

THE UNIVERSITY OF CALGARY

Adaptive Predistortion with Neural Networks

by

James McTavish

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR  
THE  
DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF ELECTRICAL AND COMPUTING  
ENGINEERING

CALGARY, ALBERTA

November, 2002

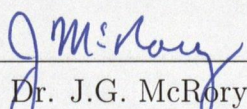
© James McTavish 2002

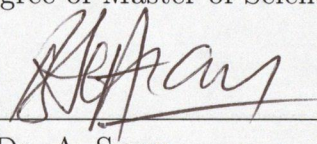


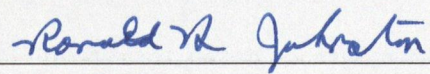
# THE UNIVERSITY OF CALGARY

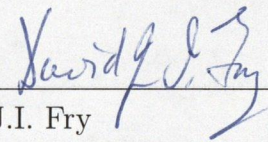
## FACULTY OF GRADUATE STUDIES


The undersigned certify that they have read, and recommend "Adaptive Pre-distortion with Neural Networks" submitted by James McTavish in partial fulfillment of the requirements for the degree of Master of Science.

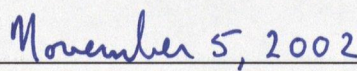
  
\_\_\_\_\_  
Supervisor, Dr. J.G. McRory  
Department of Electrical and Computer Engineering

  
\_\_\_\_\_  
Dr. A. Sesay  
Department of Electrical and Computer Engineering

  
\_\_\_\_\_  
Co-Supervisor, Dr. R.H. Johnston  
Department of Electrical and Computer Engineering

  
\_\_\_\_\_  
Dr. D.J.I. Fry  
Department of Physics and Astronomy

  
\_\_\_\_\_  
Dr. B. Davies  
Department of Electrical and Computer Engineering

  
\_\_\_\_\_  
Date



# Abstract

Amplifiers have had great demands placed on them as spectrally efficient modulation techniques become more prevalent in the industry, and bandwidth requirements increase. Not only must they have broad bandwidth, they must remain linear over a large dynamic range. Linearization has been shown to be a viable technique for improving an amplifier's performance.

Baseband predistortion is a linearization technique with excellent performance capabilities that requires very little additional RF equipment, and the system can be made adaptive resulting in reduced sensitivity to changes. Until recently the focus has been on memoryless predistortion. For narrow-band systems the memory effects of the amplifier can largely be ignored. However, as bandwidth increases the memory effects will become more significant and can't be ignored.

This thesis discusses the effect of increasing bandwidth on memoryless predistortion using neural networks. Neural networks were chosen for their ability to model black box systems by knowing only their input and output. The system can also be extended to compensate for memory using Wiener and Hammerstein models. Results from simulation and experiments are presented and compared.

## Acknowledgements

I would like to express my gratitude for my supervisors Dr. John McRory and Dr. Ron Johnston for their help and guidance. They have been a very valuable resource throughout the thesis.

I would also like to sincerely thank my colleagues at *TRLabs* for their feedback and help. Specifically I would like to thank Rob Randall and Brad Davis for their discussions on linearization and support in the labs.

Last but not least I would like to thank *TRLabs* and the University of Calgary for the financial support without which this work would not have been possible.

## Dedication

*To my family and friends.  
Their support and encouragement made this possible.*

# Table of Contents

Approval Page	ii
Abstract	iii
Acknowledgements	iv
Dedication	v
Table of Contents	vi
List of Symbols and Abbreviations	xii
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	3
1.2 Thesis Contribution . . . . .	4
1.3 Outline . . . . .	5
<b>2 RF Power Amplifier Distortion</b>	<b>6</b>
2.1 Amplifier Distortion Effects . . . . .	6
2.1.1 Linear Distortion . . . . .	7
2.1.2 NonLinear Distortion . . . . .	9
2.2 Linearity Evaluation . . . . .	13
2.2.1 1dB Compression point . . . . .	14
2.2.2 Error Vector Magnitude . . . . .	16
2.2.3 Third Order Intercept . . . . .	17
2.2.4 Adjacent Channel Power Ratio . . . . .	18
<b>3 Neural Networks</b>	<b>21</b>
3.1 Structure . . . . .	21
3.1.1 Neuron Structure . . . . .	22
3.1.2 Network Structure . . . . .	23
3.2 Output Calculation . . . . .	25
3.3 Training . . . . .	26
3.3.1 Backpropagation . . . . .	27
3.3.2 Gradient Descent . . . . .	31

<b>4</b>	<b>Linearization Techniques</b>	<b>34</b>
4.1	Output Power Reduction . . . . .	35
4.2	Feed-Back . . . . .	36
4.3	Feed-Forward . . . . .	39
4.4	LINC . . . . .	42
4.5	Envelope Elimination and Restoration . . . . .	44
<b>5</b>	<b>Predistortion</b>	<b>47</b>
5.1	Theory . . . . .	47
5.1.1	Derivation . . . . .	48
5.2	Implementation . . . . .	51
5.3	Limitations . . . . .	55
5.4	Advantages . . . . .	56
5.5	Disadvantages . . . . .	57
<b>6</b>	<b>System Overview</b>	<b>58</b>
6.1	Amplifier Model . . . . .	58
6.2	System Structure . . . . .	60
6.2.1	Nonlinear Predistortion Networks . . . . .	60
6.2.2	Linear Predistortion . . . . .	62
6.2.3	Nonlinear Training Block . . . . .	63
6.2.4	Linear Training Block . . . . .	64
<b>7</b>	<b>Simulation Results</b>	<b>66</b>
7.1	Device Under Test . . . . .	66
7.2	Network Size . . . . .	67
7.3	Nonlinear Predistortion Performance . . . . .	69
7.4	Nonlinear Model Performance . . . . .	72
7.5	Model Performance with Linear Distortion . . . . .	72
7.6	Linear and Nonlinear Predistortion Performance . . . . .	73
<b>8</b>	<b>Experimental Results</b>	<b>75</b>
8.1	Implementation Issues . . . . .	75
8.1.1	Feedback Attenuation . . . . .	75
8.2	Experimental Setup . . . . .	78
8.2.1	Experimental System . . . . .	78
8.2.2	Setup Limitations . . . . .	79
8.2.3	Setup Verification . . . . .	82

8.3	Results . . . . .	84
8.3.1	Memoryless Predistortion Performance . . . . .	84
8.3.2	Predistortion with Memory . . . . .	85
<b>9</b>	<b>Conclusions and Future Work</b>	<b>89</b>
9.1	Thesis Summary . . . . .	89
9.2	Achievements . . . . .	91
9.3	Future Work . . . . .	92
	<b>Bibliography</b>	<b>95</b>



## List of Tables

3.1 Symbols Used . . . . .	26
----------------------------	----

## List of Figures

1.1	Forecast Revenue Breakdown for Europe (source: Digitrends) .	2
2.1	Linear Response . . . . .	7
2.2	Gain Compression . . . . .	11
2.3	Intermodulation Products . . . . .	12
2.4	Distorted Spectrum . . . . .	13
2.5	AM/PM Distortion . . . . .	14
2.6	1dB Compression Point . . . . .	15
2.7	IM3 . . . . .	18
3.1	Node Structure . . . . .	23
3.2	Activation Functions . . . . .	24
3.3	Single Layer Perceptron . . . . .	25
3.4	Gradient Descent with Momentum . . . . .	33
4.1	Feedback Linearization . . . . .	37
4.2	Feed-Forward Linearization . . . . .	39
4.3	LINC Transmitter . . . . .	43
4.4	Envelope Elimination and Restoration . . . . .	45
5.1	Predistortion . . . . .	48
5.2	Predistortion Block Diagram . . . . .	48
5.3	Symbol Predistortion Vs. Signal Predistortion . . . . .	54
5.4	Predistortion Limitations . . . . .	56
6.1	Nonlinear Systems with Memory . . . . .	59
6.2	A Hammerstein System . . . . .	59
6.3	System Block Diagram. . . . .	61
6.4	Nonlinear Predistortion . . . . .	62
6.5	Nonlinear Training Block . . . . .	64
7.1	Network Size Vs. Modeling Accuracy . . . . .	68
7.2	Network Size Vs. Amplitude Distortion Modeling Accuracy . .	69
7.3	Nonlinear Predistortion Performance . . . . .	70
7.4	Distortion of QAM-16 Constellation . . . . .	71
7.5	Model Performance . . . . .	73

7.6	Predistortion with Memory Performance . . . . .	74
8.1	Effects of Feedback Attenuation . . . . .	77
8.2	Experimental Setup . . . . .	79
8.3	Modulated Spectrum Verification . . . . .	83
8.4	Demodulated Spectrum Verification . . . . .	83
8.5	Nonlinear Cancellation Vs. Bandwidth . . . . .	85
8.6	IM3 Suppression Vs. Bandwidth . . . . .	86
8.7	Predistortion With Memory . . . . .	87

## List of Symbols and Abbreviations

$A$	Signal amplitude
$A()$	Polar amplifier amplitude transfer function
ACPR	Adjacent channel power ratio
ADC	Analog to digital converter
AM/AM	Amplitude modulation to amplitude modulation
AM/PM	Amplitude modulation to phase modulation
$B_n$	Vector of biases for layer $n$
$c$	The speed of light
CDMA	Code division multiple access
$\delta_n$	Weight:Output change ratio for layer $n$
DAC	Digital to analog converter
dB	Logarithmic ratio of power (decibels)
dBm	Logarithmic ratio of power (decibels) with respect to 1 milliwatt
DC	Direct Current
DQPSK	Differential quadrature phase shift keying
DSP	Digital signal processor
DUT	Device under test
$E_n$	Vector of errors for layer $n$
EVM	Error vector magnitude
$f$	Frequency in Hertz
FIR	Finite Impulse Response
FPGA	Field Programmable Gate Array
G	Amplifier gain
GHz	Gigahertz
GPIO	General purpose interface bus
IM3	Third order intermodulation products
IMD	Intermodulation distortion
ISI	Intersymbol interference
$k_n$	Power series coefficients
kHz	Kilohertz
LINC	Linear amplification with nonlinear components
LMS	Least mean squared
MHz	Megahertz
$N_n$	Number of nodes for layer $n$

$O_n$	Vector of outputs for layer $n$
$\Psi()$	Complex amplifier transfer function
$\Psi_{LIN}()$	Linear component of $\Psi$
$\Psi_{NL}()$	Nonlinear component of $\Psi$
$\Phi()$	Complex compensation function
PA	Power amplifier
PAR	Peak to average ratio
PSD	Power spectral density
QAM	Quadrature amplitude modulation
RF	Radio Frequency
$S_n$	Vector of combined inputs for layer $n$
$\theta$	Phase rotation in radians
$\Theta()$	Polar amplifier phase transfer function
$t$	Time
$T_n$	Vector of targets for layer $n$
$v_i$	Input voltage
$v_o$	Output voltage
$\omega_n$	Radial Frequency
$\Omega()$	Polar AM/PM compensation
$\Upsilon()$	Polar AM/AM compensation
$W_n$	Matrix of weights for layer $n$

# Chapter 1

## Introduction

The society we live in has developed into an interconnected web that allows many forms of information to travel seamlessly around the globe on many networks, including the Internet and telecommunications networks. With the advancement of the Internet in the 1990s everyone was introduced to what is now known as the Information Revolution. In the late 1990s wireless communications with data services began growing rapidly.

