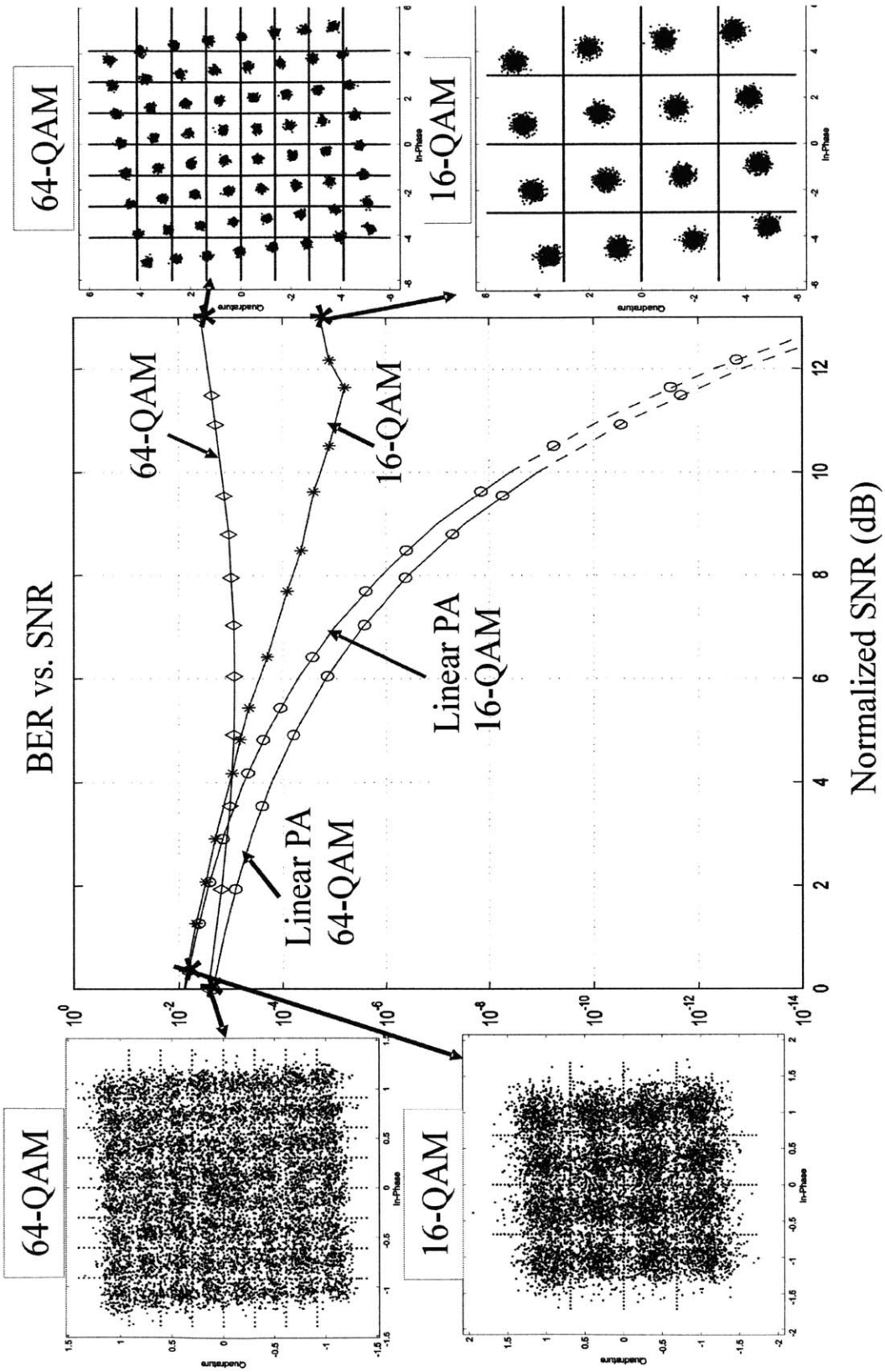

Figure 6-8: Plot of symbol-error-probability versus normalized SNR of 16-qam and 64-qam that have gone through linear and nonlinear PAs.

6.4 Application: Finding the Optimal PA Operating Point

As mentioned in Chapter 1, one of the driving motivations for quantifying the nonlinearity of a PA and its effect on BER, was to utilize such a tool to determine the nominal operating point of the PA that achieves high power-efficiency while still meeting the system performance criteria. Subsequently, such a tool helps PA designers avoid over-specifying the PA linearity and power-efficiency, which results in design of less expensive and more efficient PAs. In this section, we provide a case study demonstrating how our tool can be used to find the PA's nominal operating point.

