2009

Information Theoretic Identification and Compensation of Nonlinear Devices

Sepideh Dolatshahi
University of Massachusetts Amherst, sepid.dsh@gmail.com

INFORMATION THEORETIC IDENTIFICATION AND COMPENSATION OF NONLINEAR DEVICES

A Thesis Presented

by

SEPIDEH DOLATSHAHI

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL AND COMPUTER ENGINEERING

September 2009

Electrical and Computer Engineering

INFORMATION THEORETIC IDENTIFICATION AND COMPENSATION OF NONLINEAR DEVICES

A Thesis Presented

by

SEPIDEH DOLATSHAHI

Approved as to style and content by:

Dennis L. Goeckel, Co-chair

Hossein Pishro-Nik, Co-chair

Patrick A. Kelly, Member

C. V. Hollot, Department Head
Electrical and Computer Engineering

To my Mom for her constant sacrifices and best wishes that make my life easy

To my Father for his endless encouragements and from whom I learned the first concepts of Electrical Engineering

ACKNOWLEDGMENTS

This research project would not have been possible without the support of many people. The author wishes to express her gratitude to her advisors, Prof. Dennis L. Goeckel and Dr. Hossein Pishro-Nik who were abundantly helpful and offered invaluable assistance, support and guidance. Professor Goeckel makes a complete role model for advisor-advisee interactions. Deepest gratitude are also due to Professor Patrick A. Kelly for being such a good teacher and for his kind assistance. Special thanks also to Professor Robert W. Jackson and his student Arash Mashayekhi for providing me with the amplifier data and to Professor Brian Neil Levine for help on anonymity aspects. The author would also like to convey thanks to Analog Devices for financial support.

ABSTRACT

INFORMATION THEORETIC IDENTIFICATION AND COMPENSATION OF NONLINEAR DEVICES

SEPTEMBER 2009

SEPIDEH DOLATSHAHI

B.S., UNIVERSITY OF TEHRAN

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Dennis L. Goeckel and Professor Hossein Pishro-Nik

Breaking the anonymity of different wireless users with the purpose of decreasing internet crime rates is addressed in this thesis by considering radiometric identification techniques.

Minute imperfections and non-idealities in the different transmitter components, especially the inherent nonlinearity in power amplifiers, result in variations in their Volterra series representations which could be utilized as a signature.

For a two user scenario, signal processing algorithms based on generalized likelihood ratio test(GLRT) and the classical likelihood ratio test are introduced and the resulting receiver decision rules and performance curves are presented. These algorithms consider the

high signal to noise ratio(SNR) case where we have available the input samples completely at the receiver which is a practical assumption for most cases.

Volterra series are widely used in behavioral modeling of power amplifiers. To validate the existence of these variations in the Volterra series representation of power amplifiers, process variations are introduced as major sources. The plausibility of our techniques are justified by deriving and comparing the Volterra coefficients for the fast and slow process corners.

Finally,an information theoretic framework is presented where the amount of mutual information of the output about the Volterra coefficients represents the amount of anonymity taken from users. Here, some results for the low SNR case are presented to prove the achievability of some information about individual systems using our hardware anonymity breaking techniques.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Motivation

1.1.1 Anonymity

Modern life with the Internet as a principal part provides different crime opportunities and crime types. Sexual exploitation of children, production and dissemination of contraband music and video, intellectual property theft, murder, identity theft, financial fraud and espionage are some instances of crimes which are either created or made easier by the advent of computers and use of the internet. For example exploiting open wireless access points(AP's) hosted by private homes, businesses, and municipalities provides offenders with de facto anonymity. Fortunately the use of computers by such offenders typically results in digital evidence.

The primary artifacts used in investigation of internet crimes are the Internet Protocol(IP)address and Media Access Control address (MAC address) of the suspect's computer both of which cannot be relied on by crime investigators. Consistent IP addresses are assigned by an Internet Service Provider(ISP)and all incoming and outgoing messages are tagged by this IP address. The reason why IP addresses are not considered reliable is the proliferation of open AP's hosted by many private and public places these days. Dynamic IP addresses, which are temporarily assigned IP addresses from a pool of possible

IP addresses for the duration of that internet session or for some other specified amount of time, are most frequently assigned on LANs and broadband networks by Dynamic Host Configuration Protocol (DHCP) servers. They are used because it avoids the administrative burden of assigning specific static addresses to each device on a network. It also allows many devices to share limited address space on a network if only some of them will be online at a particular time. Being temporary they are thus not considered trustworthy evidence. MAC addresses are similarly unreliable as they are easily reconfigured by the user. Another problem is the use of encrypted network connections which can be employed at the link, network, or transport layers. Wi-Fi Protected Access (WPA), Internet Protocol Security(IPSec), and the Secure Shell protocol(SSH) are respective examples of these encryptions. with such protection utilized, it is difficult to attribute network activities to a particular user in terms of the content or the internet destination. In general for software anonymity breaking techniques usually some type of software security and protection technique can be developed and thus the results of these anonymity breaking techniques are not considered reliable evidence.

1.1.2 Imperfections at the PHY layer

In this section, we detail how we will develop new methods of collecting identifying characteristics of radio transceivers at the physical layer — even if the higher layers provide a high degree of anonymity. The paradigm behind our approach is simple. There are long-standing imperfections in the RF portion of any wireless transmitter that still exist despite decades of significant efforts by the commercial and government microwave circuits community. By exploiting these imperfections, an observer can group together RF

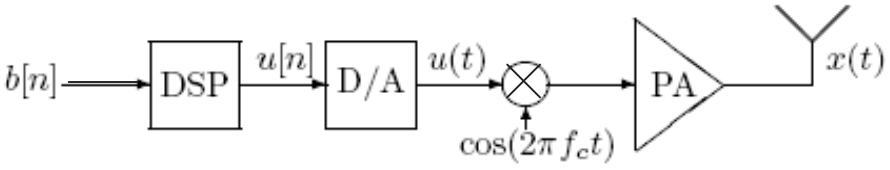

Figure 1.1. Block Diagram of a Standard Wireless Transmitter, where $b[n]$ is the sequence of bits to be transmitted, $u[n]$ and $u(t)$ are the digital and analog baseband waveforms, respectively, and $x(t)$ is the transmitted signal.

signals from one radio that would otherwise be anonymous due to changes in MAC or IP addresses.

An extremely simplified version of the transmit chain for a wireless transmitter is shown in Figure 1.1. In particular, a digital (discrete-time, discrete-amplitude) baseband signal $u[n]$ that carries the information bits is generated by a digital signal processor (DSP). This signal is then converted to an analog signal, which is upconverted to the desired carrier frequency and then amplified by the power amplifier (PA). In an ideal system, the transmitted signal would be given by $x(t) = Au(t)\cos(2\pi f_c t + \theta)$, where A is the gain of the power amplifier, $u(t)$ is the ideal analog form of $u[n]$ (i.e., the sinc$(\cdot)$-interpolated version of $u[n]$), f_c is the desired carrier frequency, and θ is the (constant) phase of the oscillator. However:

- The **digital-to-analog (D/A)** conversion suffers from the finite precision of the digital input, but, more importantly, particularly for our forensics work, the analog output for a given digital input can vary significantly across converters.

- The **oscillator**, which for our work will be defined as both the crystal and the associated phase-locked loop (PLL), attempts to generate the sinusoid $\cos(2\pi f_c t + \theta)$, but the actual frequency/phase of the sinusoid generated can vary greatly from crystal to crystal.
- **Power amplifiers**, which seek to produce a linear device that takes in $u(t)$ and puts out $Au(t)$ are often quite nonlinear — even with significant compensation. As with the D/A converter and oscillator, this variation can be significant across devices.

Our approach based on radiometric identity makes use of the minor variations in analog hardware of transmitters which manifest themselves as idiosyncratic artifacts in their emitted signals.

1.1.3 Proposed Work

First, the feasibility of this approach was investigated by measuring the output frequency of a couple of higher quality oscillators from manufacturer 1 and a couple of lower quality oscillators from manufacturer 2. The results which are included in the following sections showed empirically that there are some parameters in different transmitter components including the center frequency of the oscillator in the mixer that could be exploited to tell different users apart.

Next, we will pose the theoretical simplified problem of two nonlinear systems with different system parameters(in our case Volterra series coefficients) and we will propose algorithms to break their anonymity. The first algorithm introduced relies on Generalized Likelihood Ratio Test(GLRT) to differentiate between the two different users in the high signal to noise ratio(SNR) case and we will show its effectiveness.A more practical

algorithm which requires less storage is introduced next. This algorithm first estimates the Volterra coefficients using least squares(LS) estimation and then solves the resulting detection problem using the classical likelihood ratio test.

At last, we then need to form some justified proof for our claim that different amplifiers even from the same manufacturer have different Volterra series representation vectors, far enough to let us differentiate between them with low probability of error. This is done by using the results of simulating a simple power amplifier and taking into account the process variations.

1.2 Background

There are many vulnerabilities that we could exploit to perform our forensic work. We could use non-idealities in different transmitter components including:

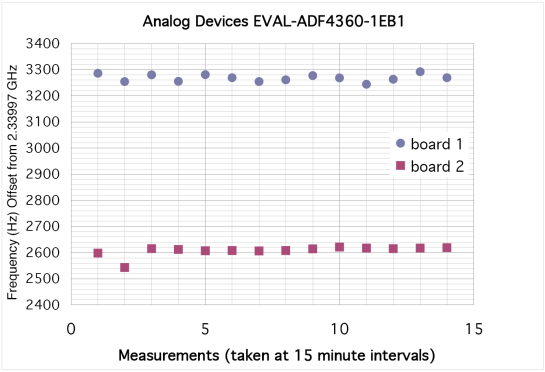

Figure 1.2. The frequencies of two oscillators measured 14 times over 3.5 hours demonstrating consistently measurable differences and identifying characteristics

Oscillator. An ideal oscillator would produce the signal $\cos(2\pi f_c t + \theta)$ at the desired f_c for a given channel. However, this is rarely the case in practice. Instead the following signal is produced:

$$y(t) = \cos(\, 2\pi(f_c + \Delta(t))t + \Theta(t)\,),$$

where $2\pi\Delta(t)t + \Theta(t)$ is the (time-varying) phase noise. The frequency offset $\Delta(t)$ has characteristics specific to a given chip, but it is time-varying due to environmental fluctuations, particularly in device temperature. When one reads the data sheets for crystal oscillators, the numbers for this frequency offset look impressive — measured in at most tens of and often a fraction of a "ppm" (part per million) [36]. However, when one considers a multi-GHz carrier, these offsets become significant.

We measured the frequency of each of two very good oscillators for the 2.4 GHz band over 3.5 hours, and the results are shown in Figure 1.2. It is very clear that one could easily group transmissions from one user together based on the carrier frequency. Furthermore, frequency compensation is a critical synchronization function in any wireless receiver [37, Chapter 6], and for small offsets, this frequency compensation is often done digitally at the receiver by estimating the frequency offset and then performing compensation by multiplication by the appropriate sinusoid on the DSP. Thus, to establish communication, any "standard" receiver for the system must be able to compute this offset within relatively tight bounds, and hence a powerful eavesdropper could also easily track it as well.

Amplifier. An ideal amplifier would produce the signal $Au(t)$ when given the input $u(t)$. However, a standard power amplifier is only linear at very low powers where it runs quite inefficiently, and thus amplifiers in small wireless cards or cell phones, which are what we wish to identify, are run in the nonlinear regime. Hence, compensation for this nonlinearity has been one of the most active of all research areas for commercial RF companies. Why not simply pre-compensate for the amplifier nonlinearity characteristics at design time? Therein lies the difficulty — every amplifier, even of the same part number,

exhibits different characteristics that have to be measured and then compensated individually for, if the amplifier linearity is to be improved [39]. Obviously, it is the uniqueness of the characteristics that we will attempt to exploit in our proposed work.

1.2.1 Other Works

There have been a number of *Radio Frequency Fingerprinting*(RFF) efforts over the years. Much of the work has been in the microwave circuit community [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23], with most of it based on transient analysis. A transient is a brief radio emission produced while the power output of an RF amplifier goes from idle to the level required for data communication. The nature of transients is such that they are difficult to detect and there is no obvious correct way to succinctly describe them. The extended RFF process, including the identification of devices, consists of four key phases. The first phase involves the extraction of features (e.g. amplitude, phase or frequency information) from the digital signal. These features are subsequently used to detect the start of the transient in the second phase. Once the end of the transient has been estimated, typically in an experimental manner, the fingerprint (features representing the transient) is obtained. Finally, the transceiver of the device is identified based on the classification of the fingerprint.In the papers on transients in the literature, different parts of this process has been altered. For instance in [16, 17]the feature used in the first phase is the amplitude while in [14] the phase information is also exploited.In [22, 15] a wavelet analysis is used to characterize the features contained in the transient.

The term RF fingerprinting, in general, refers to various PHY layer classification approaches of RF signals. We broadly classify RF features into: (i) channel-specific ones,

e.g., channel impulse response, that characterize the wireless channel; and (ii) transmitter-specific ones that are independent of the channel, e.g., signal encoding. Since channel-specific features uniquely identify the channel between the transmitter and the receiver, they have been successfully adopted in robust location distinction. There have also been a few significant recent works in the networking community, through location identification(channel-specific) [24, 25, 26, 27, 28], which would allow one to group transmissions from a stationary user. Some of these including [26, 27] are based upon the ability of the multipath environment to provide a waveform whose structure an adversary cannot measure or model accurately. The rapid decorrelation properties of the multipath channel is exploited. Temporal and spectral variability is reflected by two notions, the coherence time and coherence bandwidth of the channel. Spatial separation of one to two wavelengths is sufficient for assuming independent fading paths.