Wireless usage in general is growing at an exponential rate. More than ever people are communicating with wireless products, but the methods that they use are evolving. New cellular phones include digital cameras to take and transmit still pictures or videoconference. Sending short text messages has become a major way to communicate and the ability to receive email anywhere is becoming commonplace.

Figure 1.1 shows that the forecasts in Europe for revenue from voice communication is expected to plateau. Now that market penetration of cellular phones is almost complete, the cellular providers have to innovate in order to generate new revenue streams. New services such as text messaging are expected to become the new source of revenue growth. Further in the future raw data transfer will become another major source of income. Unlike voice



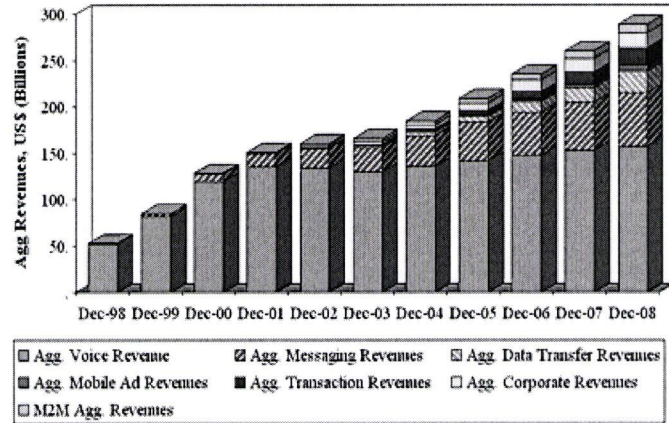


Figure 1.1: Forecast Revenue Breakdown for Europe (source: Digitrends)

communications, some of these new services require broadband data communication.

In the next few years, cellular providers will begin rolling out new communications networks, with forecast bit rates exceeding 1Mb/s. Soon cellular providers will be able to compete with more traditional Internet service providers. Not only will broad bandwidth be available, but the cellular customers will enjoy the freedom of multimedia Internet communication wherever they are. Beyond this bandwidth, future wireless links have been proposed that reach into the 1Gb/s ranges. With this bandwidth the cellular providers will be able to offer service that not only rivals wired Internet service providers, but will compete with them on an even footing.

Throughout all of this change, one thing remains constant, wireless com-

munications will continue to grow with faster connections. The engineering challenges that these broadband communications networks pose are staggering. From the wireless network infrastructure to the mobile terminal, everything must be redesigned.

## 1.1 Motivation

Traditional mobile system deployment involves several mobile terminals communicating with a network of base stations. While the mobile terminal will be in contact with a single base station, each base station will communicate with several mobile terminals simultaneously. This places great demands on the performance of the base station power amplifiers.

With current systems the amplifiers that are used, are typically linear amplifiers. Most signals are amplified without incurring much distortion. However for digital multi-user modulation schemes with high peak to average ratios, the efficiency of these amplifiers is typically below 5%. If the amplifiers are operated at a higher input signal level the efficiency will increase, however the distortion introduced by the amplifier will also increase.

Linearization techniques are methods where the distortion introduced by the amplifier is reduced, allowing for a more linear amplification. Each technique has its advantages and disadvantages. The focus for this thesis is a technique known as predistortion. Predistortion works by predicting what the distortion will be, and altering the signal applied to the input of the ampli-

fier such that the output of the amplifier is the desired signal with an overall linear gain. Until the mid 1990s most predistortion schemes treated the amplifier as a memoryless nonlinear system, that is the output at any time is only dependent on the input at that time. This assumption usually works for narrow-band systems, however, as the signal bandwidth grows, high power amplifiers often begin to exhibit the symptoms of memory.

This thesis will investigate the effects of memory on memoryless predistortion, and will introduce a method of compensating for these effects.

## 1.2 Thesis Contribution

Several predistortion methods that compensate for memory have been proposed and simulated, however, to the author's knowledge only one has actually been implemented on the bench[1]. This thesis is the first time a Hammerstein predistorter has been implemented and tested on a physical amplifier. An implementation of a Hammerstein system is proposed using neural networks to implement the nonlinear portion. Memoryless predistortion is then compared to the new proposed system.

The main contribution reported in this thesis is the test and validation of the Hammerstein predistorter concept. To the author's knowledge no other publication has shown the the results of a predistortion algorithm based on a Hammerstein predistorter tested with a physical amplifier.

### 1.3 Outline

Chapter 2 begins with an introduction to the effects of amplifier distortion and the metrics used to measure the distortion. Neural networks are introduced in Chapter 3, where their structure and the methods used to train them will be discussed. Chapter 4 is an overview of the methods used to compensate for the distortions introduced by power amplifiers. Each method will be introduced and then the advantages and disadvantages will be discussed. This continues into Chapter 5 which discusses linearization with predistortion. Here the concept of predistortion is introduced, and is discussed in detail along with some limitations and practical implementation issues. Then, combining the concept of predistortion and neural networks, the proposal for this thesis is introduced in Chapter 6. Here, neural networks are used to model the nonlinearities of the amplifier, and compensate for them. Chapter 7 simulates the various components of the proposed system and presents some results. Chapter 8 extends on the simulation by testing the system on a test-bench, and discusses some results from an experimental setup comparing them to the simulation. Finally, Chapter 9 concludes the thesis and proposes some future research work.

## Chapter 2

### RF Power Amplifier Distortion

The role of the base station RF power amplifier is to provide a signal at adequate power levels for transmission to the user terminal. Ideally the amplifier, amplifies the input signal to the desired output power without distortion, however practical amplifiers are nonlinear devices that will distort the signal. The overall distortion can be broken down into several effects each of which manifest themselves in a different way. This chapter discusses the effects of the distortion and the metrics used to measure its severity.

#### 2.1 Amplifier Distortion Effects

The distortions incurred by an amplifier can be broken down into two categories, linear and nonlinear. Linear distortions are not dependent on the input power of the signal, only the frequencies of the input signal. In contrast nonlinear distortions depend on the power of the input signal. Each of these categories has a different effect on the distortion products that manifest themselves. Linear effects typically contribute to inter-symbol interference and become a problem as the data rates and the system bandwidth increases. Non-linear distortions distort the constellation that is transmitted and can cause the wrong symbol to be detected along with causing the transmitted signal to

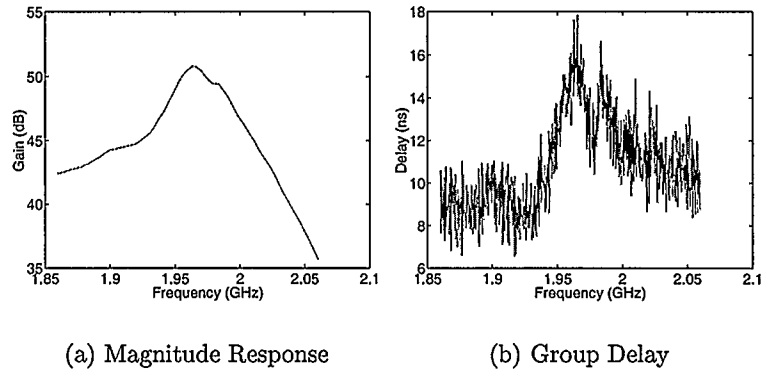


Figure 2.1: Linear Response

occupy a larger spectrum.

### 2.1.1 Linear Distortion

Since amplifiers may operate in wide bandwidths, they are subject to linear frequency dependent distortions. In general the linear distortion is composed of two components, frequency dependent gain and frequency dependent delay.

#### Frequency Dependent Gain

Ideally the amplifier's frequency response will be flat within the operating band, independent of frequency. The frequency response of a system composed of a Motorola MHW-1916 driving a Motorola MRF-282SR1 amplifier is shown in Figure 2.1. Each plot has a frequency range of 200MHz centered at 1.96 GHz. As the plot shows, the gain of the amplifier is not constant within the operational band of the amplifier.



If the system is narrow band, the change in gain with respect to frequency will not be great and can be assumed to be flat. However, since newer systems are using a much higher data rate and wider signal bandwidth, the assumption that the gain is constant over the bandwidth is not valid. With these high data rate systems, the linear distortion will manifest itself as inter-symbol interference, impacting the bit error rate of the system.

### Frequency Dependent Delay

The group delay of a system characterizes the delay through a system for a given frequency. It is defined as the negative derivative of the phase with respect to frequency, as given in (2.1).

$$\tau_g = -\frac{d\theta(\omega)}{d\omega} \quad (2.1)$$

Under ideal circumstances this group delay, like gain, would remain constant across the operating band. However as can be seen by Figure 2.1b, practical amplifiers do not have a constant group delay. Since the frequencies are delayed differently, the individual frequency components of the signal do not line up at the output correctly, resulting in signal distortion. As with frequency dependent gain, the frequency dependent delay causes inter-symbol interference which increases as the data rate increases.

The cause of group delay variation can be attributed to many sources including but not limited to frequency responses of matching networks, nonlinear capacitances of the transistors and the response of the bias networks.

### 2.1.2 NonLinear Distortion

While linear distortions only distort the signal and do not create any new frequencies, nonlinear distortions create new frequencies and cause the spectral content of the signal to grow. This spectral growth impacts directly on how many channels will fit in a given frequency range, and the overall spectral efficiency of the system. Nonlinear distortion can be broken down into two major components, amplitude distortion, and phase distortion [2].

#### Amplitude Distortion

The ideal amplifier would apply a constant gain to the input signal regardless of its level. Under the assumption that the amplifier is memoryless, the input to output relationship of the amplifier can be expressed with a power series as in (2.2)[3].

$$v_o(t) = \sum_n k_n (v_i(t))^n \quad (2.2)$$

where  $v_i$  and  $v_o$  are the input and output voltages respectively and  $k_n$  are constants. Substituting a simple sinusoidal input  $v_i(t) = A \cos(\omega t)$ , where  $A$  and  $\omega$  are the amplitude and frequency, expanding out the first five terms, applying trigonometric identities and collecting terms, yields (2.3).

$$\begin{aligned}
v_o = & \left( \frac{1}{2}k_2A^2 + \frac{3}{8}k_4A^4 \right) + \left( k_1A + \frac{3}{4}k_3A^3 + \frac{5}{8}k_5A^5 \right) \cos(\omega t) + \\
& \left( \frac{1}{2}k_2A^2 + \frac{1}{2}k_4A^4 \right) \cos(2\omega t) + \left( \frac{1}{4}k_3A^3 + \frac{5}{16}k_5A^5 \right) \cos(3\omega t) + \\
& \left( \frac{1}{8}k_4A^4 \right) \cos(4\omega t) + \left( \frac{1}{16}k_5A^5 \right) \cos(5\omega t)
\end{aligned} \tag{2.3}$$

This equation shows that a number of new harmonics are introduced, including components at DC and multiples of the fundamental frequency. Dividing the output amplitude of the fundamental frequency by  $A$  results in the gain of the fundamental frequency.

$$G = k_1 + \frac{3}{4}k_3A^2 + \frac{5}{8}k_5A^4 \tag{2.4}$$

Ideally the gain would be independent of the input power. However the gain is dependent on the input amplitude  $A$ , related by all odd coefficients equal to or greater than one. This distortion is known as AM/AM distortion. The result is a deviation from the linear gain,  $k_1$ , as power increases. The coefficient  $k_3$  governs whether the system is expansive or compressive if it is positive or negative respectively. Most practical amplifiers,  $k_3$  will be negative. The power in versus power out relationship for a Motorola MHW-1916 driving a Motorola MRF-282SR1 is shown in Figure 2.2.