Suppose that we are required to use a 16-QAM modulated signal, and the system requirements impose a maximum SNR of 8 dB and $BER \leq 10^{-3}$. We use the simulation tool described in this thesis with an approximate characterization of the PA nonlinearity to simulate the system under investigation. Next, with the aid of some matlab scripts, we perform some evaluations on the simulated signal samples through analytical expressions. Finally, we plot the symbol-error-probability versus input back-off from the 1-dB compression point (dB) for various SNRs, each giving a unique curve in Fig. 6-9. Next, we draw a horizontal line at the specified maximum tolerable BER (i.e. 10^{-3}). The intersection point of this line with the curve corresponding to $SNR = 8dB$ (as specified by the system requirements) is the nominal PA operating point that achieves the system requirements while it is as nonlinear as it possibly could be, resulting in an optimum performance/power-efficiency tradeoff.

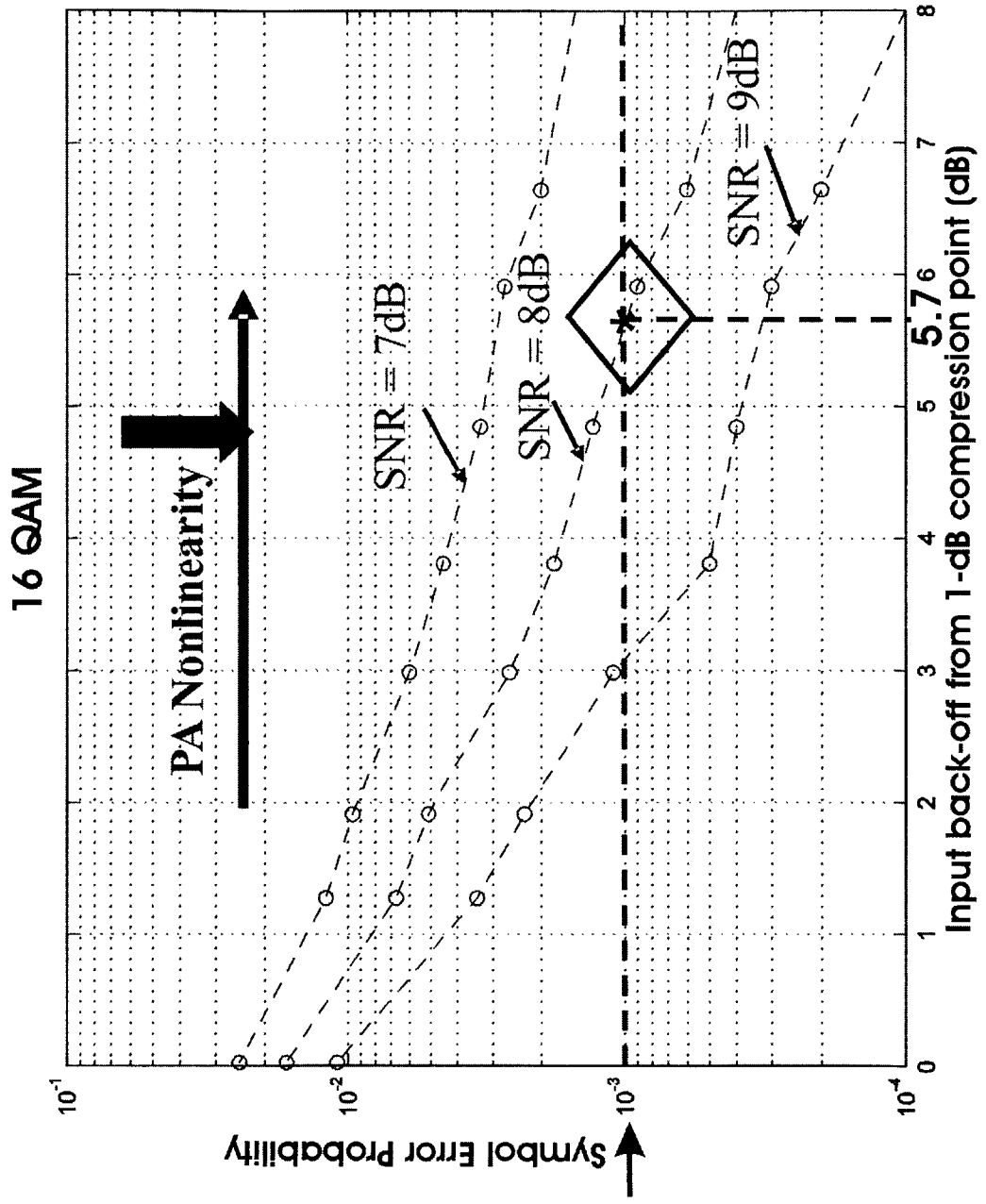

Figure 6-9: Finding the optimal PA operating point meeting the system requirement: 16-QAM modulation, maximum $SNR = 8dB$, and $BER \leq 10^{-3}$.

Chapter 7

Conclusions

In digital mobile band-limited communications systems, the use of bandwidth-efficient, time-varying envelope quadrature amplitude modulations in the presence of nonlinear power amplifiers presents a challenging tradeoff between linearity and power-efficiency. To optimize such a tradeoff, it is necessary to fully understand the effect of PA nonlinear distortion on the system performance.

More specifically, PA nonlinearities introduce distortions in the channel of interest. These in-channel distortions degrade the quality of the received signal, and are measured by the error probability (or bit-error-rate BER) at the detector.

Consequently, by developing an intuitive relationship between the PA nonlinearity and performance (BER), this thesis provides a tool that finds the nominal operating point of the PA that is at most nonlinearity (hence, achieving high power-efficiency) while still meeting the system performance criteria.

7.1 Summary