Another location distinction based technique uses the Received Signal Strength (RSS) to distinguish transmitters [25, 28]. An RSS method simply uses the RSS measured at multiple receivers as a feature vector. RSS measurements contain information about a link but vary due to small-scale and frequency-selective fading, such that its use in location distinction requires multiple measurements at different receivers. Also, in the network security application, adversaries can 'spoof' their signalprint using array antennas which send different signal strengths in the directions of different access points. Moreover, for wireless sensor networks, multi-node collaboration is expensive in terms of energy. These location based techniques assume that different transmitters remain active and do not move and thus lack the ability to actually make an identification or recognize a previously seen device that moved or sat silent for some time.

The transmitter specific RF fingerprinting techniques rely on the exploitation of device non-idealities(transmitter-specific) [30, 31, 32, 29] (see [30] for a thorough review of prior work). Per above, approaches in the RF community generally consider very specific observed transient phenomena of the RF signal. As a representative example, Remley et al [29] measure the envelope of a number of different wireless local area networks (WLAN) cards and note that the envelopes of the waveforms on an oscilloscope for different cards *look different*. At the other extreme, recent work by Brik et al [30] used machine learning techniques on collected modulation data to train data-agnostic classifiers that are then able to distinguish wireless cards that are produced by the same vendor.

On the other hand we exploit the minute imperfections in the different transmitters hardware even from the same manufacturer that manifest themselves as the difference in the Volterra series representations of for instance the power amplifier in the transmitter circuitry. We suggest signal processing detection and classification techniques and support the feasibility of out techniques theoretically. We also do not need long input vectors and we do not have big memory requirements as out procesing is considered realtime.

1.3 Contribution

1.3.1 Formal Approach

Our broad approach to device modeling, anonymity analysis, and algorithm design is significantly different than these prior efforts. In particular, the approach here is focused on a comprehensive understanding and exploitation of the phenomena being exploited for node identification. This will yield an accurate (generally statistical) model amenable to analysis by researchers at the physical layer, thus allowing us to answer fundamental ques-

tions: (1) how much anonymity is forfeited by such devices?, (2) what are the key device characteristics that cause such anonymity loss?, and, particularly important from an operational point of view, (3) how might the nodes employ countermeasures to regain some anonymity and how would such be thwarted? For example, in contrast to the recent empirical classification results of [30] on commercial 802.11 cards, our modeling and analysis could provide clear understanding of countermeasures that will be particularly effective (likely frequency offset dithering) and those that will not (likely amplifier nonlinearity modification). Provable performance is the key characteristic to our approach.

1.3.2 Thesis Outline

In Chapter 2, we will discuss a simple two-user scenario in which there are only two possible transmitters transmitting in a role and a third time one of them transmits. We would like to determine which transmitter transmitted the third signal. For the high SNR case, we propose two algorithms to distinguish different users.First, a Generalized Likelihood Ratio Test(GLRT)-based algorithm and another algorithm based on the classical likelihood ratio test are introduced. Then, the performance of this technique by writing the probability of error in terms of the difference vector of the two systems was studied. Also, if we consider the coefficients to be random variables around some certain mean, the performance of our algorithms are plotted versus the variance of these random variables. The results of the simulations show that the average probability of error decreases when the norm of the difference vector of system parameters increases and that the classical likelihood ratio test performs better than the GLRT algorithm. In addition, the classical likelihood ratio algorithm only needs to store the volterra series coefficients of the two or more possible transmitters and is thus more practical in this sense.

In Chapter 2, when we talk about the system parameters, we mean the truncated form of the Volterra series coefficients. To be able to link this to the main anonymity breaking application discussed in this introduction chapter, we need to discuss some of the sources that cause these variations in the Volterra coefficients of different amplifiers. Chapter 3 talks about the modeling side of this project and uses the results of the simulations of a simple class A amplifier and its variations due to the process variations of the NMOS transistor inside the amplifier to show how different the Volterra series coefficients could be.

Chapter 4 provides theoretical proof for the possibility of applying the hardware anonymity breaking techniques to the low SNR case and in the mean time introduces another approach to studying anonymity problems which is measuring the amount of information conveyed by the output from the system parameters.

Finally chapter 5 concludes the work done in this thesis and suggests future researches in this topic.

CHAPTER 2

IDENTIFYING AMPLIFIERS VIA VOLTERRA COEFFICIENTS

2.1 Problem statement

Assume there are only two possible users connecting to a wireless LAN. We have samples from the input and output of these two wireless cards. Call the input vectors of size (M+1) X_1 and X_2, and their output vectors of size M, Y_1 and Y_2 respectively. The transmitters in the wireless cards are in general nonlinear systems because of the nonlinear components they have including power amplifiers(PA's). Figure 2.1 shows another input vector X_3 passing through either system 1 or system 2. Assuming perfect errorless decoding of the input we have access to X_3 and Y_3, the input and output, and we would like to determine which system it went through.

2.1.1 Problem settings-high SNR case

For the high signal to noise ratio(SNR), we could accurately decode the message and thus know X at the receiver. This input vector($\underline{\boldsymbol{X}}_i$) goes through the power amplifier in the transmitter of the system to be identified which is characterized by its Volterra series coefficients.Then zero mean Gaussian channel noise is added to the signal and at the receiver the output vector $\underline{\boldsymbol{Y}}_i$ is received.

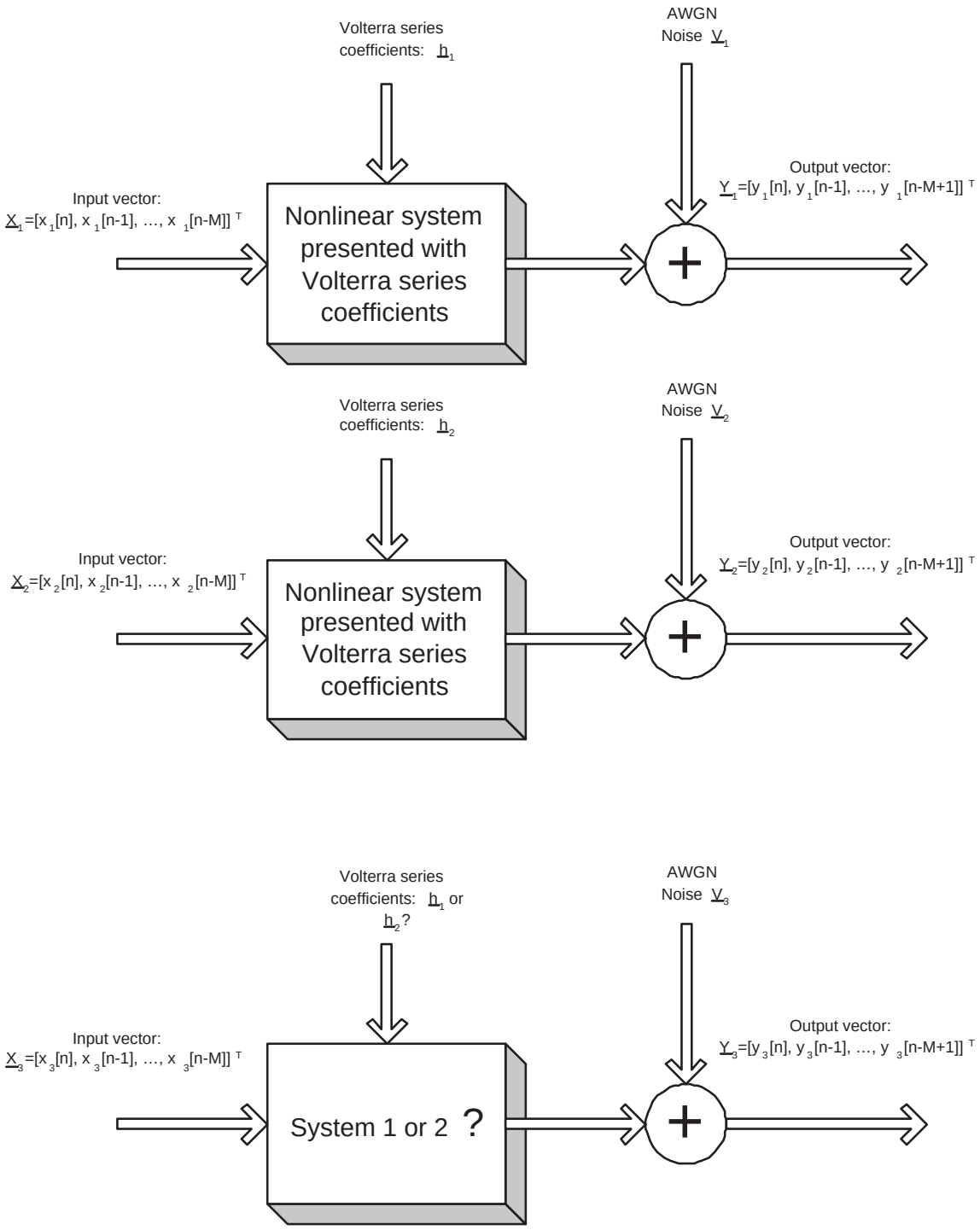

Figure 2.1. The two-system identification scenario.

$\underline{\boldsymbol{X}}_i$ is the input vector of the system i known at the receiver for i=1,2,3. We consider the $(M+1)\times 1$ input vector to the system i is drawn from a zero-mean Gaussian random process with variance σ_X^2:

$$\underline{\boldsymbol{X}}_i = \begin{bmatrix} x_i(n) \\ x_i(n-1) \\ \vdots \\ x_i(n-M) \end{bmatrix} \quad ; \quad i = 1, 2, 3.$$

The $M \times 1$ additive noise vector is also drawn from a zero-mean Gaussian random process with variance σ_n^2:

$$\underline{\boldsymbol{\nu}}_i = \begin{bmatrix} \nu_i(n) \\ \nu_i(n-1) \\ \vdots \\ \nu_i(n-M) \end{bmatrix} \quad ; \quad i = 1, 2, 3.$$

We use the Volterra series representation with memory of order 1 for our nonlinear devices, and we assume that the nonlinear systems are well-represented using the Volterra series up to order two, which is called a linear quadratic system and is in the form(see appendix A):

$$y_i(n) = \Sigma_{k_1=0}^{1} h_{i,1}(k_1) x_i(n-k_1) + \Sigma_{k_1=0}^{1}\Sigma_{k_2=0}^{1} h_{i,2}(k_1,k_2) x_i(n-k_1)x_i(n-k_2) + \nu_i(n) \tag{2.1}$$

$$y_i(n) = h_{i,1}(0)x_i(n) + h_{i,1}(1)x_i(n-1)$$

$$+h_{i,2}(0,0)x_i^2(n) + h_{i,2}(1,1)x_i^2(n-1) + h_{i,2}(0,1)x_i(n)x_i(n-1) + \nu_i(n) \qquad (2.2)$$

Note that this just simplifies notation. The concept applies to higher order Volterra series.

Thus, the system parameters vector $\underline{\boldsymbol{h_i}}$ would be an $N \times 1$ vector where $N = 5$:

$$\underline{\boldsymbol{h}}_i = \begin{bmatrix} h_{i,1}(0) \\ h_{i,1}(1) \\ h_{i,2}(0,0) \\ h_{i,2}(1,1) \\ h_{i,2}(0,1) \end{bmatrix} \quad ; \quad i = 1, 2.$$

Now in vector form we have:

$$\underline{\boldsymbol{Y}}_i = \begin{bmatrix} x_i(n) & x_i(n-1) & x_i^2(n) & x_i^2(n-1) & x_i(n)x_i(n-1) \\ x_i(n-1) & x_i(n-2) & x_i^2(n-1) & x_i^2(n-2) & x_i(n-1)x_i(n-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i(n-M+1) & x_i(n-M) & \cdots & \cdots & x_i(n-M+1)x_i(n-M) \end{bmatrix} \underline{\boldsymbol{h}}_i + \underline{\boldsymbol{\nu}}_i$$

or equivalently:

$$\underline{\boldsymbol{Y}}_i = P_i \underline{\boldsymbol{h}}_i + \underline{\boldsymbol{\nu}}_i \quad ; \quad i = 1, 2.$$

Now that we have defined the $M \times N$ matrices $P_i \quad i = 1, 2, 3$, let's define the $2M \times N$ matrices P_{13} and P_{23} formed by stacking matrices P_1 and P_3, and P_2 and P_3 respectively:

$$P_{i3} = \begin{bmatrix} P_i \\ P_3 \end{bmatrix} \quad where \quad i = 1, 2$$

$$\underline{Y}_{i3} = \begin{bmatrix} \underline{\boldsymbol{Y}}_i \\ \underline{\boldsymbol{Y}}_3 \end{bmatrix} \quad where \quad i = 1, 2$$

2.2 The proposed GLRT Estimation Procedure

2.2.1 The GLRT receiver

According to the generalized likelihood ratio test(GLRT test):

$$max_{\underline{\boldsymbol{h}}_1}\{P(\underline{\boldsymbol{Y}}_1, \underline{\boldsymbol{Y}}_3 \,|\, \underline{\boldsymbol{h}}_1, \underline{\boldsymbol{X}}_1, \underline{\boldsymbol{X}}_3)\} \underset{\underline{\boldsymbol{h}}_1}{\overset{\underline{\boldsymbol{h}}_2}{\gtrless}} max_{\underline{\boldsymbol{h}}_2}\{P(\underline{\boldsymbol{Y}}_2, \underline{\boldsymbol{Y}}_3 \,|\, \underline{\boldsymbol{h}}_2, \underline{\boldsymbol{X}}_2, \underline{\boldsymbol{X}}_3)\} \tag{2.3}$$