If a two-tone signal,  $v_i(t) = A \cos(\omega_1 t) + A \cos(\omega_2 t)$  is inserted into (2.2) and three terms of the result are expanded out, the result is shown in (2.5).

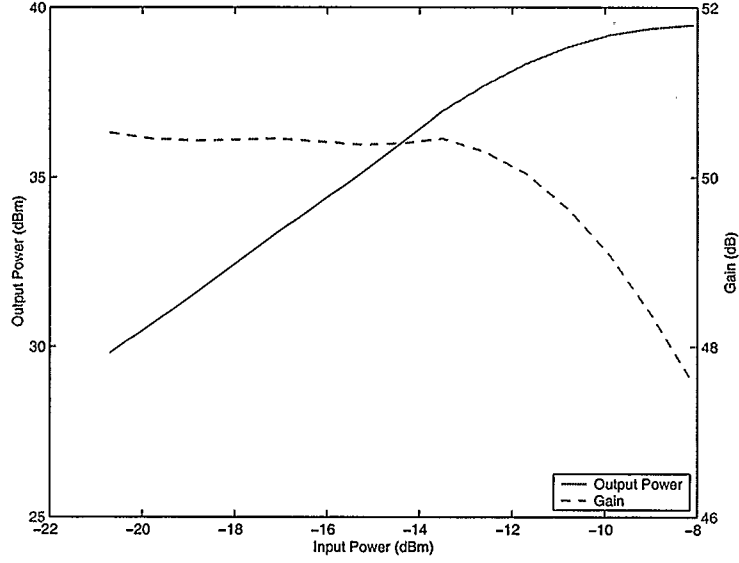


Figure 2.2: Gain Compression

$$\begin{aligned}
 v_o = & k_2 A^2 + \left( k_1 A + \frac{9}{4} k_3 A^3 \right) (\cos \omega_1 t + \cos \omega_2 t) + \\
 & k_2 A^2 [\cos (\omega_1 t + \omega_2 t) + \cos (\omega_1 t - \omega_2 t)] + \\
 & \frac{3}{4} k_3 A^3 [\cos (2\omega_1 t - \omega_2 t) + \cos (2\omega_2 t - \omega_1 t) + \cos (2\omega_1 t - \omega_2 t) + \cos (2\omega_2 t - \omega_1 t)] + \\
 & \frac{1}{2} k_2 A^2 [\cos (2\omega_1 t) + \cos (2\omega_2 t)] + \\
 & \frac{1}{4} k_3 A^3 [\cos (3\omega_1 t) + \cos (3\omega_2 t)]
 \end{aligned} \tag{2.5}$$

This shows that for a signal with two frequencies, the output will contain combinations of these frequencies called intermodulation products. As more terms are expanded, an increasing number of these intermodulation products will occur. In general these combinations will take on the form of  $m\omega_1 \pm n\omega_2$

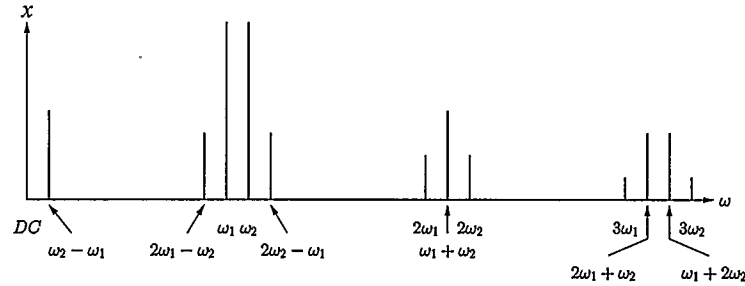


Figure 2.3: Intermodulation Products

where  $m$  and  $n$  represent any integer equal to or greater than zero. Figure 2.3 illustrates the effects of the intermodulation products in the frequency domain. As can be seen from Figure 2.3, the intermodulation that will generally cause the most problems are the third order products, which are characterized by  $m + n = 3$ . These products are the closest to the desired signal, which makes them difficult or impossible to filter out without adversely affecting the fundamental signal.

If the input signal contains more than two frequencies, the modulation products at the output will contain every combination of every input frequency. This will result in a frequency plot such as Figure 2.4 which was obtained from a system composed of a Motorola MHW-1916 driving a Motorola MRF282SR1, operating well into its nonlinear region.

### Phase Distortion

If the coefficients  $k_n$  in (2.2) are complex, this will give rise to an output phase response that is also dependent on the input amplitude. This distortion is

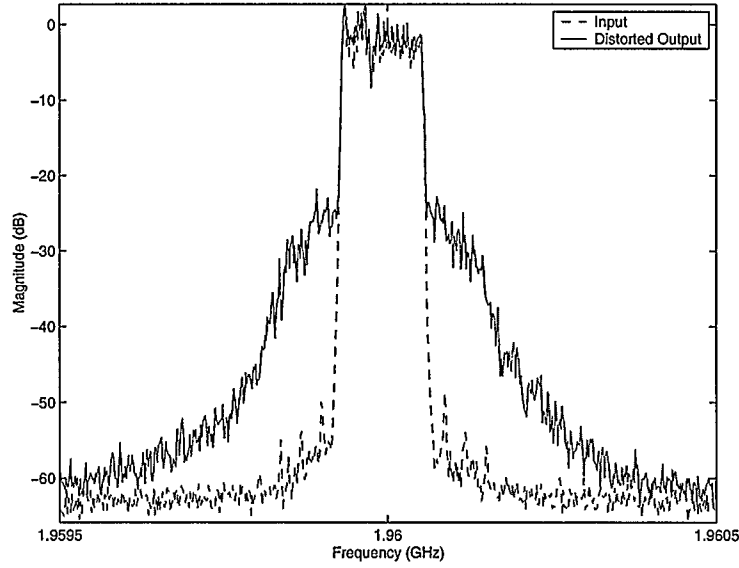


Figure 2.4: Distorted Spectrum

known as AM/PM distortion and results in a non-constant phase. This is shown in Figure 2.5, which is a plot of output phase for a given input power with a single tone input. For an input signal with multiple frequencies, the output will contain extra phase modulation similar to the intermodulation products introduced by nonlinear amplitude distortion.

## 2.2 Linearity Evaluation

To characterize and measure the effects of distortion from both linear and nonlinear sources, several metrics are employed. The simple 1dB compression point provides a measure between linear and nonlinear regions. The Error Vector Magnitude, or EVM, measures the amount of distortion applied to a



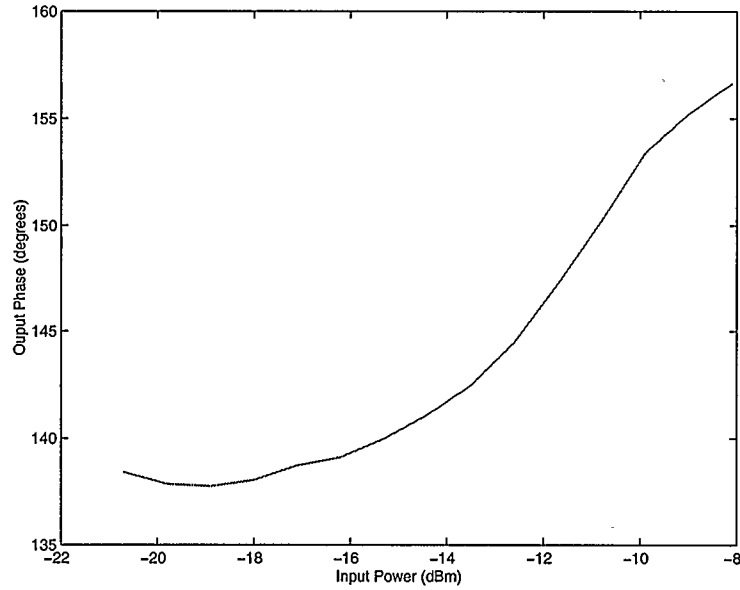


Figure 2.5: AM/PM Distortion

signal in the time domain. The IM3 point measures the level of third order intermodulation products with respect to the fundamental frequency band. Finally the Adjacent Channel Power Ratio measures the level of frequency emissions outside the signal with respect to the power of the fundamental frequencies.

### 2.2.1 1dB Compression point

One of the simplest and most commonly used boundaries between an amplifier's linear and nonlinear regions is the 1dB compression point. When the gain is plotted against the output power, the point at which the gain drops 1dB below the ideal linear gain is the 1dB compression point. The results with a

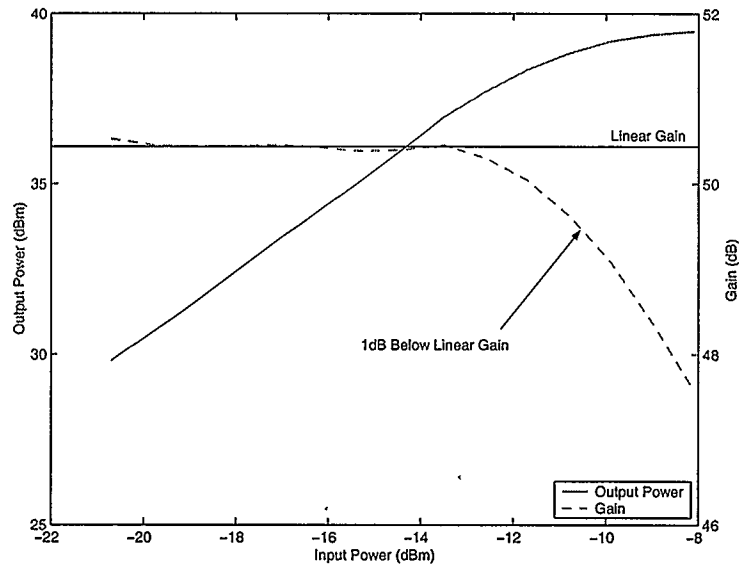


Figure 2.6: 1dB Compression Point

Motorola MHW-1916 driving a Motorola MRF-282SR1 is shown in Figure 2.6.

While no part of the amplifier's response is perfectly linear, if an amplifier's output power remains below this point, the output should remain in the region where the response is dominated by the linear gain term. If the output power approaches or rises above the 1dB compression point, the signal will enter the nonlinear region and be distorted.

As can also be seen from Figure 2.6 the linear gain value is very subjective. The most commonly used and accepted values are the average gain and the maximum gain, however any value around these can be used. Despite the different values used, the 1dB compression point will usually be within 1 to 2 dB of input power regardless of what value for linear gain is used. This

is because in general once the amplifier begins to enter its saturation region, the transition is swift. An accurate 1dB compression point is in general not required as the value is primarily used to serve as a general boundary between linear and nonlinear regions. As such, it does not characterize any of the linear distortion introduced by the amplifier.