To obtain an intuitive understanding of the relationship between circuit nonlinearity and system performance, a typical digital communications system is simulated. The simulation is performed using the CppSim simulation tool, which is introduced in Chapter 4. The simulated system consists of a modulator and a PA at the transmitter, an AWGN channel, and a demodulator and an optimum detector at the receiver. Chapter 2 provides some background on these fundamental communications system blocks. Table 7.1 below summarizes the specifications of the simulation blocks.

To obtain meaningful results, the *PA* block is modeled using measured and simulated data from an experimented PA. Moreover, we found that the best way to characterize the PA nonlinearity (that would lead to an intuitive understanding of its effect on system performance), is the AM-AM/AM-PM model. In such a model, the PA nonlinear distortions in amplitude and phase of the output signal are characterized separately. In other words, when $x(t)$ (a voltage quantity with amplitude $|x(t)|$ and phase $\angle x(t)$) is fed to a nonlinear PA, with AM-AM function $g(.)$ and AM-PM function $f(.)$, the instantaneous output voltage $y(t)$ can be found by

$$y(t) = g(|x(t)|)e^{j[\angle x(t) + f(|x(t)|)]} \tag{7.1}$$

As indicated above, a remarkable observation is that both output amplitude and phase are

Modulator:	16-QAM/64-QAM, Gray-Encoding mapping, Rectangular constellation $T_s = 10^{-6}$ secs/symbol, sampling frequency = 10 MHz (thus, 10 samples/symbol) $R_{16-QAM} = 4/T_s = 4$ (Mega bits/secs), $R_{64-QAM} = 6/T_s = 6$ (Mega bits/secs) Square-root raised cosine baseband filter, roll-off factor=1, bandwidth=$1/T_s = 1\ MHz$ Transmitted signal: baseband
PA:	Class A, at 5.8 GHz, max output power = 20 dBm (=100 mWatts), gain = 17.5 dB
Channel:	Band-limited (to signal bandwidth $W = 1\ MHz$), baseband, AWGN
Demodulator:	Matched-filter: square-root raised cosine filter (same as the baseband filter) Optimum sampler
Detector:	Optimum minimum-distance (MD) detector for a linear PA

Table 7.1: Specifications of the system simulation blocks.

distorted as functions of the input amplitude only, justifying why the PA nonlinearity affects envelope-varying signals only. Chapter 5 provides the specifications of the experimented PA together with the AM-AM/AM-PM nonlinearity behavior of the simulation *PA* block and their implementations.