Given the inputs, $\underline{\boldsymbol{h_1}}$, and $\underline{\boldsymbol{h_2}}$ are known,

$$\underline{\boldsymbol{Y}}_{i3}(Given \underline{\boldsymbol{X}}_i, \underline{\boldsymbol{X}}_3, \underline{\boldsymbol{h}}_i) \sim N(P_{i3}.\underline{\boldsymbol{h}}_i, \sigma_n^2 I_{(2M \times 2M)}), \quad i = 1, 2.$$

Or equivalently:

$$P_{\underline{\boldsymbol{Y}}_{i3}|\underline{\boldsymbol{X}}_i, \underline{\boldsymbol{X}}_3, \underline{\boldsymbol{h}}_i}(\underline{\boldsymbol{Y}}_{i3}|\underline{\boldsymbol{X}}_i, \underline{\boldsymbol{X}}_3, \underline{\boldsymbol{h}}_i) = \frac{1}{(\sqrt{2\pi}\sigma_n)^{2M}} e^{-\frac{(\underline{\boldsymbol{Y}}_{i3} - P_{i3}.\underline{\boldsymbol{h}}_i)^H (\underline{\boldsymbol{Y}}_{i3} - P_{i3}.\underline{\boldsymbol{h}}_i)}{2\sigma_n^2}} \tag{2.4}$$

Now substituting these in the main GLRT formula yields:

$$min_{\underline{\boldsymbol{h}}_1}\{(\underline{\boldsymbol{Y}}_{13}-P_{13}.\underline{\boldsymbol{h}}_1)^H(\underline{\boldsymbol{Y}}_{13}-P_{13}.\underline{\boldsymbol{h}}_1)\} \underset{\underline{\boldsymbol{h}}_2}{\gtrless} min_{\underline{\boldsymbol{h}}_2}\{\underline{\boldsymbol{Y}}_{23}-P_{23}.\underline{\boldsymbol{h}}_2)^H(\underline{\boldsymbol{Y}}_{23}-P_{23}.\underline{\boldsymbol{h}}_2) \tag{2.5}$$

Let $|e_i|^2$ be:

$$|e_i|^2=(\underline{\boldsymbol{Y}}_{i3}-P_{i3}.\underline{\boldsymbol{h}}_i)^H(\underline{\boldsymbol{Y}}_{i3}-P_{i3}.\underline{\boldsymbol{h}}_i)=|\underline{\boldsymbol{Y}}_{i3}-P_{i3}.\underline{\boldsymbol{h}}_i|^2 \quad i=1,2.$$

Minimizing $|e_i|^2$ is the same problem as the Least Squares(LS) problem where the number of equations to estimate the parameter is more than the number of parameters(in this case $\underline{\boldsymbol{h}}_i$'s with $S=5$ elements). Thus we could apply the results of the LS problem:

$$\underline{\boldsymbol{h}}_{i,OPT}=(P_{i3}^HP_{i3})^{-1}P_{i3}^H\underline{\boldsymbol{Y}}_{i3}^H \quad ,i=1,2. \tag{2.6}$$

$$|e_{i,OPT}|^2=\underline{\boldsymbol{Y}}_{i3}^H(I_{2M\times 2M}-P_{i3}(P_{i3}^HP_{i3})^{-1}P_{i3}^H)\underline{\boldsymbol{Y}}_{i3} \quad ,i=1,2. \tag{2.7}$$

At the receiver the GLRT decision rule will be:

$$\underline{\boldsymbol{Y}}_{13}^H(I_{2M\times 2M}-P_{13}(P_{13}^HP_{13})^{-1}P_{13}^H)\underline{\boldsymbol{Y}}_{13} \underset{\underline{\boldsymbol{h}}_2}{\gtrless} \underline{\boldsymbol{Y}}_{23}^H(I_{2M\times 2M}-P_{23}(P_{23}^HP_{23})^{-1}P_{23}^H)\underline{\boldsymbol{Y}}_{23}\} \tag{2.8}$$

2.2.2 Performance analysis

To determine the performance of this method and its ability to differentiate between two different transmitters, we should find the probability of error in terms of some form of distance between the two system parameters vectors $\underline{\boldsymbol{h_1}}$ and $\underline{\boldsymbol{h_2}}$.

$$P_e = Pr\{\underline{h}_1\}.Pr\{\text{GLRT results: } \underline{h}_2|\underline{h}_1\} + Pr\{\underline{h}_2\}.Pr\{\text{GLRT results: } \underline{h}_1|\underline{h}_2\} \quad (2.9)$$

Because of the symmetry:

$$P_e = Pr\{\text{GLRT results: } \underline{h}_2|\underline{h}_1\} \quad (2.10)$$

We are interested in $Pr\{max_{\underline{\boldsymbol{h_1}}}\{P(\underline{\boldsymbol{Y_1}}, \underline{\boldsymbol{Y_3}} \,|\, \underline{\boldsymbol{h_1}}, \underline{\boldsymbol{X_1}}, \underline{\boldsymbol{X_3}})\} < max_{\underline{\boldsymbol{h_2}}}\{P(\underline{\boldsymbol{Y_2}}, \underline{\boldsymbol{Y_3}} \,|\, \underline{\boldsymbol{h_2}}, \underline{\boldsymbol{X_2}}, \underline{\boldsymbol{X_3}})\}|\underline{\boldsymbol{h_1}}\}$
Or:

$$P_e = Pr\{min_{\underline{\boldsymbol{h_1}}}\{(\underline{\boldsymbol{Y}}_{13} - P_{13}.\underline{\boldsymbol{h}}_1)^H(\underline{\boldsymbol{Y}}_{13} - P_{13}.\underline{\boldsymbol{h}}_1)\} < min_{\underline{\boldsymbol{h_2}}}\{(\underline{\boldsymbol{Y}}_{23} - P_{23}.\underline{\boldsymbol{h}}_2)^H(\underline{\boldsymbol{Y}}_{23} - P_{23}.\underline{\boldsymbol{h}}_2)\}|\underline{\boldsymbol{h}}_1\} \quad (2.11)$$

According to the definition of $|e_{i,opt}|^2$ in (2.7):

$$P_e = Pr\{|e_{1,opt}|^2 < |e_{2,opt}|^2|\underline{\boldsymbol{h}}_1\} \quad (2.12)$$

$$P_e = Pr\{\underline{\boldsymbol{Y}}_{13}^H(I_{2M\times 2M} - P_{13}(P_{13}^H P_{13})^{-1}P_{13}^H)\underline{\boldsymbol{Y}}_{13} < \underline{\boldsymbol{Y}}_{23}^H(I_{2M\times 2M} - P_{23}(P_{23}^H P_{23})^{-1}P_{23}^H)\underline{\boldsymbol{Y}}_{23}|\underline{\boldsymbol{h}}_1\} \quad (2.13)$$

Knowing that the third system was actually system 1 with system parameters $\underline{\boldsymbol{h_1}}$ or equivalently substituting $\underline{Y}_{i3}$ by:

$$\underline{\boldsymbol{Y}}_{13} = \begin{bmatrix} \underline{\boldsymbol{Y}}_1 \\ \underline{\boldsymbol{Y}}_3 \end{bmatrix} = \begin{bmatrix} P_1\underline{\boldsymbol{h}}_1 + \underline{\boldsymbol{v}}_1 \\ P_3\underline{\boldsymbol{h}}_1 + \underline{\boldsymbol{v}}_3 \end{bmatrix} = P_{13}\underline{\boldsymbol{h}}_1 + \begin{bmatrix} \underline{\boldsymbol{v}}_1 \\ \underline{\boldsymbol{v}}_3 \end{bmatrix} \tag{2.14}$$

$$\underline{\boldsymbol{Y}}_{23} = \begin{bmatrix} \underline{\boldsymbol{Y}}_2 \\ \underline{\boldsymbol{Y}}_3 \end{bmatrix} = \begin{bmatrix} P_2\underline{\boldsymbol{h}}_2 + \underline{\boldsymbol{v}}_2 \\ P_3\underline{\boldsymbol{h}}_1 + \underline{\boldsymbol{v}}_3 \end{bmatrix} = \begin{bmatrix} P_2\underline{\boldsymbol{h}}_2 \\ P_3\underline{\boldsymbol{h}}_1 \end{bmatrix} + \begin{bmatrix} \underline{\boldsymbol{v}}_1 \\ \underline{\boldsymbol{v}}_3 \end{bmatrix} \tag{2.15}$$

yields:

$$|e_1|^2(Given\underline{\boldsymbol{h}}_1) = (P_{13}\underline{\boldsymbol{h}}_1 + \underline{\boldsymbol{v}}_{13})^H(I_{2M\times 2M} - P_{13}(P_{13}^H P_{13})^{-1}P_{13}^H)(P_{13}\underline{\boldsymbol{h}}_1 + \underline{\boldsymbol{v}}_{13}) \tag{2.16}$$

$$= \underline{\boldsymbol{v}}_{13}^H\underline{\boldsymbol{v}}_{13} - \underline{\boldsymbol{v}}_{13}^H P_{13}(P_{13}^H P_{13})^{-1}P_{13}^H\underline{\boldsymbol{v}}_{13}) \tag{2.17}$$

where

$$\underline{\boldsymbol{v}}_{i3} = \begin{bmatrix} \underline{\boldsymbol{v}}_i \\ \underline{\boldsymbol{v}}_3 \end{bmatrix} \quad , i = 1, 2.$$

And, also,

$$|e_2|^2 = (\begin{bmatrix} P_2\underline{\boldsymbol{h}}_2 \\ P_3\underline{\boldsymbol{h}}_1 \end{bmatrix} + \underline{\boldsymbol{v}}_{23})^H(I_{2M\times 2M} - P_{23}XP_{23}^H)(\begin{bmatrix} P_2\underline{\boldsymbol{h}}_2 \\ P_3\underline{\boldsymbol{h}}_1 \end{bmatrix} + \underline{\boldsymbol{v}}_{23}) \tag{2.18}$$

$$= A + B + C + D \tag{2.19}$$

where:

$$X = (P_{23}^H P_{23})^{-1} = (P_2^H P_2 + P_3^H P_3)^{-1} \tag{2.20}$$

$$A = \begin{bmatrix} P_2\underline{\boldsymbol{h}}_2 \\ P_3\underline{\boldsymbol{h}}_1 \end{bmatrix}^H (I_{2M\times 2M} - P_{23}XP_{23}^H) \begin{bmatrix} P_2\underline{\boldsymbol{h}}_2 \\ P_3\underline{\boldsymbol{h}}_1 \end{bmatrix} \tag{2.21}$$

$$= (\underline{\boldsymbol{h}}_2 - \underline{\boldsymbol{h}}_1)^H P_2^H P_2 X P_3^H P_3 (\underline{\boldsymbol{h}}_2 - \underline{\boldsymbol{h}}_1) \tag{2.22}$$

$$= \underline{\boldsymbol{d}}^H P_2^H P_2 X P_3^H P_3 \underline{\boldsymbol{d}} \tag{2.23}$$

where:

$$\underline{\boldsymbol{d}} = \underline{\boldsymbol{h}}_2 - \underline{\boldsymbol{h}}_1 \tag{2.24}$$

$$B = \underline{\boldsymbol{v}}_{23}^H \begin{bmatrix} I_{M\times M} - P_2XP_2^H & -P_2XP_3^H \\ -P_3XP_2^H & I_{M\times M} - P_3XP_3^H \end{bmatrix} \begin{bmatrix} P_2\underline{\boldsymbol{h}}_2 \\ P_3\underline{\boldsymbol{h}}_1 \end{bmatrix} \tag{2.25}$$

$$= \underline{\boldsymbol{v}}_{23}^H \begin{bmatrix} P_2XP_3^HP3 \\ -P_3XP_2^HP2 \end{bmatrix} \underline{\boldsymbol{d}} \tag{2.26}$$

Because of symmetry,

$$C = B^H = \underline{\boldsymbol{d}}^H \begin{bmatrix} P_2XP_3^HP3 \\ -P_3XP_2^HP2 \end{bmatrix} \underline{\boldsymbol{v}}_{23} \tag{2.27}$$

$$D = \underline{\boldsymbol{v}}_{23}^H \begin{bmatrix} I_{M\times M} - P_2XP_2^H & -P_2XP_3^H \\ -P_3XP_2^H & I_{M\times M} - P_3XP_3^H \end{bmatrix} \underline{\boldsymbol{v}}_{23} \tag{2.28}$$

Rewrite (2.16) in the form:

$$E = |e_1|^2 = \begin{bmatrix} \underline{\boldsymbol{v}}_1 \\ \underline{\boldsymbol{v}}_3 \end{bmatrix}^H \begin{bmatrix} I_{M\times M} - P_1X^*P_1^H & -P_1X^*P_3^H \\ -P_3X^*P_1^H & I_{M\times M} - P_3X^*P_3^H \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{v}}_1 \\ \underline{\boldsymbol{v}}_3 \end{bmatrix} \tag{2.29}$$

,where $X^* = (\begin{bmatrix} P_1 \\ P_3 \end{bmatrix})^H (\begin{bmatrix} P_1 \\ P_3 \end{bmatrix})^{-1} = (P_1^H P_1 + P_3^H P_3)^{-1}$

Now (2.13) results:

$$P_e = P\{E > A + B + C + D\} \tag{2.30}$$

Using some algebraic manipulations, this simplifies to:

$$Pr\{(\begin{bmatrix} \underline{\boldsymbol{V}}_1 \\ \underline{\boldsymbol{V}}_2 \\ \underline{\boldsymbol{V}}_3 \end{bmatrix} - \begin{bmatrix} \underline{\mathbf{0}} \\ P_2.\underline{\boldsymbol{d}} \\ \underline{\mathbf{0}} \end{bmatrix})^H \begin{bmatrix} -(I - P_1 X^* P_1^H) & 0 & P_1 X^* P_3^H \\ 0 & (I - P_2 X P_2) & P_2 X^* P_3^H \\ P_3 X^* P_1^H & -P_3 X^* P_2^H & P_3(X - X^*) P_3^H \end{bmatrix} .(\begin{bmatrix} \underline{\boldsymbol{V}}_1 \\ \underline{\boldsymbol{V}}_2 \\ \underline{\boldsymbol{V}}_3 \end{bmatrix} - \begin{bmatrix} \underline{\mathbf{0}} \\ P_2.\underline{\boldsymbol{d}} \\ \underline{\mathbf{0}} \end{bmatrix}) < 0\} \tag{2.31}$$