### 2.2.2 Error Vector Magnitude

The error vector magnitude, or EVM, is used to quantify the amount of distortion applied to a signal by a system. The distortion is measured in the time domain and includes both linear and nonlinear distortion effects. The error vector is defined as the difference between the desired signal and the distorted signal. In the case of a complex signal with in-phase and quadrature components, this difference will take the form of a vector. The EVM is defined as the average magnitude of the error vector with respect to the magnitude of the input signal. The general definition is provided in (2.6).

$$EVM = \frac{1}{M} \sum_{n=1}^M \frac{|x - \tilde{x}|}{|x|} \quad (2.6)$$

If the signal is complex with in-phase and quadrature components, EVM is defined as

$$EVM = \frac{1}{M} \sum_{n=1}^M \sqrt{\frac{(x_i[n] - \tilde{x}_i[n])^2 + (x_q[n] - \tilde{x}_q[n])^2}{x_i^2[n] + x_q^2[n]}} \quad (2.7)$$

where  $x[n]$  is the desired signal and  $\tilde{x}[n]$  is the distorted signal. The subscripts  $i$  and  $q$  indicate the in-phase and quadrature components respectively. This definition normalizes the EVM to the expected signal, if the EVM is to be normalized to the peak magnitude of the ideal complex signal

$$EVM = \frac{1}{M} \sum_{n=1}^M \sqrt{\frac{(x_i[n] - \tilde{x}_i[n])^2 + (x_q[n] - \tilde{x}_q[n])^2}{x_{imax}^2 + x_{qmax}^2}} \quad (2.8)$$

where  $x_{imax}$  and  $x_{qmax}$  represent the maximum values of the in-phase and quadrature values of the desired signal respectively.

### 2.2.3 Third Order Intercept

From (2.5) it can be shown that the power of the third order intermodulation frequency is a cube of the fundamental power, and it increases in power three times faster than the fundamental. The third order intercept point is defined as the point where the fundamental and the third order intermodulation powers if they continued in a linear fashion, are equal. Due to saturation and power limitations, this point can never be reached, however it can be extrapolated from the data of a two-tone test as in Figure 2.7. When the data from the fundamental power of one of the two input tones, and the intermodulation power are plotted with respect to input power and then linearly extrapolated, they will cross for some output power. The output power where this would occur is the third order intercept point, also known as the IM3 point. From this measurement, the level of third order intermodulation frequencies can be

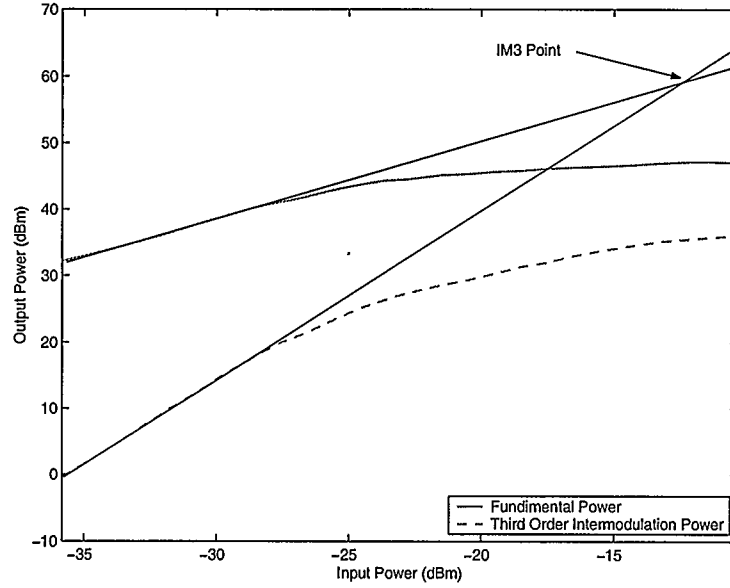


Figure 2.7: IM3

estimated for a given output power.

#### 2.2.4 Adjacent Channel Power Ratio

The Adjacent Channel Power Ratio, or ACPR, is a measure of the power that is outside the main channel with respect to the power that is contained in it. This specification is used to ensure that the power of a signal does not interfere with adjacent channels. It is defined as the ratio of the power contained in a portion of the adjacent channel to the power contained in the main channel. To determine the ACPR the power spectral density of the signal needs to be evaluated. The PSD of a signal is defined in (2.9).

$$P_x(f) = |fft(s)|^2 \quad (2.9)$$

Using the PSD the power of the adjacent channel and the main channel can be evaluated from (2.10) and (2.11) respectively.

$$P_{adj} = \int_{f_o - \frac{B_{adj}}{2}}^{f_o + \frac{B_{adj}}{2}} P_x(f) df \quad (2.10)$$

$$P_{main} = \int_{\frac{B_{main}}{2}}^{\frac{B_{main}}{2}} P_x(f) df \quad (2.11)$$

The bandwidth of the main and adjacent channels are  $B_{main}$  and  $B_{adj}$  respectively, and the frequency offset between the carrier frequency and the adjacent channel frequency is  $f_o$ . (2.10) represents the power contained in the upper adjacent channel. The lower adjacent channel does not necessarily have to be equal. If the upper and lower adjacent channels differ in their power levels, the average of the upper and lower is used. The ACPR is then defined as the ratio of the two powers.

$$ACPR = 10 \log \left( \frac{P_{adj}}{P_{main}} \right) \quad (2.12)$$

The primary purpose of the ACPR metric is to define the interference that will be contributed to adjacent channels in an FDMA system. Regulatory bodies will specify a maximum ACPR that is allowed. The specifications will include what values of  $B_{main}$ ,  $B_{adj}$ , and  $f_o$  should be used in the measurement.

Since linear systems can't generate new frequencies, and the ACPR measures the interference outside the original signals bandwidth, only nonlinear distortions will contribute to the overall ACPR. The primary contribution to ACPR is the intermodulation products generated by the nonlinear distortion in the amplifier.

## Chapter 3

### Neural Networks

Neural networks provide a flexible tool to model any arbitrary nonlinear system. They have several advantages over other modeling techniques, including their ability to have their structure tailored to the complexity of the problem, modeling arbitrary systems without knowledge of the inner workings of the system, and generalizing the input to output relationship of the system.

The seminal work on neural networks and parallel distributed processing is by David E. Rumelhart and James L. McClelland in “Parallel Distributed Processing”, Volumes 1 and 2 [4, 5]. Unless otherwise stated, the information in this chapter is from these books. The reader is strongly encouraged to consult it for a more detailed view of neural networks.

#### 3.1 Structure

The structure of a neural network can be very arbitrary. A network is composed of computational nodes called neurons. Each neuron can have several inputs and can output to several other neurons. In their most general form, a neural network could be any number of neurons with interconnections between arbitrary pairs. In practice, neural networks usually are composed of neurons in regular structures. This section will first discuss the structure of the neuron



itself, followed by various structures of neural networks.

### 3.1.1 Neuron Structure

The general structure of a neuron is shown in Figure 3.1. Each neuron is composed of two parts, the input combination and the activation function. Since each neuron can have several inputs, each input must be weighted, combined into a single value and biased to provide a value for the node to operate on. Several common methods exist to accomplish this. The dot-product is the most common, where the inputs are simply weighted, summed, and then biased to form a single answer. Other more exotic methods do exist but will not be discussed in this thesis.

The activation function is, in general, a nonlinear function with several properties. As some learning functions use derivatives of the activation function, it is advantageous for the activation function to be continuously differentiable. It is also advantageous for it to have a output limiting property, where any input will result in a bounded output, commonly in the interval of  $\begin{bmatrix} -1 & 1 \end{bmatrix}$ . This type of function is known as a squashing function. Finally for the purposes of this thesis an activation function should be single valued, that is only one possible output value for any single input.

Common activation functions include sigmoidal based functions as in Figure 3.2a and radial based functions as in Figure 3.2b. Activation functions can also be knowledge based. A knowledge based activation function uses an

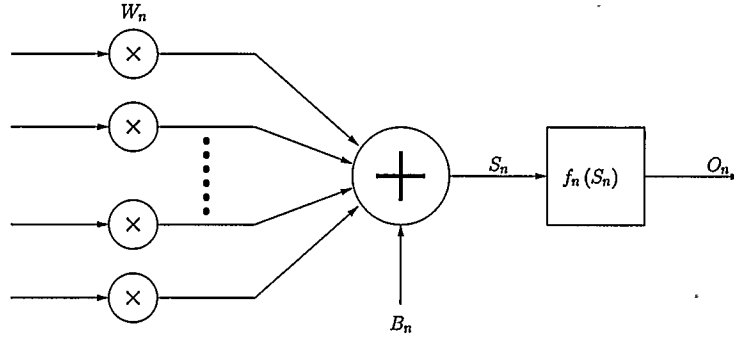


Figure 3.1: Node Structure

empirical model to combine the inputs. Knowledge based networks use the flexibility of neural networks to enhance already known empirical models [6]. Recently the use of wavelet theory has been combined with neural networks [7, 8].

The flexibility in the neuron comes from the parameters used in the input combination, and the ability to adjust them to change the node's input to output response. For example, if the input combination function was the dot-product function, the parameters to be adjusted would be the weights applied to the input and the overall node bias,  $B_n$ .

### 3.1.2 Network Structure

The typical arrangement of a neural network is a series of neurons arranged into layers. Each layer is then fully connected to the other layers. These layers can be placed in any fashion, including in a combination of parallel and serial. Recurrent networks can be formed where these layers feed back to previous

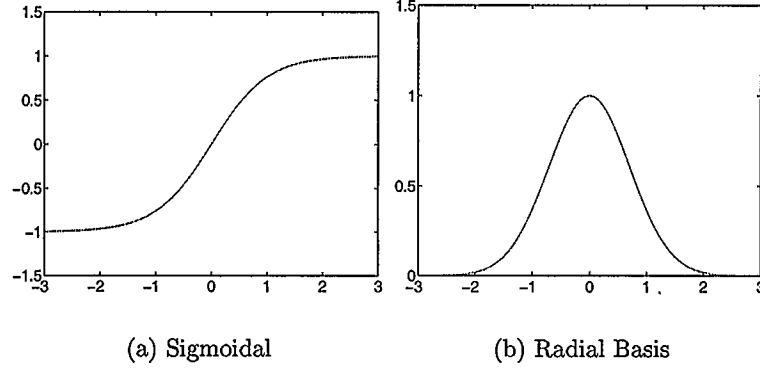


Figure 3.2: Activation Functions

layers forming feed back loops [9]. Feed-forward neural networks are a subclass where the layers are connected serially. The input is provided to the first layer, each layer will only feed to the next layer, and the last layer will provide the output from the network. The category of networks that this thesis is most concerned with is a type of feed-forward neural network known as the single layer perceptron, shown in Figure 3.3.

Since the networks discussed in this paper are only required to provide a single output, the output layer is composed of a single neuron. If the output neuron uses a sigmoid or radial function, the output of the network will be fundamentally limited to the range of values that the activation function can provide. In order to achieve an output that can be in any arbitrary range, a linear activation function is typically used in the output layer. Usually the activation function is a linear input to output relationship with no gain. Since the activation function output is not bounded, the parameters in the input

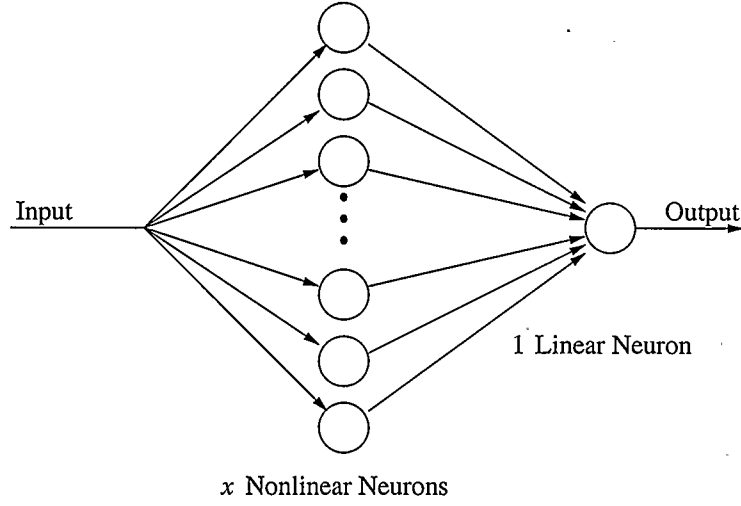


Figure 3.3: Single Layer Perceptron

combination function can be adjusted to provide any arbitrary output level.