Next, Chapter 6 provides both qualitative and quantitative characterizations of the effect of the PA nonlinearity on BER. The intuition (or qualitative) understanding of this relationship is obtained by looking at how the PA nonlinearity distorts the shape of the input signal constellation. It is observed that the AM-AM nonlinearity amplifies low-amplitude signals linearly while it introduces more and more gain compression in signals with increasing amplitude levels (in the nonlinear region). In addition, the AM-PM nonlinearity introduces non-constant phase-shifts to signals of different amplitude levels; for this particular AM-PM characteristics, more phase-shifts are added to the phase of signals with higher amplitudes. The net result is a distorted signal constellation, in which some signals, in the presence or even absence of the AWGN channel, are detected in decision regions different from where they were originally transmitted. Hence, the effect of the PA nonlinearity on the signal constellation shape clearly reveals how it degrades the system BER.

Once we have gained an intuition for how the PA nonlinearity affects the BER, a quantitative analysis of such an effect is performed. This is done by plotting the simulated BER vs. the signal-to-noise ratio (SNR), and proposing a way that finds the PA operating point corresponding to each SNR. In other words, we have obtained a method that maps PA operating points to points on the BER vs. SNR curve. Hence, we can find the system performance BER that results from operating the PA at any point in its transfer characteristics.

Consequently, by having a tool that enables us to relate PA operating points to their resulting BERs, we can find the maximum nonlinear point that meets the maximum allowable BER and other system requirements. In fact, PA designers can use this tool - that consists of a system simulation followed by matlab scripts that construct the final PA operating-point/BER relationship - to design PAs that achieve the optimal linearity/efficiency performance.

7.2 Future Work

In this thesis, we focused primarily on the effect of the PA nonlinearity in distorting the signal in the channel of interest. However, as discussed in Chapter 3, in response to a time-varying envelope signal, besides in-channel distortion, the PA nonlinearity introduces interference in the adjacent channels. These adjacent-channel distortions are caused by spectral broadening (or spectral regrowth) of the modulated signal spectrum, and are commonly quantified by the adjacent-channel power ratio (ACPR). Hence, the next step is to establish a qualitative and quantitative understanding of the effect of PA nonlinearity on the system performance, bandwidth efficiency, or spectral regrowth.

In addition, to examine the direct effect of the PA nonlinearity on BER, we simulated an ideal communications system in which the nonlinear PA is the only source of non-ideality. Once such an effect is intuitively and quantitatively well-understood, the next step would be to upgrade the system elements to more realizable models of the physical systems of interest. Indeed, as we emphasized the ability to relate this circuit nonlinearity to a system performance is such a powerful tool in designing more effective systems that the extension of this work to more realizable systems is beneficial. There are many ways to pursue this objective, some of which are mentioned next.

One of the most obvious upgrades to the present ideal system includes replacing the AWGN channel with a more realistic channel model, such as one with a time-invariant or time-varying linear or nonlinear filter. In wireless systems, the constantly changing physical characteristics of the media (e.g. due to the moving scatterers) results in a *fading multipath* channel model with randomly time-variant impulse responses [1, p. 800]. For instance, when there are a large number of scatterers in the medium contributing to the received signal, the Central Limit Theorem results in modeling the channel impulse response by a Gaussian process. Depending on the presence or absence of fixed signal reflectors or scatterers (giving rise to a non-zero or zero mean process), the complex Gaussian channel is statistically modeled as a *Ricean fading* or a *Rayleigh fading* channel respectively.

In addition, besides thermal noise (characterized by an AWGN) that arises from the channel and electronics in receiver, there exist other noise sources in the system that can be included in the system simulation. One of the most general noise sources is the circuit noise (also modeled as an AWGN) at the transmitter that is added to the signal before going through the power amplifier.

As mentioned in Chapter 4, anther simplified assumption in our system simulation is the use of an optimum receiver for a linear PA. However, if we instead design a receiver optimal or nearly optimal for a nonlinear PA, the performance (BER) of the system in the presence of the nonlinear PA is expected to improve substantially. In fact, this is an important issue that must be investigated since it directly affects the maximum tolerable nonlinearity. There are many creative solutions for designing such an optimum receiver, one of which was suggested by Professor Muriel Medard (EECS department, MIT). In such a method, the idea is to use our knowledge of the way a particular PA nonlinearity distorts the signal constellation shape to define new (MD) decision threshold boundaries (set as bisectors of lines connecting pairs of adjacent symbols) on the distorted constellation. Hence, the employment of such a decoder results in elimination of the BER degradation due to the PA nonlinearity. However, such a decoder is optimum for a specific operating point of the PA while typically the PA operating point is varied according to the system link budget and the channel characteristics. One solution to this problem is to design optimum decoders

for several PA operating points, and subsequently choose the most suitable decoder to the actual PA operating point. In fact, depending on the particular application, one might come up with other variants of this solution to meet system requirements, e.g. on power, size and complexity.