Defining the new matrices and vectors:

$$\mathbf{P} = \begin{bmatrix} -(I - P_1 X^* P_1^H) & 0 & P_1 X^* P_3^H \\ 0 & (I - P_2 X P_2) & -P_2 X P_3^H \\ P_3 X^* P_1^H & -P_3 X P_2^H & -P_3(X - X^*) P_3^H \end{bmatrix}$$

$$\underline{\boldsymbol{B}} = \begin{bmatrix} \underline{\mathbf{0}} \\ P_2.\underline{\boldsymbol{d}} \\ \underline{\mathbf{0}} \end{bmatrix}$$

$$\underline{\boldsymbol{V}} = \begin{bmatrix} \underline{\boldsymbol{V}}_1 \\ \underline{\boldsymbol{V}}_2 \\ \underline{\boldsymbol{V}}_3 \end{bmatrix}$$

we get:

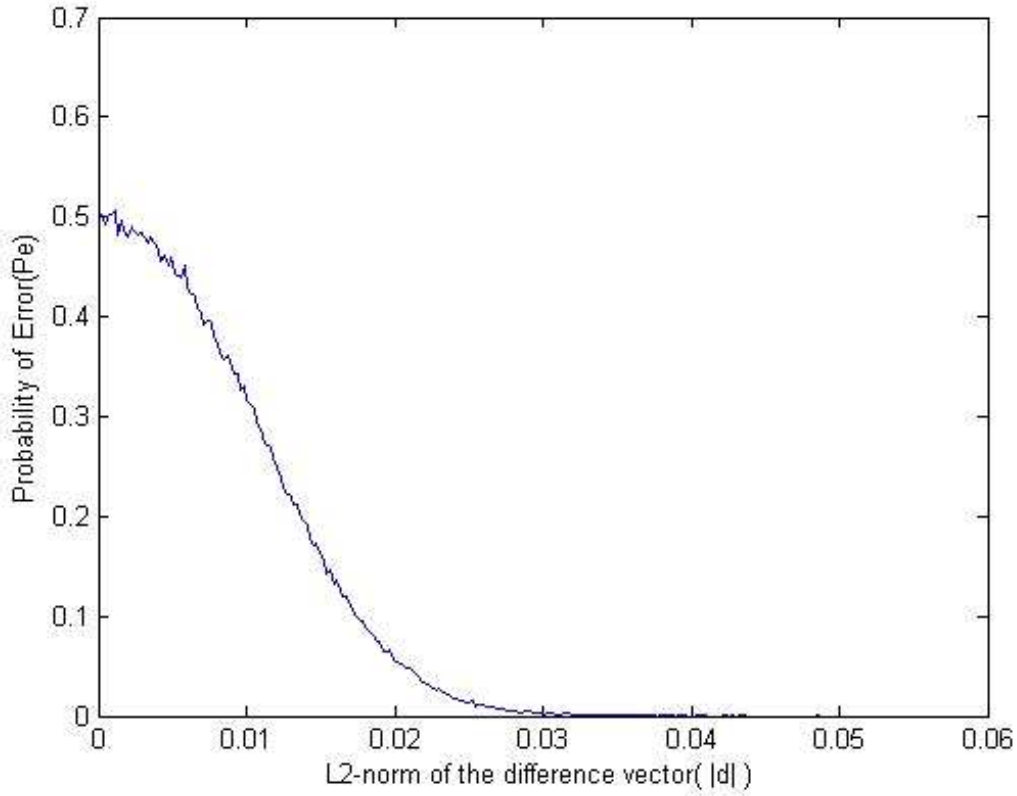

Figure 2.2. Probability of error vs. the norm of the difference vector$\underline{\boldsymbol{d}}$

$$Pr\{(\underline{\boldsymbol{V}}_{(3M\times 1)} + \underline{\boldsymbol{B}})^{H}\mathbf{P}_{(3M\times 3M)}(\underline{\boldsymbol{V}} + \underline{\boldsymbol{B}}) < 0\} \tag{2.32}$$

Figure 2.2 shows the P_e versus the L_2-norm of the difference vector $||\underline{\boldsymbol{d}}||_2$ averaged over 10000 input vectors $\underline{\boldsymbol{X_1}}$,$\underline{\boldsymbol{X_2}}$, and $\underline{\boldsymbol{X_3}}$ of size 100 and also the noise vectors $\underline{\boldsymbol{v_1}}$, $\underline{\boldsymbol{v_2}}$, and $\underline{\boldsymbol{v_3}}$ for the SNR=30.

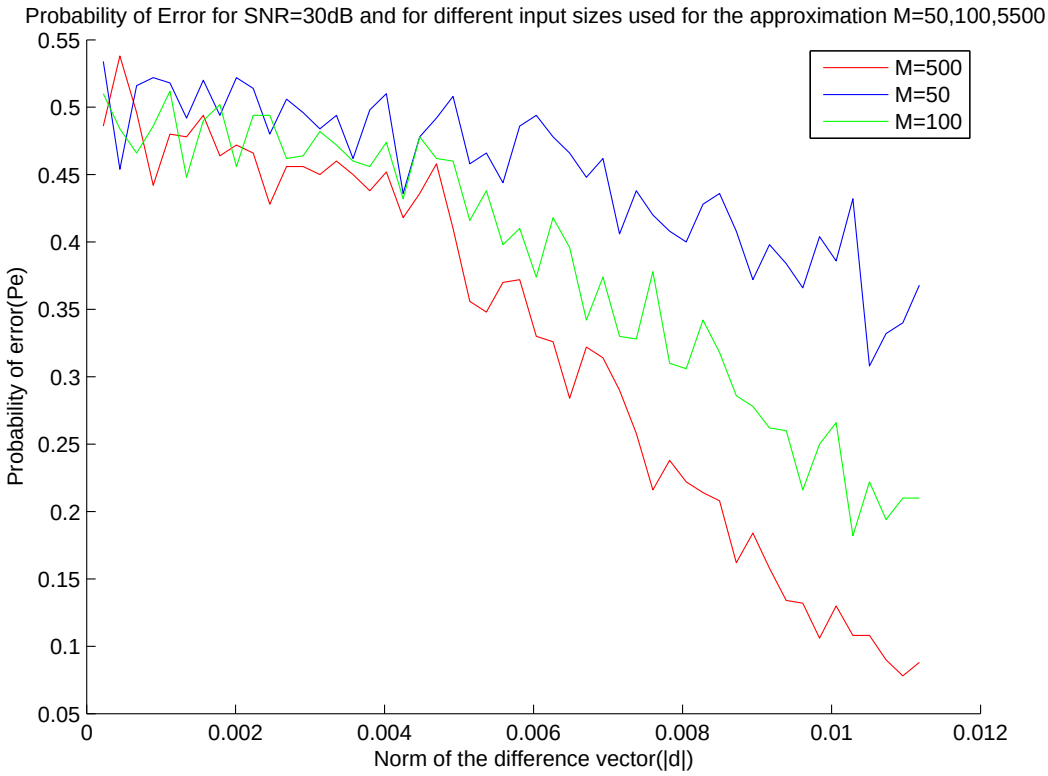

Figure 2.3. Probability of error vs. the norm of the difference vector for different input sizes$\underline{\boldsymbol{d}}$

2.2.3 Supplemental performance curves

We could improve the result by using more input samples in our GLRT detection algorithm. Also, apparently the probability of error decreases as the signal to noise ratio(SNR) is increased. This way for the same amount of the norm of the difference vector($|d|$), we get less probability of error in both cases.

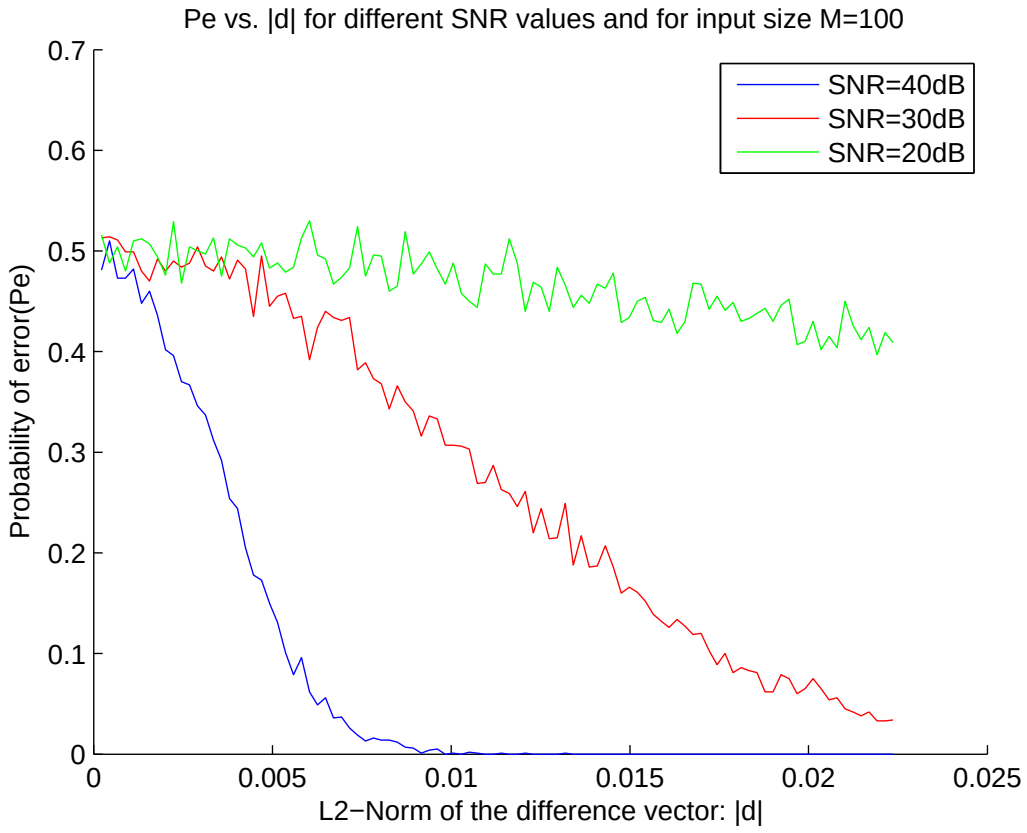

Figure 2.4. Probability of error vs. the norm of the difference vector for different values of SNR$\underline{\boldsymbol{d}}$

Figure 2.3 shows P_e versus norm of the difference vector using different input sizes for the estimation. Figure 2.4 shows P_e versus norm of the difference vector for different values of the SNR.

2.3 A Simplified Algorithm Based on the Classical Likelihood Ratio Test

We have three input/output vector pairs $(\underline{\boldsymbol{X}}_1, \underline{\boldsymbol{Y}}_1)$, $(\underline{\boldsymbol{X}}_2, \underline{\boldsymbol{Y}}_2)$, and $(\underline{\boldsymbol{X}}_3, \underline{\boldsymbol{Y}}_3)$, where the first two pairs of input/output vectors are from nonlinear systems 1 and 2, and we would like to determine which system does the third input/output pair belong to.

2.3.1 Receiver decision rule

It is more practical to store the estimated system coefficients(truncated Volterra series representation of the system)$\underline{\boldsymbol{h_i}}$ $i = 1, 2, 3$ rather than having to store all the input and output data of the first and second transmitters. Then with some suitable distance criterion we should determine if $\underline{\boldsymbol{h_3}}$ is closer to $\underline{\boldsymbol{h_1}}$ or $\underline{\boldsymbol{h_2}}$.

$$\underline{\boldsymbol{Y}}_i = \begin{bmatrix} x_i(n) & x_i(n-1) & x_i^2(n) & x_i^2(n-1) & x_i(n)x_i(n-1) \\ x_i(n-1) & x_i(n-2) & x_i^2(n-1) & x_i^2(n-2) & x_i(n-1)x_i(n-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i(n-M+1) & x_i(n-M) & \cdots & \cdots & x_i(n-M+1)x_i(n-M) \end{bmatrix} \underline{\boldsymbol{h}}_i + \underline{\boldsymbol{e}}_i$$

or equivalently:

$$\underline{\boldsymbol{Y}}_i = P_i \underline{\boldsymbol{h}}_i + \underline{\boldsymbol{e}}_i \quad ; \; i = 1, 2.$$

where $\underline{\boldsymbol{e}}_i$: $\quad i = 1, 2, 3.$ is the estimation error.