### 3.2 Output Calculation

For the duration of this paper the input combination function will be defined to be the dot-product function as defined in (3.1).

$$S_n = W_n O_{n-1} + B_n \quad (3.1)$$

The symbols used are shown in Table 3.1.

The output of a node is defined by

$$O_n = f_n(S_n) \quad (3.2)$$

Symbol	Definition
$N_n$	The number of nodes in layer $n$
$S_n$	The combination of inputs for layer $n$ that will be used as the input to the activation function. Dimensions $(N_n \times 1)$
$W_n$	A matrix of weights for layer $n$ . Dimensions $(N_n \times N_{n-1})$
$O_n$	The matrix of outputs of the activation functions from layer $n$ , or for $n = 0$ , the inputs to the network. Dimensions $(N_n \times 1)$
$B_n$	A matrix of biases for layer $n$ . Dimensions $(N_n \times 1)$
$T_n$	A matrix of target outputs for layer $n$ . Dimensions $(N_n \times 1)$
$E_n$	A matrix of output errors for layer $n$ . Dimensions $(N_n \times 1)$

Table 3.1: Symbols Used

where  $f_n$  is the activation function for layer  $n$ . The calculation is completed iteratively beginning with the first layer,  $n = 1$ , and continuing to the last layer. Each layer first uses (3.1) to combine the inputs from the previous layer with the weights and biases of the current layer to provide  $S_n$ . Then  $S_n$  is used in (3.2) to determine the output of the layer,  $O_n$ . The output from layer  $n$  then becomes the input to layer  $n+1$ . The outputs from the last layer become the output from the network.

### 3.3 Training

For the neural network to model any useful system, the parameters must be set accordingly. The process of adjusting weights and biases in the network

to model a system is known as training. Training methods can be mostly divided into two type of training, supervised and unsupervised. Supervised training uses the output of the network and a known expected value to adjust the parameters. Unsupervised training does not use an expected signal, and the network derives the adjustments from the input and output directly.

### 3.3.1 Backpropagation

Backpropagation, also known as the generalized delta rule, is an algorithm that allows the change in weights inside the network to be calculated from the difference between the result from the network and an expected value. Since the algorithm requires an expected or target value, any training algorithm that uses backpropagation can be classified as a supervised learning algorithm. Backpropagation works by propagating the error relative to an expected value at the output, backwards through the network. This can be represented by the derivative of the output error with respect to the parameters for that layer  $\frac{\partial E_n}{\partial W_n}$ . The value  $E_n$  is sometimes known as the performance function. The derivative can not be calculated directly, however application of the chain rule can break the derivative into a form that can.

The first break is shown in (3.3).

$$\frac{\partial E_n}{\partial W_n} = \frac{\partial E_n}{\partial S_n} \frac{\partial S_n}{\partial W_n} \quad (3.3)$$

Since  $S_n$  is the combination of the output from layer  $n - 1$  before it reaches

the activation function which, in this derivation, is defined as the dot product of the weights  $W_n$  and the outputs from the previous layer  $O_{n-1}$ ,  $S_n$  can be expressed as

$$S_n = W_n O_{n-1} + B_n \quad (3.4)$$

From this the derivative is as shown in (3.5).

$$\frac{\partial S_n}{\partial W_n} = O_{n-1} \quad (3.5)$$

To make future calculations simpler,  $\delta_n$  is substituted for  $\frac{\partial E_n}{\partial S_n}$  and the result from (3.5) is substituted into (3.3), the problem simplifies to

$$\frac{\partial E_n}{\partial W_n} = \delta_n O_{n-1} \quad (3.6)$$

Then to calculate  $\delta_n$ , the chain rule is applied again in (3.7).

$$\delta_n = \frac{\partial E_n}{\partial S_n} = \frac{\partial E_n}{\partial O_n} \frac{\partial O_n}{\partial S_n} \quad (3.7)$$

The first partial derivative is the derivative of the error at the output of the layer with respect to the combination of the inputs and weights before the activation function. A common error function is the squared error shown in (3.8).

$$E_n = \frac{1}{2} (T_n - O_n)^2 \quad (3.8)$$

where  $T_n$  is the target value for layer  $n$ . The  $\frac{1}{2}$  multiplier simplifies the equations, as the derivative with respect to the output reduces to

$$\frac{\partial E_n}{\partial O_n} = -(T_n - O_n) = O_n - T_n \quad (3.9)$$

The relation between  $S_n$  and  $O_n$  is defined by the activation function  $f_n$ .

$$O_n = f_n(S_n) \quad (3.10)$$

The derivative of the output becomes the derivative of the activation function with respect to the input,  $S_n$ .

$$\frac{\partial O_n}{\partial S_n} = \frac{\partial}{\partial S_n} f_n(S_n) \quad (3.11)$$

Combining the derivatives thus far yields

$$\delta_n = (O_n - T_n) \frac{\partial}{\partial S_n} f_n(S_n) \quad (3.12)$$

resulting in the final definition of  $\frac{\partial E_n}{\partial W_n}$  shown in (3.13).

$$\frac{\partial E_n}{\partial W_n} = (O_n - T_n) \frac{\partial}{\partial S_n} f_n(S_n) O_n \quad (3.13)$$

For internal layers, a target is not available, and the error for that layer can not be calculated directly. As a result if layer  $n$  is an internal layer, the partial derivative  $\frac{\partial E_n}{\partial O_n}$  can not be calculated directly. The chain rule must be used again to break it up into derivatives that can be calculated.



$$\frac{\partial E_n}{\partial O_n} = \frac{\partial E_n}{\partial S_{n+1}} \frac{\partial S_{n+1}}{\partial O_n} \quad (3.14)$$

However  $\frac{\partial E_n}{\partial S_{n+1}}$  has already been calculated in the previous layer and defined as  $\delta_{n+1}$  in (3.7). Again, assuming that the inputs to a node are combined with the weights as a dot product as in (3.4), results in (3.15).

$$\frac{\partial S_{n+1}}{\partial O_n} = W_{n+1} \quad (3.15)$$

Combining results from (3.14), (3.15) and (3.11) to form the general case of  $\delta_n$  results in (3.16).

$$\delta_n = \frac{\partial E_n}{\partial S_n} = \frac{\partial E_n}{\partial S_{n+1}} \frac{\partial S_{n+1}}{\partial O_n} \frac{\partial O_n}{\partial S_n} = \delta_{n+1} W_{n+1} \frac{\partial}{\partial S_n} f_n(S_n) \quad (3.16)$$

Combining (3.16) and (3.6) gives the final result.

$$\frac{\partial E_n}{\partial W_n} = \delta_n O_{n-1} = \delta_{n+1} W_{n+1} \frac{\partial}{\partial S_n} f_n(S_n) O_{n-1} \quad (3.17)$$

This can be shown in an alternate form as

$$\frac{\partial E_n}{\partial W_n} = \delta_n O_{n-1} \quad (3.18)$$

By applying (3.6) and (3.7) to the last layer of the network, then using Equations (3.18) and (3.16) on the last internal layer to the first internal layer, the algorithm can calculate the derivative of the error with respect to any given parameter.

The derivation thus far assumes that all values will be real, and contain no imaginary components. However, the backpropagation algorithm can easily be adapted to complex values with complex activation functions [10].

### 3.3.2 Gradient Descent

Gradient descent uses the partial derivative,  $\frac{\partial E_n}{\partial W_n}$ , to linearly adjust the weights along the n-dimensional error surface along the maximum slope.

$$\Delta W_n \propto -\frac{\partial E_n}{\partial W_n} = -\eta \frac{\partial E_n}{\partial W_n} \quad (3.19)$$

The parameter  $\eta$  is defined as the learning rate. The learning rate governs the size of the step taken down the minimum slope. If the region is linear the step size can be increased in order to move faster towards the minimum error with fewer steps. However if the learning rate is too large, and the error surface around the current location isn't linear, the weight step may increase the overall error. However lowering the learning rate also increases the number of steps required to converge on the minimum error. Since gradient descent algorithm uses the backpropagation algorithm which relies on a target value, gradient descent and its derivatives are classified as supervised training algorithms.

### Gradient Descent with Adaptive Learning Rate

Adaptive learning rate alters the learning rate to achieve the best of both worlds. After the weight change has been made, the performance is recalcu-

lated. If the performance increases, the learning rate is increased by a constant factor. If the performance is decreased, and the decrease exceeds a threshold, then the learning rate is decreased by a constant factor.

### **Gradient Descent with Momentum**

The error surface may contain several local minima and the global minima. To avoid settling into a local minima, momentum can be added to the gradient descent.

$$\Delta W_n(t) = \alpha \Delta W_n(t-1) - \eta \frac{\partial E_n}{\partial W_n} \quad (3.20)$$

where  $\alpha$  is the momentum factor. Since the next weight change is a factor of the previous weight change, the adaptation path will tend to overshoot the local minima, possibly carrying forward towards a global minima, as illustrated in Figure 3.4. However, a larger momentum factor will increase the convergence time, as time will be spent oscillating towards the minima.

### **Gradient Descent with Adaptive Learning Rate and Momentum**

By combining adaptive learning rates, and momentum, the best of all worlds can be achieved. The neural network will converge on a minima faster due to the adaptive learning rate, and avoid local minima by overshooting due to momentum.

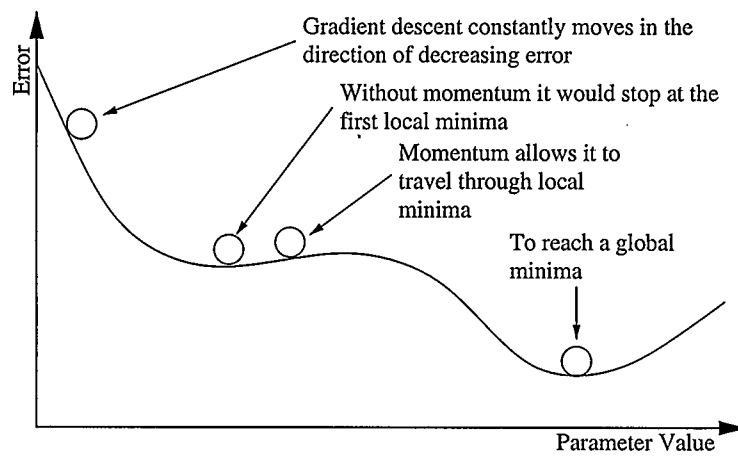


Figure 3.4: Gradient Descent with Momentum

## Chapter 4

### Linearization Techniques

As introduced in Chapter 2, power amplifiers introduce distortion into the amplified signal. The distortion can cause, but is not limited to, spectral regrowth, constellation warping and inter-symbol interference. Measures must be taken in order to reduce these effects. The methods of reducing the distortions are known as linearization. In this chapter, various linearization techniques will be introduced, and their advantages and disadvantages will be evaluated.