In addition to the above implementation solution where the knowledge of the distorted signal constellation shape at the PA output is used to eliminate the PA nonlinear distortion, there exists another solution that can use this information. In this method, the information on the PA output constellation is transmitted through a feedback system to the input of the PA in order to pre-distort the input signal constellation such that the new output constellation closely resembles the ideal (linear PA) output. However, once again the implementation of this pre-distortion/linearization technique or a more relaxed variant of it depends on its complexity and feasibility.

References

[1] J. G. Proakis, *Digital Communications - Fourth Edition*, McGraw-Hill, 2000.

[2] B. Razavi, *RF microelectronics*, Prentice Hall, 1998.

[3] S. Benedetto, and E. Biglieri, *Principles of Digital Transmission With Wireless Applications,* Kluwer Academic/Plenum Publishers, 1999.

[4] R. G. Gallager, *Class Notes for Principles of Digital Communications*, Massachusetts Institute of Technology, Fall 2001.

[5] N. Pothecary, *Feedforward Linear Power Amplifiers,* Artech House, 1999.

[6] D. Forney, *Class Notes for Digital Communications*, Massachusetts Institute of Technology, 2000.

[7] S. C. Cripps, *RF Power Amplifiers for Wireless Communications,* Artech House, 1999.

[8] R. J. Westcott, "Investigation of Multiple F.M./F.D.M. Carriers through a Satellite T.W.T. Operating Near to Saturation," *Proceedings IEE*, Vol. 114, No. 6, June 1967.

[9] J. C. Fuenzalida, O. Shimbo, L. W. Cook, "Time-Domain Analysis of Intermodulation Effects Caused by Nonlinear Amplifiers," *COMSAT Technical Review*, Vol. 3, No. 1, Spring 1973.

[10] A. L. Berman, and E. I. Podraczky, "Experimental Determination of Intermodulation Distortion Produced in a Wideband Communications Repeater," *IEEE International Convention Record*, Pt. 2, Vol. 15, pp. 69-88, 1967.

[11] A. A. M. Saleh, "Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers," *IEEE Transactions on Communications*, Vol. COM-29, No. 11, November 1981.

[12] C. M. Thomas, M. Y. Weidner, and S. H. Durrani, "Digital Amplitude-Phase Keying with M-ary Alphabets," *IEEE Transactions on Communications*, Vol. COM-22, No. 2,pp. 168-180, Feb. 1974.

[13] M. Honkanen, and S. Haggman, "New Aspects on Nonlinear Power Amplifier Modeling in Radio Communication System Simulation," IEEE, 1997.

[14] A. Ghorbani, and M. Sheikhan, "The Effect of Solid State Power Amplifiers (SSPAs) Nonlinearities on MPSK and M-QAM Signal Transmission," *Sixth International Conference on Digital Processing of Signals in Communications*, pp. 193-197, 1991.

[15] M. H. Perrott, *CppSim Behavioral Simulation Package*, http://www.mtl-mit.edu/ perrott, 2002.

[16] X. Tang, M. Alouini, A. J. Goldsmith, "Effect of Channel Estimation Error on M-QAM BER Performance in Rayleigh Fading," *IEEE Transactions on Communications*, Vol. 47, No. 12, Dec. 1999.

[17] A. Pham, *Biasing Techniques for Linear Power Amplifiers*, Master's thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2002.

[18] N. A. Mathews, and A. H. Aghvami, "M-ary C.P.S.K. Signaling over Two-Link Nonlinear Channels in Addtiive Gaussian Noise," *IEE Proceedings*, Vol. 127, Pt. F. No. 5, October 1980.

[19] A. H. Aghvami, "Performance Analysis of 16-ary QAM Signaling through Two-Link Nonlinear Channels in Additive Gaussian Noise," *IEE Proceedings*, Vol. 131, Pt. F. No. 4, pp. 403-406, July 1984.

[20] R. Giubilei, "Correspondence: Comments on Semianalytic BER Evaluation by Simulation for Noisy Nonlinear Bandpass Channels," *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 1, Jan. 1989.