The $M \times N$ matrices $P_i, \quad i = 1, 2, 3$ can be determined easily from the vectors $\underline{\boldsymbol{X}}_i, \ \ i = 1, 2, 3.$ A standard metric is to minimize the squared error:

$$|\underline{\boldsymbol{e}}_i|^2 = |\underline{\boldsymbol{Y}}_i - P_i \underline{\boldsymbol{h}}_i|^2, \ \ i = 1, 2, 3.$$

This is the classical Least-squares(LS) problem and the solution is:

$$\underline{\boldsymbol{h}}_{i,opt} = P_i^{\dagger} \underline{\boldsymbol{Y}}_i, \quad i = 1, 2, 3.$$

$$\underline{\boldsymbol{h}}_{i,opt} = (P_i^H P_i)^{-1} P_i^H \underline{\boldsymbol{Y}}_i, \quad i = 1, 2, 3.$$

Now that we have the estimated $\underline{\boldsymbol{h}}_i, \quad i = 1, 2, 3.$ we should solve the classical decision problem using the likelihood ratio:

$$\Lambda(\underline{\boldsymbol{h}}_3) \triangleq \frac{P_{\underline{\boldsymbol{h}}_3|\underline{\boldsymbol{h}}_1}(\underline{\boldsymbol{h}}_3|\underline{\boldsymbol{h}}_1)}{P_{\underline{\boldsymbol{h}}_3|\underline{\boldsymbol{h}}_2}(\underline{\boldsymbol{h}}_3|\underline{\boldsymbol{h}}_2)}$$

On the other hand, given $\underline{\boldsymbol{h}_i}$,

$$\underline{\boldsymbol{h}}_3 = (P_3^H P_3)^{-1} P_3^H \underline{\boldsymbol{Y}}_i, \quad i = 1, 2.$$

$$\underline{\boldsymbol{h}}_3 = (P_3^H P_3)^{-1} P_3^H P_3 \underline{\boldsymbol{h}}_i + (P_3^H P_3)^{-1} P_3^H \underline{\boldsymbol{\nu}}_3 = \underline{\boldsymbol{h}}_i + (P_3^H P_3)^{-1} P_3^H \underline{\boldsymbol{\nu}}_3, \quad i = 1, 2.$$

If $\underline{\boldsymbol{X}}$ is a Gaussian random vector, then so is $A\underline{\boldsymbol{X}} + \underline{\boldsymbol{b}}$for any $r \times n$ vector A and any r-vector b. Symbolically, we write:

$$X \sim N(\underline{\boldsymbol{m}}, C) \quad \Rightarrow \quad A\underline{\boldsymbol{X}} + \underline{\boldsymbol{b}} \sim N(A\underline{\boldsymbol{m}} + \underline{\boldsymbol{b}}, ACA^H)$$

then:

$$P_{\underline{\boldsymbol{h}}_3|\underline{\boldsymbol{h}}_i}(\underline{\boldsymbol{h}}_3|\underline{\boldsymbol{h}}_i) = \frac{1}{\sqrt{det(C)}(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(\underline{\boldsymbol{h}}_3 - \underline{\boldsymbol{h}}_i)^H C^{-1} (\underline{\boldsymbol{h}}_3 - \underline{\boldsymbol{h}}_i)} \tag{2.33}$$

where C is the covariance matrix and can simply be derived as:

$$C = ((P_3^H P_3)^{-1} P_3^H) * \sigma_n^2 I_{M\times M} * ((P_3^H P_3)^{-1} P_3^H)^H$$

$$C = \sigma_n^2 (P_3^H P_3)^{-1}$$

Equation (2.33) yields:

$$P_{\underline{\boldsymbol{h}}_3|\underline{\boldsymbol{h}}_i}(\underline{\boldsymbol{h}}_3|\underline{\boldsymbol{h}}_i) = \frac{1}{\sqrt{det(C)}(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma_n^2}(\underline{\boldsymbol{h}}_3 - \underline{\boldsymbol{h}}_i)^H (P_3^H P_3)(\underline{\boldsymbol{h}}_3 - \underline{\boldsymbol{h}}_i)} \tag{2.34}$$

At the receiver the estimated Volterra coefficients $\underline{\boldsymbol{h}}_1$ and $\underline{\boldsymbol{h}}_2$, are stored and every time an output vector $\underline{\boldsymbol{Y}}_3$ is received, we decode the corresponding $\underline{\boldsymbol{X}}_3$ and estimate $\underline{\boldsymbol{h}}_3$ using the results of the LS problem. Then we calculate $\Lambda(\underline{\boldsymbol{h}}_3)$ and decide whether it was $\underline{\boldsymbol{h}}_1$ or $\underline{\boldsymbol{h}}_2$:

$$e^{-\frac{1}{2\sigma_n^2}(\underline{\boldsymbol{h}}_3 - \underline{\boldsymbol{h}}_1)^H (P_3^H P_3)(\underline{\boldsymbol{h}}_3 - \underline{\boldsymbol{h}}_1)} \underset{\underline{\boldsymbol{h_2}}}{\overset{\underline{\boldsymbol{h_1}}}{\gtrless}} e^{-\frac{1}{2\sigma_n^2}(\underline{\boldsymbol{h}}_3 - \underline{\boldsymbol{h}}_2)^H (P_3^H P_3)(\underline{\boldsymbol{h}}_3 - \underline{\boldsymbol{h}}_2)}$$

Equivalently:

$$(\underline{\boldsymbol{h}}_3 - \underline{\boldsymbol{h}}_1)^H (P_3^H P_3)(\underline{\boldsymbol{h}}_3 - \underline{\boldsymbol{h}}_1) \underset{\underline{\boldsymbol{h_1}}}{\overset{\underline{\boldsymbol{h_2}}}{\gtrless}} (\underline{\boldsymbol{h}}_3 - \underline{\boldsymbol{h}}_2)^H (P_3^H P_3)(\underline{\boldsymbol{h}}_3 - \underline{\boldsymbol{h}}_2)$$

Figure 2.5 shows the simulated probability of error versus the norm of the difference vector for the classical likelihood ratio test algorithm. The GLRT performance curve is also included for comparison reasons.

2.4 Volterra coefficients as random variables

In this section we consider the Volterra coefficients to be random variables with variance σ_h^2 around their mean and run simulations to find the curves of P_e vs. the variance of the coefficients(σ_h^2).

We also consider another algorithm which we call the naive algorithm in which our naive detection system outputs the estimated coefficients vector $h_{opt,i}$, i=1,2 which is closest to the third estimated coefficients vector $h_{opt,3}$ with the L_2-norm criterion for measuring the closeness:

$$||\underline{\boldsymbol{h}}_{opt,3} - \underline{\boldsymbol{h}}_{opt,2}|| \underset{\underline{\boldsymbol{h_1}}}{\overset{\underline{\boldsymbol{h_2}}}{\gtrless}} ||\underline{\boldsymbol{h}}_{opt,3} - \underline{\boldsymbol{h}}_{opt,1}||$$

Figure 2.6 shows the performance of the GLRT algorithm, the classical likelihood ratio algorithm and the naive algorithm in one figure. The classical likelihood ratio test has the

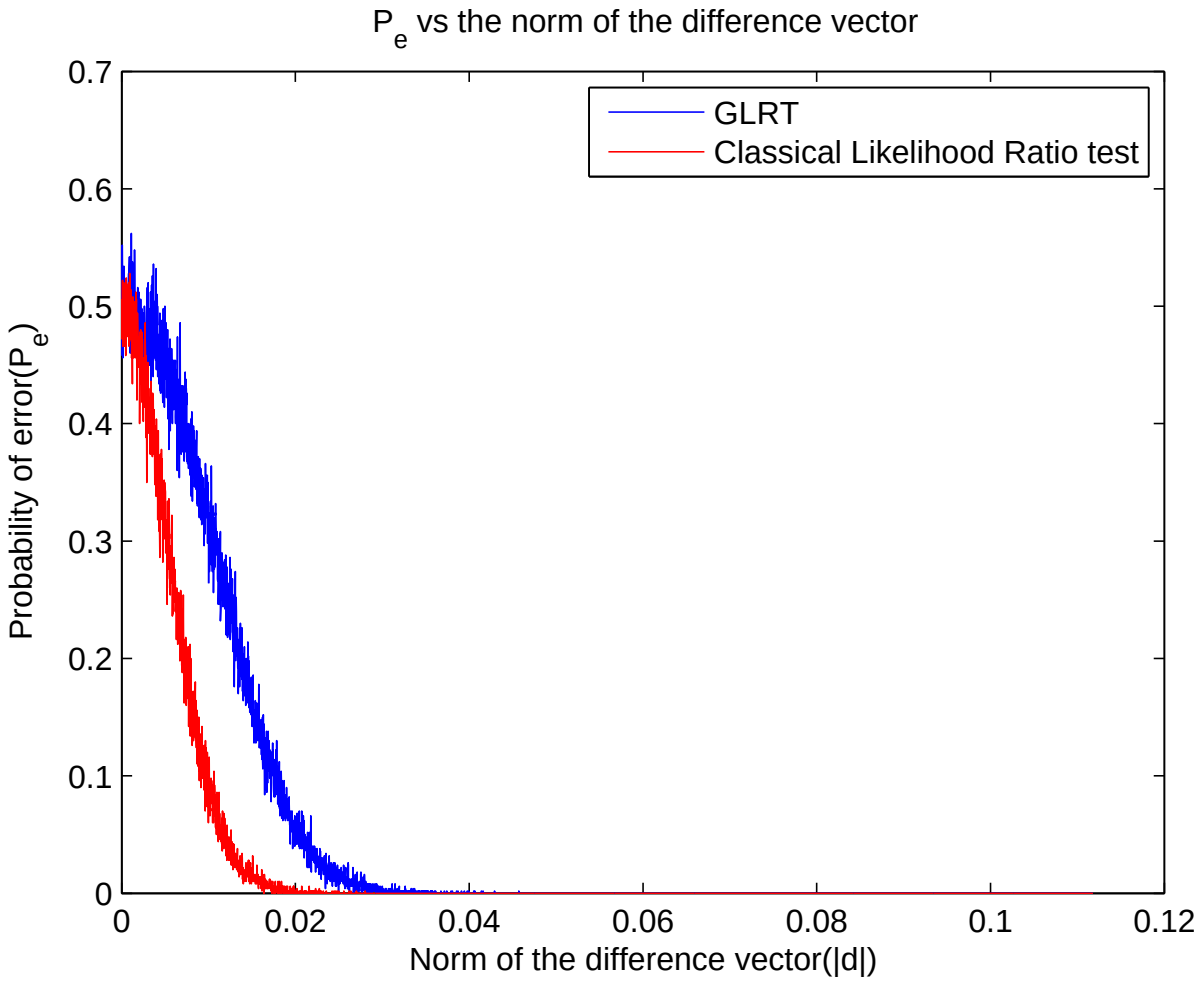

Figure 2.5. Probability of error vs. the norm of the difference vector for SNR = 30dB

$|\underline{d}|$

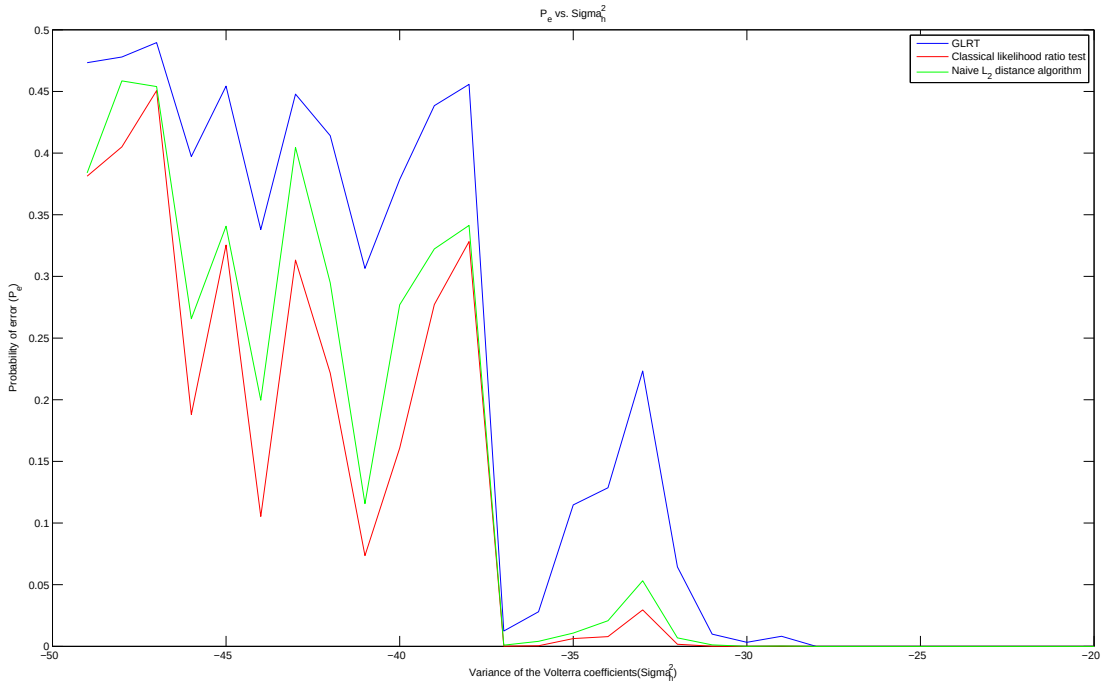

Figure 2.6. Probability of error vs. the variance of the Volterra coefficients σ_h^2

best simulated results, while the GLRT has the greatest probability of error of the three. The GLRT makes use of the whole data all at once , while in the other methods the Volterra coefficients vector is determined first as described above.

CHAPTER 3

MODELING OF POWER AMPLIFIERS

In this chapter we will validate the use of Volterra series representation for the study of nonlinear system components including power amplifiers and the possible sources of variations in the Volterra series vector of different amplifiers.This way we will provide the link between the theoretic analysis of the previous chapters and the practical anonymity breaking application.

Behavioral modeling techniques provide a convenient and efficient means to predict system-level performance without the computational complexity of full circuit simulation or physics-level analysis of nonlinear systems, thereby significantly speeding up the analysis process. General Volterra series based models have been successfully applied for radio frequency (RF) power amplifier (PA) behavioral modeling. Working with Volterra series presentations provides RF circuit designers with the insight that enables them to trace the defects in their designs and modify the circuit parameters or the circuit elements. Many instances of the use of Volterra series can be found in the literature for modeling and distortion calculation of nonlinear devices([42, 43, 40, 41, 47, 48, 49, 50, 51, 52, 53]), and also studying the response of nonlinear systems to noise as in [44, 45, 46].