Many linearization techniques have been published in the literature. In general linearization can be categorized by where the correction is introduced in relation to the amplifier, before, during or after. Feedback and predistortion introduce the correction before the amplifier, while feed-forward and LINC introduce it after the amplification. Envelope elimination and restoration techniques introduces the correction during the actual amplification.

Linearization techniques are evaluated on several points of merit including, but not limited to:

- Effectiveness is a primary consideration. Usually measured by how much the intermodulation products are reduced as a result of the linearization technique.

- The bandwidth the system is able to linearize while maintaining a certain level of reduction.
- Additional hardware increases system cost and complexity.
- Implementation issues must be considered including sensitivity to parameter changes, environment and aging.

## 4.1 Output Power Reduction

The simplest form of amplifier linearization is output power reduction. Since for lower output powers the amplifier is linear, the input power is tuned so that the signal remains in the linear region. The other way that a signal can be kept in the linear region of the amplifier is to use a larger amplifier, thereby increasing the maximum power level that will be amplified linearly.

### Advantages

The primary advantage to this method is that it is very simple. Reducing the input power only involves placing an attenuator on the input to the amplifier. Reducing the input power does reduce the intermodulation products significantly since the third order intermodulation products will fall 3dB for every 1dB of reduced output power.

### Disadvantages

By backing off the output power to the amplifier, the output signal may not be at an adequate level for transmission. More active methods of linearization, which are discussed below, allow the input signal to enter the nonlinear region, and still maintain a linear output. As a result while this method is simple, without increasing the size of the amplifier, it does not allow for the power levels that other linearization methods do. To obtain larger output powers with linear amplification using this method a larger, oversized amplifier is required.

The primary purpose of this method is to reduce the nonlinear distortions. As linear distortion is independent of the input power, this method does not address linear distortions introduced by the amplifier. While small bandwidth signals will not be affected as the frequency response can be assumed to be flat across the operational region, wideband signals will experience inter-symbol interference. This method can be augmented with equalization to compensate for the linear distortions.

## 4.2 Feed-Back

Feed-back linearization, patented in 1938 by H.S. Black [11], is commonly used in low frequency applications. The system works by scaling the output of the amplifier and subtracting it from the input. By introducing a phase shifted version of the error to the input of the amplifier, the distortion introduced by

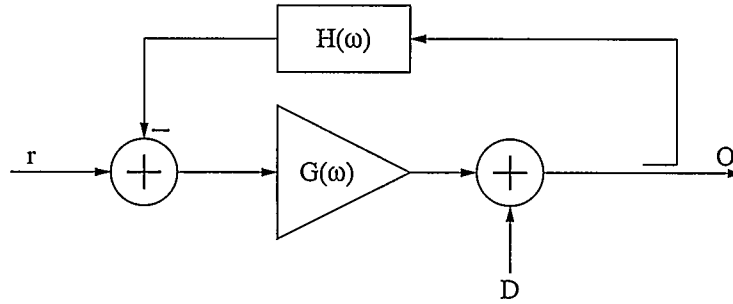


Figure 4.1: Feedback Linearization

the amplifier will be reduced. Figure 4.1 shows a block diagram. If we assume that the distortion introduced by the amplifier is summed with the signal at the output of the amplifier, then the output is defined as

$$O = \frac{rG(\omega)}{1 - G(\omega)H(\omega)} + \frac{D}{1 - G(\omega)H(\omega)} \quad (4.1)$$

This shows that the overall distortion is reduced by  $\frac{1}{1 - G(\omega)H(\omega)}$ , however the gain applied to the input signal is also reduced by the same amount.

### Advantages

The feedback system is very simple to implement, requiring very little additional hardware.

### Disadvantages

Since the output of the amplifier is being subtracted from the input, feedback systems will pay a gain penalty. As (4.1) shows, the gain of the system will not only be dependent on the gain of the amplifier, but the attenuation placed in



the feedback circuit. To compensate for this an alternate system was proposed in [12] that removes the input signal from the feedback path. Since the feedback signal does not contain the input signal, the overall gain of the system is not affected.

As frequency increases, system stability and bandwidth must be considered. For feedback to work, the combination of the input signal and the feedback signal must be subtractive. This is accomplished by adding an odd multiple of  $\pi$  radians phase rotation to the feedback signal. Delay lines are commonly used to form the proper phase shift. The approximate phase shift for a delay line of length  $L$  is

$$\theta = \frac{2\pi Lf}{c} \quad (4.2)$$

As the operational frequency increases the phase shift introduced by the feedback path linearly increases. The result is that a delay line will only provide the correct phase shift for a single frequency. As the signal moves away from that calibrated frequency, the phase shift will increase or decrease linearly with frequency. This results in the correct cancellation at the input of the amplifier for only a narrow band of frequencies which becomes narrower as the required phase shift increases.

At higher frequencies it becomes difficult if not impossible to have an electrical length that corresponds to  $\pi$  radians of phase, and the feedback path must become  $3\pi$  or even  $5\pi$  radians. The effect of the delay increasing is that

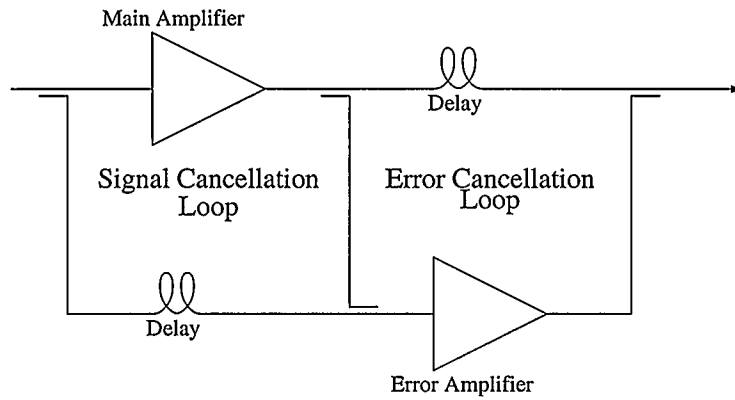


Figure 4.2: Feed-Forward Linearization

the open-loop gain will start to increase for certain frequencies. If the open loop gain for a frequency exceeds unity, the system will become unstable and oscillate.

### 4.3 Feed-Forward

The feed-forward method linearizes the signal by removing the error from the amplified signal. This is accomplished by generating the error signal and mixing it back into the output of the amplifier in such a way that the error is canceled out. The general feed-forward system is shown in Figure 4.2.

The system is composed of two loops, the signal cancellation loop and the error cancellation loop. The signal cancellation loop is responsible for canceling the signal out of the distorted amplified signal, leaving only the error. Then in the error cancellation loop the error is used to cancel the error from the

amplified signal.

In the signal cancellation loop the input signal is first amplified with the main amplifier, and then split. Some of the power in the amplified signal enters the error cancellation loop, while part of the signal is attenuated until it is at the same level as the input signal. A phase shifted version of the original undistorted signal is added to the attenuated signal. The combination of the distorted signal and the phase shifted undistorted signal cancel out the desired signal, leaving only the error introduced by the amplifier distortion.

The error cancellation loop then amplifies the error signal up to an appropriate level with the error amplifier and combines it with a phase shifted version of the distorted signal output from the main amplifier with the amplified error signal. This combination cancels out the error leaving only the undistorted amplified version of the input signal.

### **Advantages**

Feed forward systems are extremely effective and can reduce intermodulation products by 40dB or more. Since no feedback paths exist, stability is not an issue. The linear distortions are also present in the error signal, and when the error is subtracted from the output signal the linear distortions are also corrected.

### Disadvantages

There are several disadvantages associated with feed-forward amplification. A feed-forward system requires several additional RF components, including a second amplifier. The error signal contains a much higher peak to average ratio than the input signal, a larger bandwidth, and the required output power of the error amplifier is not significantly lower than the main amplifier [13]. In general the error amplifier is required to have similar capabilities to the main amplifier, so a second main amplifier is used.

The RF equipment must also be calibrated very carefully. A. Javed and B. Syrett [14] showed that in order to achieve 30dB of cancellation that an amplitude balance of 0.3dB and a phase balance of 2 degrees is required. This generally makes the system installation specific, and requires that it be recalibrated if any part of the system changes. To combat this an adaptive controlled feed-forward system was proposed in [15] which has become a commonly used technique for commercial systems.

The stringent requirements on the phase balance also have a direct impact on the amount of cancellation for wider bandwidth signals. As with feedback systems, delay lines are used to adjust the phase of the input signal and amplified signal to achieve the desired cancellation. From (4.2) the phase applied to the signal will be frequency dependent. Once the phase for a specific frequency drifts by more than 2 degrees, the cancellation will be at best 30dB. As the distance from the calibrated frequency increases, the phase will continue

to drift further, and cancellation will suffer. A.M. Smith and J.K. Cavers proposed a method that has two delay paths through the signal cancellation path in [16], which significantly widened the bandwidth, at the cost of additional RF hardware and additional tuning required.

#### 4.4 LINC

Linear amplification using nonlinear components, or LINC, was introduced in [17]. LINC works under the assumption that amplifiers do not distort constant envelope signals. To exploit this assumption the input signal, generally with both phase and amplitude information, is split into two phase modulated constant envelope signals. Since the new signals have a constant envelope, efficient nonlinear amplifiers can be used without generating intermodulation products. Figure 4.3 contains a block diagram of a LINC setup. The component separator generates the two constant envelope signals in such a way that when they are combined together, they produce an amplified version of the input signal. This is accomplished by generating an error term  $e(t)$  and then generating the two output signals as

$$S_1 = S_i(t) + e(t) \quad (4.3)$$

$$S_2 = S_i(t) - e(t) \quad (4.4)$$

such that  $S_1$  and  $S_2$  have a constant envelope. Thus when they are combined

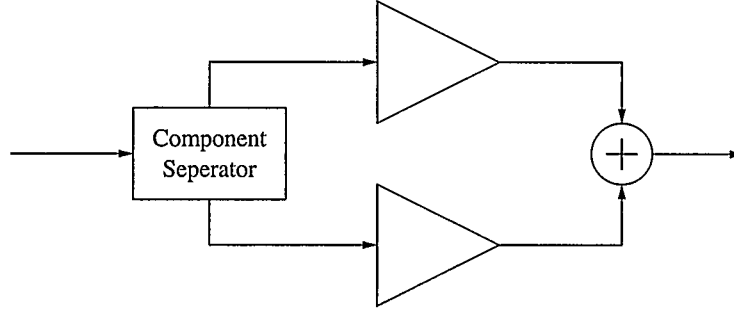


Figure 4.3: LINC Transmitter

$$S_o = S_1(t) + S_2(t) = 2S_i(t) \quad (4.5)$$

the combination results in a linear output.

### Advantages

Since the amplifier's input signal is a constant envelope signal, the amplifiers can be chosen on output power and efficiency, without regard to linearity. The cost of the additional amplifier is generally more than offset by the ability to use cheaper amplifiers.

### Disadvantages

Like feed-forward linearization, LINC is difficult to implement. The cancellation will be very sensitive to any differences in either arm. Each arm must exhibit both the same phase and magnitude response for the system to work properly. The amplifier specifications must also be broad bandwidth since the input signals  $S_1$  and  $S_2$  contain more bandwidth than the input signal  $S_i$  as a

result of the separation. The severity of the signal spreading is dependent on the characteristics of the input signal [18]. Modulation schemes that do not have zero crossings such as  $\pi/4$  DQPSK do not exhibit the signal spreading as much as 64-QAM which produces zero crossings.