Weakly nonlinear systems are systems whose response to external inputs can be described by a few terms in a functional series expansion such as a Volterra series. On the

other hand, excessively nonlinear systems like the class D power amplifiers have transfer functions that cannot be well characterized by low order Volterra functional series.To model the PA's with strong nonlinearities and long memory effects, the general Volterra model involves a great number of coefficients. In this respect, some simplified Volterra series based models for RF power amplifiers have been proposed.

For our purpose of identification, if we consider no complexity limits, either weakly or excessively nonlinear amplifiers like the class D amplifiers used in cellular phone transmitters, could be characterized by enough Volterra series coefficients. It is important how well Volterra series characterize amplifiers, as any model mismatch will affect the performance of the the algorithm presented in previous chapter as it is based on the assumption that a system is characterized by Volterra series representation. Fortunately the broad use of Volterra series for modeling and studying amplifiers in the circuit community proves it a suitable representation.

Next we need to determine how much variation there is between different amplifiers of the same type. This will determine how well we are able to distinguish amplifiers according to the P_e versus the norm of the difference vector curves derived.

3.1 Process Variations

To be able to validate the effectiveness of our anonymity breaking techniques, we need to determine how the Volterra series presentations of power amplifiers, even though from the same manufacturer, differ in practice by introducing one source of variation. Variations in fabrication process, ambient temperature and supply voltage affect the electrical performance of the transistors. For example a higher temperature and a lower supply volt-

age make the transistor operate slower.See Appendix B for a brief explanation of process corners.

The first important source that makes different amplifiers in different transmitters have distinct Volterra series coefficients which we could make use of to distinguish them is the parameter variations in production. According to [54], MOS transistors of which the amplifier circuits are made always exhibit broad variations in major device parameters among production lots. As a result, a wide range of devices are measured and parameters are extracted to characterize the statistical variations. The most notable parameter variations include channel length, threshold voltage, and gate-oxide thickness variations. Additional models are added to the model library based on the extremes of these key parameters. These models are called process corners in that they capture parameters that make the circuit unusually fast or unusually slow.

To show the variation of the Volterra series coefficients at these process corners, a simple one-transistor class A amplifier was simulated. Given a sinusoidal input at the frequency ω_0 the output of the circuit was measured at frequencies ω_0, $2\omega_0$, $3\omega_0$, and $4\omega_0$. From this data considering the simplified memoryless fourth order Volterra series representation, we derived these coefficients for the slow(worst performance, lowest power),nominal (typical performance, typical power), and fast (best performance, highest power) NMOS transistor. Figure 3.1 shows the simple amplifier simulated and the Tables 3.1, 3.2, and 3.3 show the output at frequencies ω_0, $2\omega_0$, $3\omega_0$, and $4\omega_0$ for the three typical, slow, and fast corners of operation of the NMOS transistor in the amplifier circuit. In each case the available source power P_{avs} is determined so that the optimum power is delivered to the load. Corresponding estimated Volterra series coefficients are derived in Table 3.4. Now

we can estimate the norm of the difference vector between the fast and slow corners which are the $\pm 3\sigma$ of the limits of the bell curve of the probability density function from its mean. The resulting L_2-norm of the difference is 0.0390 which probability of error quite close to zero. according to the P_e curves of chapter 2, we have a very low probability of error for this norm of the difference vector.

Table 3.1. Typical NMOS corner

P_{avs}	freq	**Typical NMOS**	
Pdel_opt=0.23W		**Pdel_opt @Pavs=9.95dBm**	
P_{avs}	freq	mag(vload)[40, ::]	phase(vload)[40, ::]
10.00	1.90E+09	4.8073	-54.778
10.00	3.80E+09	0.0789	-105.760
10.00	5.70E+09	0.0053	82.483
10.00	7.60E+09	0.0010	101.338

Table 3.2. Slow NMOS corner

Pavs	freq	**Slow NMOS**	
Pdel_opt=0.23W		**Pdel_opt @Pavs=10.653dBm**	
Pavs	freq	mag(vload)[43, ::]	phase(vload)[43, ::]
10.75	1.90E+09	4.8215	-55.841
10.75	3.80E+09	0.0717	-105.394
10.75	5.70E+09	0.0099	75.561
10.75	7.60E+09	0.0023	94.778

In these simulations the temperature and supply voltage effects are not considered. also, in practice class AB amplifiers are used in power amplifier circuits which are even less nonlinear but more efficient which makes our study of class A amplifiers a worst case analysis and results in a better probability of error.

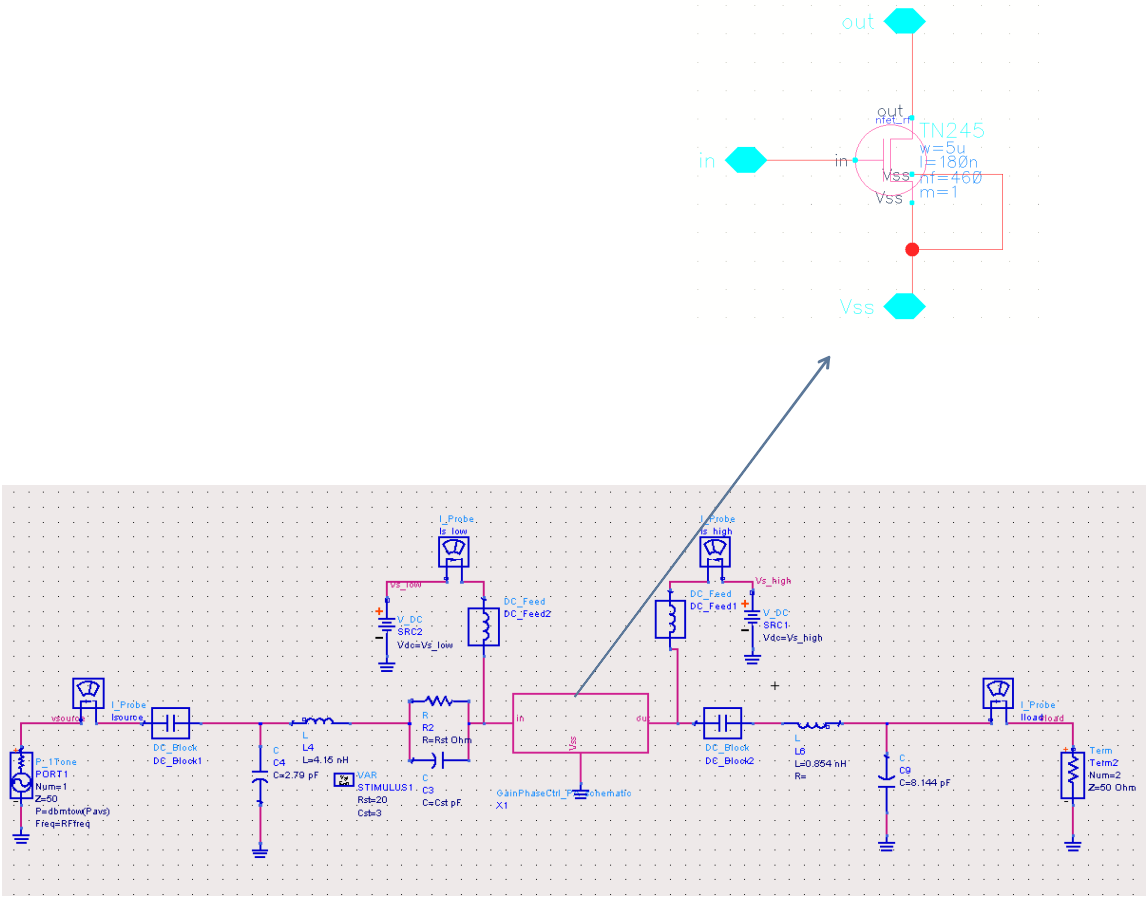

Figure 3.1. Class A amplifier circuit.

Table 3.3. Fast NMOS corner

Pavs	freq	**Fast NMOS**	
Pdel_opt=0.23W		**Pdel_opt @9.246**	
Pavs	freq	mag(vload)[37, ::]	phase(vload)[37, ::]
9.25	1.90E+09	4.7954	-53.298
9.25	3.80E+09	0.0771	-103.505
9.25	5.70E+09	0.0038	91.811
9.25	7.60E+09	0.0004	108.759

Table 3.4. Estimated Volterra Series coefficients

	Slow	**Typical**	**Fast**
h_1	4.791829	4.79136	4.783927
h_2	0.125176	0.149879	0.150715
h_3	0.039517	0.02124	0.015336
h_4	0.01832	0.007841	0.003528

Although the difference vector is computed between the $\pm 3\sigma$ limits which is between the extremes. In practice two devices can vary with some certain probability that can be calculated from their bell curve.

CHAPTER 4

BREAKING ANONYMITY IN THE LOW SNR CASE

In the previous chapters we have considered the high SNR case where the input was decoded at the output and the anonymity breaking algorithms discussed assumed having access to the input vector as well. Possible future work could be done in suggesting anonymity breaking techniques and algorithms for the low SNR case. Here in this chapter we only show from an information theoretic perspective that the output of a nonlinear system conveys information about the system Volterra coefficients and thus hardware based techniques could be devised for anonymity breaking purposes.

We will provide information theoretic formulas and bounds to the performance of the physical layer anonymity breaking techniques explained in Chapter 1. Because of the nonlinear nature of the formulas in this chapter, the multiple integrals in the formulas rarely lead to neat formulations even with the simplest possible assumptions. But, for the simplified case of the zero memory linear quadratic case and for the low signal to noise ratio case, the closed form formulas are calculated.

4.1 Zero-memory Linear Quadratic System

Using the simplified Volterra series which only consists of the first convolution term(A.9)which is the linear part and the second double summations(A.10), the quadratic part, what results

is called a linear quadratic form. In addition, we consider the zero memory case where the Volterra series simply looks like Taylor series around point zero.

$$y(n) = h_1(0)x(n) + h_2(0,0)x(n)^2 + \nu(n) \tag{4.1}$$

What we want to find is the the amount of (Shannon)information the output of the nonlinear system has about the system parameters, which is the mutual information of the output and system coefficients which in this case are $h_1(0)$ and $h_2(0,0)$.Call the coefficients vector $\underline{\boldsymbol{h}} = [h_1(0)\, h_2(0,0)]^T$. What we are actually interested in is how the mutual information increases when we have access to more output points. Let's consider the information of one output about the system parameters, I(y(n);$\underline{\text{h}}$):

$$I(y(n); \underline{\boldsymbol{h}}) = h(y(n)) - h(y(n)|\underline{\boldsymbol{h}}) \tag{4.2}$$

$$I(y(n); \underline{\boldsymbol{h}}) = h(\underline{\boldsymbol{h}}) - h(\underline{\boldsymbol{h}}|y(n)) \tag{4.3}$$

We use equation (4.2) to calculate the mutual information as it is easier to deal with in our case.

$$h(y(n)|\underline{\boldsymbol{h}}) = E_{\underline{\boldsymbol{h}}}(h(y(n)|\underline{\boldsymbol{h}} = \underline{H}) \tag{4.4}$$

where

$$h(y(n)|\underline{\boldsymbol{h}} = \underline{H}) = \int_{-\infty}^{+\infty} f_{y(n)|\underline{\boldsymbol{h}}=\underline{H}}(y(n)|\underline{\boldsymbol{h}} = \underline{H}) log(f_{y(n)|\underline{\boldsymbol{h}}=\underline{H}}(y(n)|\underline{\boldsymbol{h}} = \underline{H}))dy(n)$$

and $E_{\underline{\boldsymbol{h}}}$ means expectation with respect to $\underline{\boldsymbol{h}}$.

$$h(y(n)) = \int_{-\infty}^{+\infty} f_{y(n)}(y(n)) log(f_{y(n)}(y(n)))dy(n) \tag{4.5}$$

Thus, to be able to find $h(y(n))$ and $h(y(n)|\underline{\boldsymbol{h}})$ we first need to find $f_{y(n)}(y(n))$ and $f_{y(n)|\underline{\boldsymbol{h}}}$ which are the probability density functions of $y(n)$ and $y(n)$ given $\underline{\boldsymbol{h}}$.