## 4.5 Envelope Elimination and Restoration

Envelope elimination and restoration, also known as the Kahn technique, is not in itself a linearization technique, however it is used with conjunction with other techniques such as predistortion to provide linear amplification. It works much like LINC in that the signal that is amplified is reduced into a constant envelope signal. This is accomplished by separating the input into the magnitude and phase components. The phase is modulated onto an RF carrier producing a constant envelope signal at the desired frequency. The signal is then amplitude modulated and amplified by adjusting the DC power supply to the amplifier according to the magnitude of the input envelope. Since the amplifier power supply has a direct impact on the amplitude of the output signal, which impacts gain, modulating the DC power supply in turn modulates the output signal. The overall system is shown in Figure 4.4.

### Advantages

Like LINC, envelope elimination and restoration amplifies constant envelope signals with nonlinear amplifiers. This allows the use of lower cost, efficient,

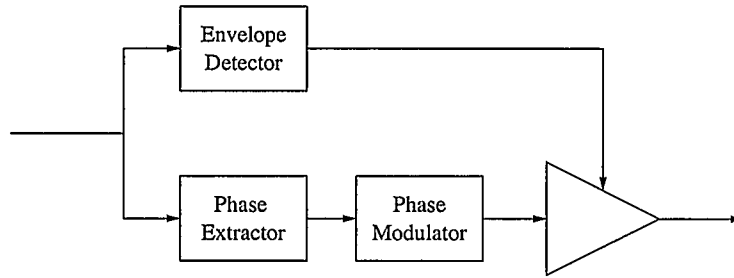


Figure 4.4: Envelope Elimination and Restoration

highly nonlinear amplifiers. Unlike LINC, there is no combination of two high power signals. This allows for much higher efficiencies than a LINC setup. This method was demonstrated by F.H. Raab in [19]. The results show amplification with 40 - 50% efficiency at 20 Watts of output, with excellent linearity.

### Disadvantages

Unfortunately envelope elimination and restoration has several disadvantages. One difficulty is how the output signal becomes amplitude modulated. Using the DC power supply for the power amplifier to amplitude modulate has several issues including linearity, dynamic range and bandwidth. The relation of the DC power supply level to output amplitude can be very nonlinear. To overcome this other linearization techniques have to be used. Predistortion was used in [19]. For high peak to average ratio signals the dynamic range of the DC power supply level to output amplitude may become an issue if not a limitation. Since the signal is separated into magnitude and phase compo-



nents, the magnitude component's bandwidth will be several times larger than the input signal, requiring the DC supply to have a large baseband bandwidth. In general, because of all the requirements and the non-ideal characteristics of an envelope elimination and restoration system, it is one of the most difficult linearization techniques to implement.

## Chapter 5

### Predistortion

Predistortion has become a hot research topic in the last decade. With the power of digital signal processors increasing and their price decreasing, some previously proposed predistortion methods are now becoming possible. One of the earliest publications on predistortion is Davis' patent in 1981 [20]. Since then most publications either reference Saleh's 1983 article [21] linearizing a QAM-64 modulated signal, or Cavers' 1990 article [22]. The novelty of Cavers' method was that very little processing power and memory were required in order to achieve excellent results. The technique was enhanced by Stapleton and Cavers in 1991 in [23]. Since then predistortion has become a very broad research topic with many variations.

#### 5.1 Theory

Predistortion is a method of compensating for amplifier distortion before the amplifier distorts the signal. By estimating what the distortion for a given signal will be, the system can estimate what correction will be required so that the output signal of the amplifier will be the desired signal. This correction is then applied to the signal before the amplifier. The general concept is to place a system before the amplifier, with a transfer characteristic that is

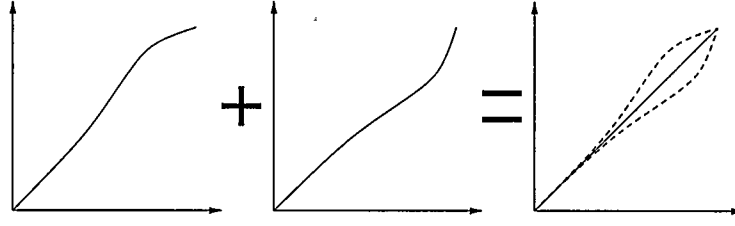


Figure 5.1: Predistortion

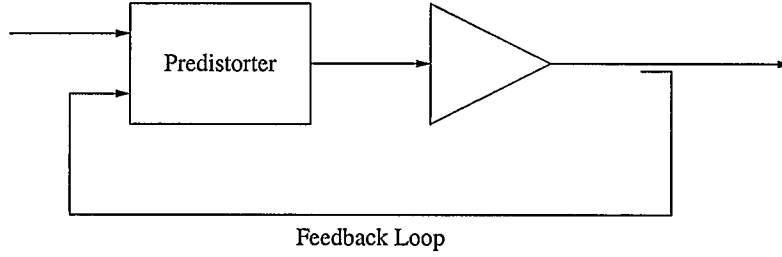


Figure 5.2: Predistortion Block Diagram

the inverse of the amplifier. The cascade of the predistortion system with the amplifier results in a system which is linear. Figure 5.1 illustrates this concept. Predistortion systems will generally follow the form of Figure 5.2.

### 5.1.1 Derivation

For complex input data, let  $X$  represent the complex input. The output of the amplifier will be  $Y$  such that

$$Y = \Psi(X) \quad (5.1)$$

where  $\Psi$  represents the complex transfer characteristic of the amplifier. Placing a predistortion function  $\Phi(X)$  into the system results in

$$Y = \Psi (\Phi (X)) \quad (5.2)$$

The goal of predistortion is to produce an output signal that is a scaled version of the input

$$Y = GX \quad (5.3)$$

Where  $G$  is a linear gain term. Substituting (5.3) into (5.2) and manipulating the equation a little results in

$$\Phi (X) = \Psi^{-1} (GX) \quad (5.4)$$

### **Polar Representation**

Amplifier distortion is commonly described in terms of polar coordinates. Modeling amplifiers in terms of amplitude and phase distortion has been shown to be very accurate for a wide range of amplifiers [2]. If the input is represented as magnitude and phase components

$$X = re^{i\theta} \quad (5.5)$$

The output of the amplifier becomes

$$Y = Re^{i(\rho+\theta)} \quad (5.6)$$

where  $R$  and  $\rho$  are calculated using

$$R = A(r) \quad (5.7)$$

$$\rho = \Theta(r) \quad (5.8)$$

The functions  $A$  and  $\Theta$  are the amplitude and phase transfer characteristics of the amplifier respectively. Predistorting the inputs results in an estimate for input that will result in the desired output

$$\hat{r}e^{i\hat{\theta}} = \Upsilon(r) e^{i(\Omega(r)+\theta)} \quad (5.9)$$

where  $\Upsilon$  is the AM/AM predistortion block and  $\Omega$  is the AM/PM predistortion block. Placing predistortion in the system and following the same derivation results in

$$\Upsilon(r) = A^{-1}(Gr) \quad (5.10)$$

$$\Omega(r) = -\Theta(\Upsilon(r)) = -\Theta(A^{-1}(Gr)) \quad (5.11)$$

### Cartesian Representation

If the input is represented by its in-phase and quadrature components in a Cartesian coordinate system

$$X = x_i + x_q \quad (5.12)$$

and the amplifier output becomes

$$Y = y_i + y_q \quad (5.13)$$

where

$$y_i = \Psi_i(x_i, x_q) = \text{Re} \{ \Psi(x_i + ix_q) \} \quad (5.14)$$

$$y_q = \Psi_q(x_i, x_q) = \text{Im} \{ \Psi(x_i + ix_q) \} \quad (5.15)$$

The function  $\Psi$  is the same characteristic from (5.4) but separated into real and imaginary components. The predistortion functions then become

$$\hat{x}_i = \Phi_i(x_i, x_q) = \Psi_i^{-1}(x_i, x_q) = \text{Re} \{ \Psi^{-1}(x_i + ix_q) \} \quad (5.16)$$

$$\hat{x}_q = \Phi_q(x_i, x_q) = \Psi_q^{-1}(x_i, x_q) = \text{Im} \{ \Psi^{-1}(x_i + ix_q) \} \quad (5.17)$$

## 5.2 Implementation

Predistortion implementations can take on several forms, depending on which stage the predistortion is done, which signal representation is used, and if the predistortion is adaptive, what is the work function.

### **Polar vs. Cartesian**

The choice of how to represent the signal divides predistortion into two categories, polar or Cartesian. Each has advantages and disadvantages. As can be seen from the derivation, the polar system benefits from a far simpler representation of the distortion. However input signals are usually provided in Cartesian in-phase and quadrature components rather than polar. If a polar predistorter is used, the input must be converted from Cartesian to polar, and the output must be converted back to Cartesian.

### **Baseband Vs. RF**

A predistortion function can operate on either the signal at baseband or at the RF modulated signal. Baseband predistortion offers the simplicity of being able to operate on a signal in either the analog domain or the digital domain. With modern DACs and ADCs that are capable of hundreds of megasamples or even gigasamples per second, wideband signals can be brought into the digital domain and predistorted in a DSP. The drawback to baseband predistortion is that the feedback signal must be demodulated back down to baseband. This can be avoided with techniques such as bandpass sampling [24], which has some noise penalties. While predistorting the RF modulated signal doesn't require that the output of the amplifier be demodulated back down to baseband, it currently precludes the possibility of digital predistortion. Currently DACs, ADCs, and DSPs are not fast enough to perform digital RF predistortion. Attempts have been made at analog predistortion such as the RF Cuber [25,

26], but they are difficult to make adaptive.

### **Symbol Predistortion Vs. Signal Predistortion**

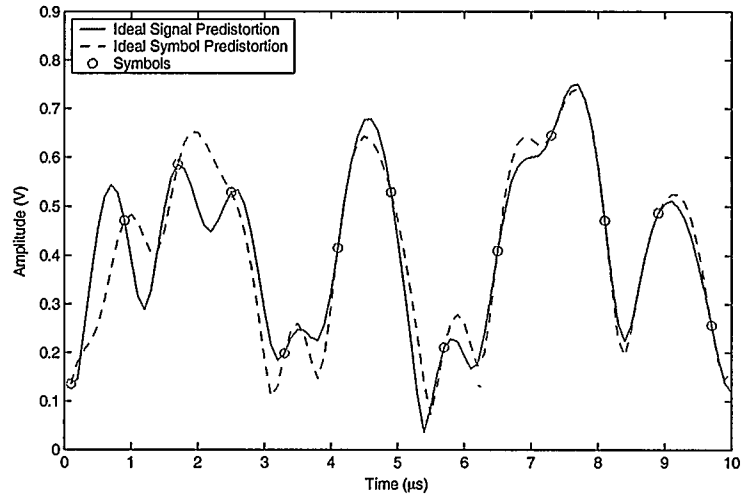
Assuming digital baseband distortion, the predistortion can either work with the raw data symbols, or the signal after the pulse shaper. If the predistortion operates on the pulse-shaped signal, the signal must be sampled well above the symbol rate. This raises the throughput requirements of signal predistortion over symbol predistortion. The drawback with symbol predistortion is that the transition between symbols is not addressed. Symbol predistortion makes sure that the symbols will be transmitted perfectly and the resulting eye-diagram will have wide open eyes. However the transitions are still left to the pulse shaper.