The input $x(n)$ and the system coefficients are random variables. The input is a binary random variable with equal probability for -1 and 1, and the system coefficients and the additive noise are Gaussian random variables with mean zero and variances $\sigma_{h_1}^2$, $\sigma_{h_2}^2$, and σ_n^2 respectively. $y(n)$ given $\underline{\boldsymbol{h}}$ and $x(n)$ is a Gaussian random variable with mean $h_1(0)x(n) + h_2(0,0)x(n)^2$ and variance σ_n^2. Given $x(n)$, $\underline{\boldsymbol{h}}$

$$y(n) \sim N(h_1(0)x(n) + h_2(0,0)x(n)^2, \sigma_n^2)$$

Or equivalently:

$$f_{y(n)|\underline{\boldsymbol{h}},x(n)}(y(n)|\underline{\boldsymbol{h}}, x(n)) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(y(n)-h_1(0)x(n)-h_2(0,0)x(n)^2)^2}{2\sigma_n^2}} \tag{4.6}$$

and

$$f_{y(n)|\underline{\boldsymbol{h}}}(y(n)|\underline{\boldsymbol{h}}) = E_{x(n)}\left(\frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(y(n)-h_1(0)x(n)-h_2(0,0)x(n)^2)^2}{2\sigma_n^2}}\right) \tag{4.7}$$

$$f_{y(n)}(y(n)) = E_{\underline{\boldsymbol{h}}}\left(E_{x(n)}\left(\frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(y(n)-h_1(0)x(n)-h_2(0,0)x(n)^2)^2}{2\sigma_n^2}}\right)\right) \tag{4.8}$$

4.2 Low SNR approximation

In this section we use the results of [1]to find the mutual information for the case of binary inputs. The author in [1] finds the capacity of the binary-input additive gaussian noise channel:

$$Y = X + \nu \tag{4.9}$$

The capacity is the maximum amount of mutual information for different input probability mass functions which is when the inputs -1 and 1 are equiprobable. For the low SNR case, [1] finds:

$$I(X;Y) = h(Y) - h(Y|X) \tag{4.10}$$

Since H(Y) in [1] is the same as $h(y(n)|\underline{h} = \underline{H})$ for our problem, we can also easily find $h(Y|X)$:

$$h(y(n)|\underline{h} = \underline{H}) = h(Y) = I(X;Y) + h(Y|X) = \tag{4.11}$$

$$= log_2(e)\frac{h_1^2}{2\sigma_n^2} + o(\frac{h_1^3}{\sigma_n^3}) + E_X(h(Y|X=x)) = log_2(e)\frac{h_1^2}{2\sigma_n^2} + E_X(log_2(e) + log_2(\sqrt{2\pi}\sigma_n)) \tag{4.12}$$

$$= log_2(e)\frac{h_1^2}{2\sigma_n^2} + \frac{1}{2}log_2(e) + log_2(\sqrt{2\pi}\sigma_n) \tag{4.13}$$

In our problem we have $Y = h_1(0)x(n) + h_2(0,0)x(n)^2 + \nu$ where $x(n)^2 = 1$. As $Y = h_1(0)(x(n) + \frac{h_2(0,0)}{h_1(0)}) + \nu$ is the scaled shifted version of $Y = X + \nu$, and also as shifting does not change the amount of information, we just need to take into account the scaling factor which shows itself as a h_1^2 coefficient in $I(X;Y)$ and does not change $h(Y|X)$.

Therefore,

$$h(y(n)|\underline{\boldsymbol{h}}) = E_{\underline{\boldsymbol{h}}}(h(y(n)|\underline{\boldsymbol{h}} = \underline{H})) = log_2(e)\frac{\sigma_1^2}{2\sigma_n^2} + \frac{1}{2}log_2(e) + log_2(\sqrt{2\pi}\sigma_n) \tag{4.14}$$

According to (4.1), given $x(n)$

$$y(n) \sim N(0, \sigma_n^2 + \sigma_1^2 x_n^2 + \sigma_2^2) \tag{4.15}$$

and as $x(n)$ shows itself only in the form of $x(n)^2 = 1$. Thus $y(n)$ has the same pdf, and the entropy for the normal random variables is simply:

$$h(y(n)) = log_2(\sqrt{2\pi(\sigma_n^2 + \sigma_1^2 + \sigma_2^2)}) + \frac{1}{2}log_2(e) \tag{4.16}$$

and finally the mutual information of the output and the zero memory linear quadratic system parameters equals:

$$I(y(n); \underline{\boldsymbol{h}}) = h(y(n)) - h(y(n)|\underline{\boldsymbol{h}}) = log_2(\frac{\sqrt{\sigma_n^2 + \sigma_1^2 + \sigma_2^2}}{\sigma_n}) - log_2(e)\frac{\sigma_1^2}{2\sigma_n^2} \tag{4.17}$$

In the equation 4.18 if we consider $\sigma_1^2 = \sigma_2^2$ and also if we call $\frac{\sigma_1^2}{\sigma_n^2} = x$, then we will have:

$$I(y(n); \underline{\boldsymbol{h}}) = \frac{1}{2}log_2(\frac{1 + 2x}{e^x}) \tag{4.18}$$

For the low SNR case we have a lot of noise or equivalently large σ_n^2 and thus $0 < x < 1$ which results in $I(y(n); \underline{\boldsymbol{h}}) > 0$ which means the output has some non-zero information about the system Volterra coefficients.

Now if we increase the number of outputs and form the vector of outputs:

$$\underline{\boldsymbol{Y}} = [y(n)\ y(n-1)\ ...\ y(n-M+1)]^T$$

where

$$y(n-i) = h_1(0)x(n-i) + h_2(0,0)^2x(n-i)^2 + \nu\ ,\ i = 0, ..., M-1$$

and the vector of system parameters $\underline{h} = [h_1(0)\ h_2(0,0)]$ Now as we considered the system parameters known(deterministic) for finding $h(y(n)|\underline{h} = \underline{H})$, if we increase the number of outputs, still $y(n-i)$ given $\underline{h}$ are independent and thus:

$$h(\underline{\boldsymbol{Y}}|\underline{\boldsymbol{h}}) = E_{\underline{\boldsymbol{h}}}(h(\underline{\boldsymbol{Y}}|\underline{\boldsymbol{h}} = \underline{\boldsymbol{H}})) = M(log_2(e)\frac{\sigma 1^2}{2\sigma^2} + \frac{1}{2}log_2(e) + log_2(\sqrt{2\pi}\sigma_n)) \quad (4.19)$$

But for finding $I(\underline{\boldsymbol{Y}}; \underline{h})$ we still need to find h($\underline{\boldsymbol{Y}}$) which is not as easy as $y(n-i)$'s are not independent.

4.3 The General Case

The general input output relationship for a nonlinear system with limited memory L and only considering the terms until and including the N-fold summations is of the form:

$$y(n) = \Sigma_{r=1}^{N}\Sigma_{k_1=0}^{L-1}...\Sigma_{k_r=0}^{L-1}h_r(k_1,...,k_r)x(n-k_1)...x(n-k_r) \quad (4.20)$$

$$= \Sigma_{k_1=0}^{L-1}h_1(k_1)x(n-k_1) \quad (4.21)$$

$$+\Sigma_{k_1=0}^{L-1}\Sigma_{k_2=0}^{L-1}h_2(k_1,k_2)x(n-k_1)x(n-k_2) \quad (4.22)$$

$$+...$$

$$+\Sigma_{k_1=0}^{L-1}...\Sigma_{k_N=0}^{L-1}h_N(k_1,...,k_N)x(n-k_1)...x(n-k_N), \quad (4.23)$$

and also after observing M output samples $\underline{\boldsymbol{Y}} = [y_n\ y_{n-1}\ ...\ y_{n-2}]^T$, we will have:

$$f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{h}},\underline{\boldsymbol{X}}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{h}},\underline{\boldsymbol{x}}) = \frac{1}{(\sqrt{2\pi}\sigma_n)^M} e^{\Sigma_{i=0}^{M-1} \frac{(y(n-i)-h_1(0)x(n-i)-h_2(0,0)x(n-i)^2)^2}{2\sigma_n^2}} \tag{4.24}$$

This is because given $x(n-i)$'s and $\underline{\boldsymbol{h}}$ $y(n-i)$'s are i.i.d. Gaussian.

$$f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{h}}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{h}}) = E_{\underline{\boldsymbol{x}}}\left(\frac{1}{(\sqrt{2\pi}\sigma_n)^M} e^{\Sigma_{i=0}^{M-1} \frac{(y(n-i)-h_1(0)x(n-i)-h_2(0,0)x(n-i)^2)^2}{2\sigma_n^2}}\right) \tag{4.25}$$

$$f_{\underline{\boldsymbol{Y}}}(\underline{\boldsymbol{y}}) = E_{\underline{\boldsymbol{h}}}\left(E_{\underline{\boldsymbol{x}}}\left(\frac{1}{(\sqrt{2\pi}\sigma_n)^M} e^{\Sigma_{i=0}^{M-1} \frac{(y(n-i)-h_1(0)x(n-i)-h_2(0,0)x(n-i)^2)^2}{2\sigma_n^2}}\right)\right) \tag{4.26}$$

$$h(\underline{\boldsymbol{Y}}) = \int\int ... \int_{-\infty}^{+\infty} f_{\underline{\boldsymbol{Y}}}(\underline{\boldsymbol{y}}) log_2(f_{\underline{\boldsymbol{Y}}}(\underline{\boldsymbol{y}}))d\underline{\boldsymbol{y}} \tag{4.27}$$

$$h(\underline{\boldsymbol{y}}|\underline{\boldsymbol{h}} = \underline{H}) = \int\int ... \int_{-\infty}^{+\infty} f_{\underline{\boldsymbol{Y}}|\underline{\boldsymbol{h}}=\underline{H}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{h}} = \underline{H}) log(f_{\underline{\boldsymbol{y}}|\underline{\boldsymbol{h}}=\underline{H}}(\underline{\boldsymbol{y}}|\underline{\boldsymbol{h}} = \underline{H}))d\underline{\boldsymbol{y}} \tag{4.28}$$

$$h(y(n)|\underline{\boldsymbol{h}}) = E_{\underline{\boldsymbol{h}}}(h(y(n)|\underline{\boldsymbol{h}} = \underline{H}) \tag{4.29}$$

To find neat formulations or bounds to the above equations we need to solve multiple integrals of the function (4.25) or (4.24) which are fairly complicated functions the integrals of which do not have closed form solutions.

CHAPTER 5

CONCLUSION

In this thesis, with the motivation of digital forensics and breaking anonymity, an approach to identify users is presented. The main idea is to make use of the minute imperfections in the different components of the transmitters' hardware even for the case where they are made by the same manufacturer.

First the feasibility and effectiveness of this approach is shown empirically by measuring the output frequency of some oscillators from the same manufacturer.We saw that although the two oscillators were low noise oscillators from the same manufacturer they had different center frequencies and this difference could be deployed along with other parameters to identify different users.

Next,for the high signal to noise ratio(SNR) case, where the input is fully recovered at the receiver, two algorithms based on the generalized likelihood ratio test(GLRT) and classical likelihood ratio test were proposed and the effectiveness of these algorithms was shown by drawing the P_e versus the norm of the difference vector between the two sets of system parameters. The nonlinear systems are presented by volterra series coefficients, the appropriateness of which was addressed in the modeling chapter. Also one source of the difference in Volterra series coefficients of different transmitters which is the process variations and how it causes variations in the Volterra series representation of different

power amplifiers was studied. In practice class AB amplifiers are used in the transmitter systems which are more nonlinear than the class A amplifier studied.

Finally, for the low SNR case, only the fact that some certain amount of information about the system coefficients from the output could be derived, was proved, leaving the task of devising new algorithms capable of this to future researchers.

Future researches could be done by improvising new separation techniques for the low SNR case. Also the apparent generalization of the introduced algorithms for the high SNR case to the scenario where there are more than two possible users and the resulting performance curves is another possible future research topic. Also, the possible countermeasures and studying the susceptibility of our techniques to these possible countermeasures could be done as another complementary research project.

APPENDIX A

VOLTERRA SERIES REPRESENTATION

A linear, causal system with memory can be described by the convolution representation:

$$y(t) = \int_{-\infty}^{+\infty} h(\tau)x(t-\tau)d\tau \tag{A.1}$$

$$y(n) = \Sigma_{m=-\infty}^{\infty} h(m)x(n-m) \tag{A.2}$$

where $x(t)$ is the input, $y(t)$ the output, and $h(t)$ the impulse response of the system. A nonlinear system without memory can be described with a Taylor series:

$$y(t) = \Sigma_{n=1}^{+\infty} a_n[x(t)]^n \tag{A.3}$$

$$y(n) = \Sigma_{n=1}^{+\infty} a_n[x(n)]^n \tag{A.4}$$

where, again, $x(t)$ is the input and $y(t)$ the output. The a_n are the Taylor series coefficients. A Volterra series combines the above two representations to describe a nonlinear system with memory:

$$y(t) = \Sigma_{n=1}^{\infty} \frac{1}{n!} \int_{-\infty}^{+\infty} du_1 ... \int_{-\infty}^{+\infty} du_n g_n(u_1, ..., u_n) \prod_{r=1}^{n} x(t-u_r) \tag{A.5}$$

$$= \frac{1}{1!}\int_{-\infty}^{+\infty} du_1 g_1(u_1)x(t-u_1) \tag{A.6}$$

$$+\frac{1}{2!}\int_{-\infty}^{+\infty} du_1 \int_{-\infty}^{+\infty} du_2 g_2(u_1,u_2)x(t-u_1)x(t-u_2) \tag{A.7}$$

$$+\frac{1}{3!}\int_{-\infty}^{+\infty} du_1 \int_{-\infty}^{+\infty} du_2 \int_{-\infty}^{+\infty} du_3 g_3(u_1,u_2,u_3)x(t-u_1)x(t-u_2)x(t-u_3) \tag{A.8}$$

$$+...$$

$x(t)$ is the input, $y(t)$ is the output, and the $g_n(u_1,...,u_n)$ are called the Volterra kernels of the system or simply the kernels . The u_i are time variables. For $n = 1$, $g_1(u_1)$ will be recognized as the familiar impulse response (A.1); thus, g_n for $n > 1$ are rather like "higher-order impulse responses". These serve to characterize the various orders of nonlinearity . The first few terms of (A.5) have been explicitly written out; (A.6) is the familiar convolution integral (A.1), and (A.7) and (A.8) may be thought of as two-fold and three-fold convolution. (A.5) is an infinite sum of n-fold convolution integrals. The leading $\frac{1}{n!}$ is omitted by most authors.For causal systems the lower bound of all integrals equals zero. Equation (A.9) shows the discrete-time version of (A.5),which we use throughout this thesis:

$$y(n) = \Sigma_{r=1}^{\infty}\Sigma_{k_1=-\infty}^{\infty}...\Sigma_{k_r=-\infty}^{\infty} h_r(k_1,...,k_r)x(n-k_1)...x(n-k_r) \tag{A.9}$$

$$= \Sigma_{k_1=-\infty}^{\infty} h_1(k_1)x(n-k_1) \tag{A.10}$$

$$+\Sigma_{k_1=-\infty}^{\infty}\Sigma_{k_2=-\infty}^{\infty} h_2(k_1,k_2)x(n-k_1)x(n-k_2) \tag{A.11}$$

$$+...$$

In the above equations for causal discrete-time systems the lower bound of the summations becomes zero and for the systems with limited memory L, the upper bound of summations

should be replaced with L. Also we can approximate a nonlinear system with an N^{th} order approximation having the sum of first i-fold summations where $i \leq N$:

$$y(n) = \Sigma_{r=1}^{N} \Sigma_{k_1=0}^{L-1} ... \Sigma_{k_r=0}^{L-1} h_r(k_1, ..., k_r) x(n-k_1) ... x(n-k_r) \tag{A.12}$$

$$= \Sigma_{k_1=0}^{L-1} h_1(k_1) x(n-k_1) \tag{A.13}$$

$$+ \Sigma_{k_1=0}^{L-1} \Sigma_{k_2=0}^{L-1} h_2(k_1, k_2) x(n-k_1) x(n-k_2) \tag{A.14}$$

$$+ ...$$

$$+ \Sigma_{k_1=0}^{L-1} ... \Sigma_{k_N=0}^{L-1} h_N(k_1, ..., k_N) x(n-k_1) ... x(n-k_N) \tag{A.15}$$

APPENDIX B

CMOS PROCESS CORNERS

Transistors have uncertainty in their process parameters and also due to environmental variations. Variations towards a shorter L_{eff}(effective channel length), lower V_t(threshold voltage), and a thinner t_{ox}(oxide thickness) make the device faster and the device gets slower if we have variations in the opposite direction. Also environmental variations like higher V_{DD} and lower temperature makes the device faster.

A process corner represents a three sigma(standard deviation) variation from nominal doping concentrations (and other parameters) in transistors on a silicon wafer. This variation may occur for many reasons, such as minor changes in the humidity or temperature changes in the clean-room between wafers, or due to the position of the die relative to the center of the wafer. Apart from the typical corner, there are fast and slow corners, where the carrier mobilities are higher and lower than normal, respectively.

According to the normal probability density function that characterizes the process, fast and slow corners are considered within $\pm 3\sigma$ of the mean, which makes it improbable(with probability less than 1 percent) for them to be outside these values.

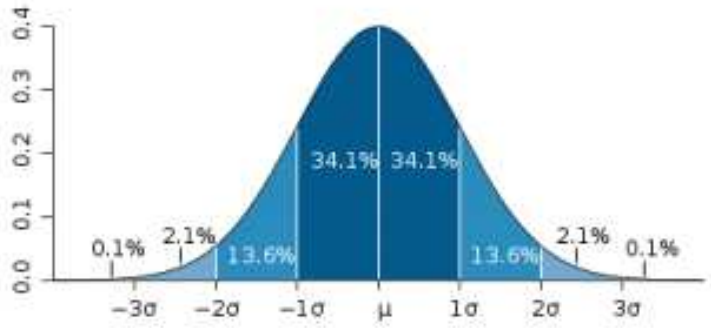

Figure B.1. Normal probability distribution curve.

BIBLIOGRAPHY

[1] Giorgio Tarrico, "On the Capacity of the binary Input Gaussian and Rayleigh Fading Channels,"*Telecommunication system Letter,*Vol. 7, No. 2, pp. 201-208, March-April 1996.

[2] V. Brik , S. Banerjee, M. Gruteser, S. Oh,"Wireless Device Identification with Radiometric Signatures," *Proceedings of the 14th ACM international conference on Mobile computing and networking* , pp. 116-1271, 2008.

[3] K.A. Remley, C.A. Grosvenor, R.T. Johnk, D.R. Novotny, P.D. Hale, M.D. McKinley, A. Karygiannis, and E. Antonakakis, "Electromagnetic Signatures of WLAN Cards and Network Security," *Proceedings of ISSPIT,*2005.

[4] O. Ureten and N. Serinken, "Wireless Security Through RF Fingerprinting," *Canadian Journal of Electrical and computer Engineering*, vol. 32, no. 1, pp. 27-33, Winter 2007.

[5] J. Dudczyk, J. Matuszewski, M. Wnuk, "Applying the radiated emission to the specific emitter identification,Microwaves, Radar and Wireless Communications," *Proceedings of 15th International Conf.MIKON-2004* ,

[6] L.E. Langley, "Specific emitter identification (SEI) and classical parameter fusion technology", *WESCON/93,* Sept 28-30, 1993.

[7] M. Barbeau, J. Hall, and E. Kranakis, "Detecting Impersonation Attacks in Future Wireless and Mobile Networks," *MADNES,* 2006.

[8] J. Hall, M. Barbeau, and E. Kranakis, "Radio frequency fingerprinting for intrusion detection in wireless networks," *Defendable and Secure Computing*, 2005.

[9] J. Hall,"Detection of rogue devices in wireless networks," *PhD thesis,* 2006.

[10] H.C. Choe, C.E. Poole, A.M. Yu, and H.H. Szu, "Novel identification of intercepted signals from unknown radio transmitters," *Proc. of SPIE,* vol. 2491, 504, 1995.

[11] K.B. Rasmussen and S. Capkun, "Implications of Radio Fingerprinting on the Security of Sensor Networks," *Proceedings of IEEE SecureComm* 2007.

[12] C. L. Corbett, R. A. Beyah, J. A. Opeland, "Using Active Scanning to Identify Wireless NICs," *Proceedings of the 7th IEEE Workshop on information Assurance, U.S. Military Academy, West Point, NY,* pp. 21-23 June 2006.

[13] J. Hall, M.Barbeau, and E. Kranakis, "Enhancing intrusion detection in wireless networks using radio frequency fingerprinting (extended abstract)," *3rd IASTED International Conference on Communications, Internet and Information Technology (CIIT),* November 2004. 2006.

[14] J. Hall, M.Barbeau, and E. Kranakis, "Detection of transient in radio frequency fingerprinting using signal phase," *IASTED International Conference on Wireless and Optical Communications,* May 2003.

[15] R.D. Hippenstiel and Y. Payal, "Wavelet Based Transmitter Identification," *Proc. Information Symposium on Signal Processing and its Applications, Gold Coast, Australia,* 1996.

[16] O. Ureten and N. Serinken, "Bayesian detection of radio transmitter turn-on transients," *Proc. NSIP99,* 1999, pp. 830-834.

[17] O. Ureten and N. Serinken, "Detection of radio transmitter turn-on transients," *Electronics Letters,* vol 35, 1999, pp. 1996-1997.

[18] O.Ureten,, and N. Serinken, "Detection, characterisation and classification of radio transmitter turn-on transient signals," *Proc. of the NATO ASI on Multisensor Data Fusion,* 2002, pp. 611-616, Kluwer Academic Publishing.

[19] N.Serinken and O. Ureten, "Generalized Dimension characterization of Radio Transmitter Turn-On Transients," *IEE Electronics Letters,* vol. 36, no. 12, pp. 1064-1064, June 2000.

[20] O. Ureten and N. Serinken, "Bayesian Detection of WiFi Transmitter RF Fingerprints," *IEE Electronics Letters,* vol. 41, no. 6, pp. 373-374, March 2005.

[21] D. Shaw, and W. Kinsner, "Multifractal Modelling of Radio Transmitter Transients for Classification," *Proc. Conference on Communications, Power and Computing,* pp. 306-312.,1997.

[22] J. Toonstra and W. Kinsner , " Transient Analysis and Genetic Algorithms for Classification,," *IProc. IEEE WESCANEX,* 1995.

[23] Kawalec A., et. al., "Mixed Method Based on Intrapulse Data and Radiated Emission to Emitter Sources Recognition",*15th Int'l Conf on Microwaves, Radar and Wireless Comm (MIKON),* May 2004.

[24] Chen Y., et al., "Detecting and Localizing Wireless Spoofing Attacks," ,*IEEE Conf on Sensor, Mesh and Ad Hoc Comm and Nets (SECON),* pp. 193-202, June 2007.

[25] D. B. Faria and D. R. Cheriton, "Radio-layer security: Detecting identity-based attacks in wireless networks using signalprints," *IProc. 5th ACM Workshop on Wireless Security (WiSe'06),* pp. 43-52, Sep. 2006.

[26] Z. Li, W. Xu, R. Miller, and W. Trappe, "Securing wireless systems via lower layer enforcements," *Proceedings of the 5th ACM workshop on Wireless security,*pp. 33-42, 2006.

[27] N. Patwari and S.K. Kasera, "Robust location distinction using temporal link signatures," *ACM MOBICOM,* pp. 111-122, 2007.

[28] Sheng Y., et al., "Detecting 802.11 MAC Layer Spoofing Using Received Signal Strength,"*IEEE 27th Annual Conference on Computer Communications (INFOCOM),* April 2008

[29] K. Remley et al, "Electromagnetic Signatures of WLAN Cards and Security," *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology,* December 2005.

[30] V. Brik, S. Banerjee, M. Gruteser, S. Oh,"Wireless Device Identification with Radiometric Signatures," *Proceedings of the 14th ACM international conference on Mobile computing and networking*, pp.116-127, 2008.

[31] K.I. Talbot, P.R. Duley, and M.H. Hyatt, "Specific Emitter Identification and Verification," *Technology Review,* 2003.

[32] T. Daniels, M. Mina, and S. F. Russell,"Short Paper: A Signal Fingerprinting Paradigm for General Physical Layer and Sensor Network Security and Assurance," *Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SECURECOMM),* 2005.

[33] B. Razavi, *Principles of Data Conversion System Design,* EEE Press, 1995.

[34] Data Sheet, Analog Devices AD9776A/AD9778A/AD9779A, http://www.analog.com/UploadedFiles/Data_Sheets/AD9776A_9778A_9779A.pdf

[35] Data Sheet, Analog Devices AD9780/AD9781/AD9783, http://www.analog.com/UploadedFiles/Data_Sheets/AD9780_9781_9783.pdf

[36] "Basic Technology of Quartz Crystal Oscillators", http://www.4timing.com/techoscillator.htm.

[37] R. Gitlin, J. Hayes, and S. Weinstein, *Data Communication Principles,* Springer, 1992.

[38] Georgi I. Radulov, Markus Heydenreich, Remco W. van der Hofstad, Johannes A. Hegt, and Arthur H. M. van Roermund, "Brownian-Bridge-Based Statistical Analysis of the DAC INL Caused by Current Mismatch," *IEEE Transactions on Circuits and Systems II: Express Briefs,* vol.54, no.2, pp.146-150, Feb. 2007

[39] A. Menon, *M.S. Thesis: Power Amplifier Linearization and Implementation,* University of Massachusetts, August 2007.

[40] E. Ngoya, and A. Soury," Modeling memory effects in nonlinear subsystems by dynamic Volterra series," *Behavioral Modeling and Simulation, Proceedings of the 2003 International Workshop,* pp. 28-33, 2003.

[41] E. Ngoya, N. Le Gallou, et al," Accurate RF and Microwave System Level Modeling of Wide Band Nonlinear Circuits," *IEEE MTT-S Digest,* Vol. 1, pp. 79-82, 2000.

[42] P. Wambacq and W. Sansen, *The Distortion analysis of analog integrated circuits,* Kluwer, 1998.

[43] B Hernes, T. Sather *Design criteria for low distortion in feedback opamp circuits,* Springer, 2003.

[44] E. Bedrosian, S. O. Rice,"The output properties of Volterra systems(nonlinear systems with memory) driven by harmonic and Gaussian inputs," *Proc. IEEE,* Vol. 59, No. 12, pp. 1688-1707, Dec. 1971.

[45] M. Rudko and D. Weiner,"Volterra systems with random inputs: A formalized approach," *IEEE trans. communications,* Vol. COM-26, No. 2, pp. 217-225, Feb. 1978.

[46] M. Schetzen, *The Volterra and Wiener theories of nonlinear systems,* John Wiley and sons, 1980.

[47] S. Narayanan,“ Transistor distortion analysis using Volterra series representation,” *The Bell System Technical J.,* Vol. 46, No. 12, pp. 991-1024, May/June 1967.

[48] S. Narayanan,“Application of Volterra series to intermodulation distortion analysis of transistor feedback apmlifiers,” *IEEE trans. Circuit theory,* Vol.CT-17, No. 4, pp. 518-527, Nov. 1970.

[49] R. Meyer, M. Shemsa, and R. Eschenbach,“ Cross modulation and intermodulation in amplifiers at high frequencies,” *IEEE J solid-state Circuits,* Vol.SC-76, No. 1, pp. 16-23, Feb 1972.

[50] W. Sansen, *Ph.D. Dissertation: Optimum design of integrated variable-gain amplifiers,* University of California Berkeley, 1972.

[51] L.O. Chua, C. Y. Ng.,“ Frequency-domain analysis of nonlinear systems: formulation of transfer functions,” *IEE J. electronic Circuits and systemst,* Vol. 3,No. 6, pp. 257-269, Nov. 1979.

[52] T. Wand and T. J. Brazil,“ The Estimation of Volterra Transfer Functions with Applications to RF Power Amplifier Behavior Evaluation for CDMA Digital Communication ,” *IEEE MIT-S International Microwve Symposium Digest,* Vol. 1, pp. 425-428, 2000.

[53] J. Dooley, B. OBrien and T. J. Brazil,“ Behavioral Modeling of RF Power Amplifiers Using Modified Volterra Series in the Time Domain,” *IEEE High frequency Postgraduate student Colloquiumt,* pp. 169-174, Sept. 2004.

[54] David A. Hodges, Horace G. Jackson, Resve A. Saleh, *Analysis and design of digital integrated circuits,* McGraw-Hill,pp. 125-127, 2003.