Figure 5.3 illustrates this problem. Figure 5.3a shows the amplified output with signal and symbol predistortion. Both methods pass through the symbols perfectly, however the routes taken are very different. Figure 5.3b shows the power spectral density for both types. The end result with symbol predistortion is that the constellation is transmitted clearly, but intermodulation products will be far higher than the signal predistortion case.

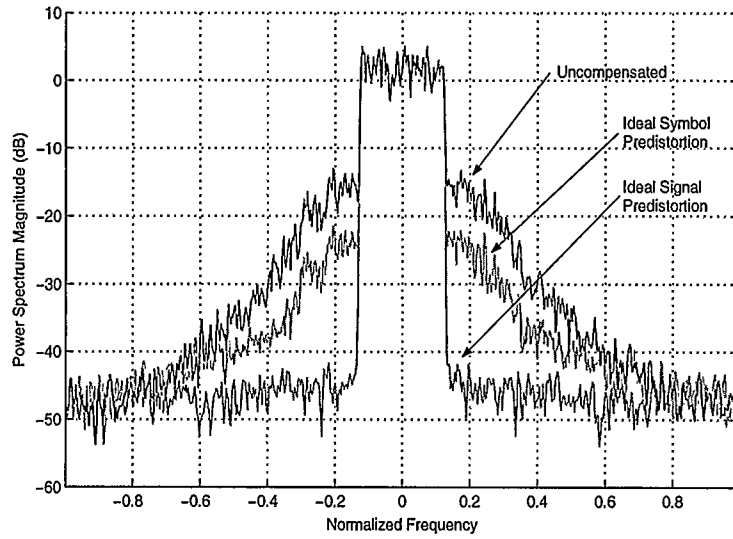
### **Maximum Linearity Vs. Minimum Intermodulation Products**

The work function of a predistorter further categorizes the various predistortion techniques. A work function is performance or error function that an adaptation algorithm will either maximize or minimize to achieve the goal. A





(a) Time Domain



(b) Frequency Domain

Figure 5.3: Symbol Predistortion Vs. Signal Predistortion

predistortion system can have its performance function tuned to either maximum linearity, or minimum out-of-band distortion performance [27]. The difference usually comes from what the adapting function is observing. In [28, 23] the predistorter is updated based on the measured ACPR, while in [29] the overall linearity of the signal is used as the work function.

In theory, a maximally linear amplifier will not exhibit any intermodulation products, however they will result from any deviation in linearity. If the sole criteria for adaption is the ACPR, then the intermodulation products will be heavily reduced, but the in-band distortion could still be present. The result is that maximally linear predistorters will have good in-band and adjacent channel performance, and predistorters built to minimize intermodulation products will have better ACPR performance with worse in-band distortion performance.

### 5.3 Limitations

Unlike feed-forward linearization and other linearization techniques that introduce power at the output of the amplifier, predistortion does not introduce any new power. As a result, the overall gain of a predistortion system has to remain at or below the amplifier's gain for any given input value. Figure 5.4 illustrates the constraints of predistortion.

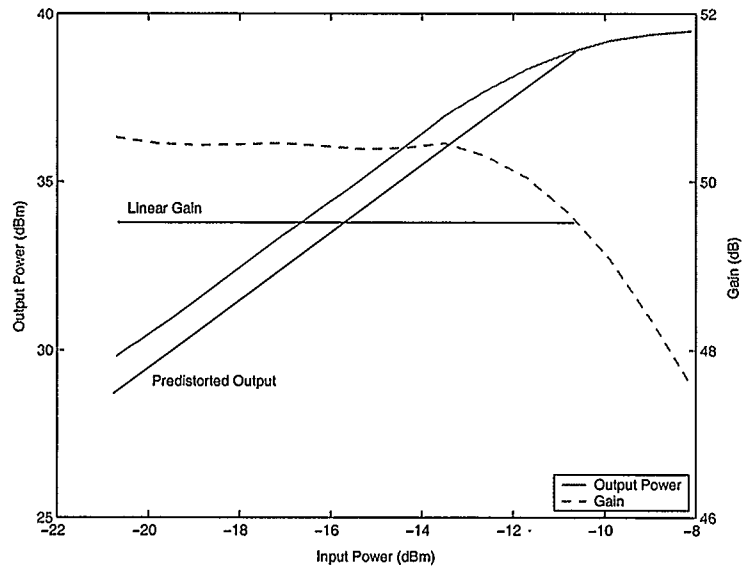


Figure 5.4: Predistortion Limitations

## 5.4 Advantages

Since the predistortion occurs before the amplifier, not as much additional RF equipment is required when compared to other linearization techniques. If the system is made adaptive, the only additional RF hardware required is an output coupler and a demodulation stage. The other components of a predistortion system are less expensive baseband components and include DACs and ADCs, and the actual predistortion block which is usually a digital signal processor.

## 5.5 Disadvantages

The primary disadvantage of predistortion is the computational complexity involved. Some systems are implemented as large multidimensional tables, and interpolation which involves a lot of memory. In the systems proposed by Cavers [22] and Stapleton and Cavers [23], the processing power and memory requirements were both drastically reduced. The issue of computational complexity is also being alleviated with the advancement of high speed digital signal processors and FPGAs.

Until the 90's, the subject of amplifier memory in predistortion was not addressed in literature. In 1993, Benvenuto, Piazza and Uncini published a paper using a neural network to perform predistortion, which took into account the memory of the system it was to linearize [30]. Later in 2001, Opere addressed the subject in his master's thesis [1].

One outstanding research issue with predistortion is the issue of broad bandwidth. As the bandwidth of the input signal increases, the effectiveness of the predistortion tends to decrease. This subject will be further addressed in Chapter 8.

## Chapter 6

### System Overview

The predistortion system proposed in this thesis is detailed in this chapter. First some background is provided about the model used for the amplifier, and how that relates to predistortion. Then the system architecture is described in detail.

#### 6.1 Amplifier Model

A common approach for designing a predistortion system is to determine an accurate model for the amplifier and then inverting the model. Several methods to model amplifiers exist, with varying levels of accuracy and complexity.

One of the most common ways to model an amplifier is with a simple lookup table [21]. This scheme works with two tables, one for the AM/AM distortion and one for the AM/PM distortion. The tables are populated with the data from a one-tone test. From these tables the output value can be interpolated from any input value. The disadvantage is that this is only accurate for a narrow band system that does not exhibit linear distortions or memory.

A Volterra series is a popular method to model a system with memory [14]. The primary disadvantage of using a Volterra series model for predistortion is that a direct inverse does not exist. Approximations can be made using

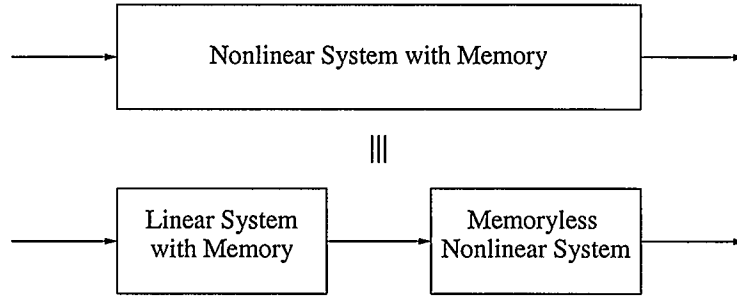


Figure 6.1: Nonlinear Systems with Memory

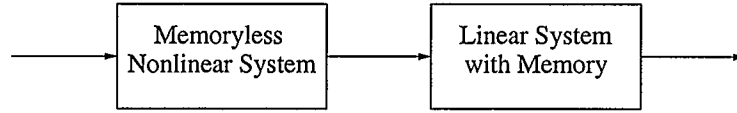


Figure 6.2: A Hammerstein System

polynomials [31] and the  $p$ th-order inverse approximations, both of which are very complicated however.

Modeling nonlinear systems with memory can be very difficult. One method to simplify the process is to break the system into two pieces, a nonlinear system without memory, and a linear system with memory placed before the nonlinear system [32] as shown in Figure 6.1. This type of model is known as a Wiener model and was proposed in [33].

The advantage of modeling the amplifier using a Wiener model is that the corresponding predistortion system is a nonlinear system without memory followed by a linear system with memory. This structure, known as a Hammerstein system, is shown in Figure 6.2.

## 6.2 System Structure

In this system two stages of correction will be applied, linear and nonlinear. Both stages will have their performance function tuned for maximum linearity. The block diagram for the system is given in Figure 6.3. The input to the system is assumed to be a complex baseband signal. Following the Hammerstein model, the signal is first passed through a neural network to compensate for the memoryless nonlinear distortion. Then a linear FIR filter is used as the linear system with memory. From there the signal is converted to an analog baseband representation, modulated to the desired carrier frequency and amplified.

The system receives its feedback from the output of the amplifier, which is demodulated, and fed back into two training blocks. One block is responsible for isolating the linear distortion while the other is responsible for isolating the nonlinear distortion. All of the blocks are discussed in further detail below.

### 6.2.1 Nonlinear Predistortion Networks

The nonlinear predistortion networks compensate for the static memoryless nonlinear response of the amplifier. The input to the system is  $x(t)$ , and output a nonlinear compensated signal  $v(t)$ . The response of this block should represent the inverse response of the amplifier as described in (5.4).

As introduced in Section 5.1.1, predistortion can be performed in Cartesian or polar coordinates. The polar coordinate models are far easier to implement,

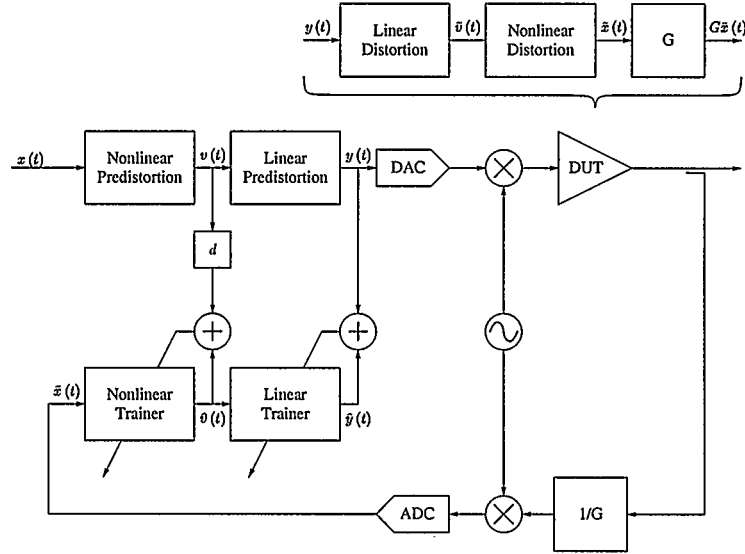


Figure 6.3: System Block Diagram.

but require extra conversions between polar and Cartesian coordinates. These conversions can easily become a bottleneck in any system. However since an actual system is not being implemented in this thesis, computational complexity is not a concern. Thus polar coordinates will be used for their implementation simplicity. The predistortion block is based on polar coordinates and is shown in Figure 6.4. The block diagram shows the implementation of (5.10) and (5.11) in a physical form using neural networks.

Neural networks are excellent at modeling nonlinear systems as black boxes, knowing only the input and output. Since the amplifier is being modeled as a linear and nonlinear system, the systems can not be separated and studied individually.



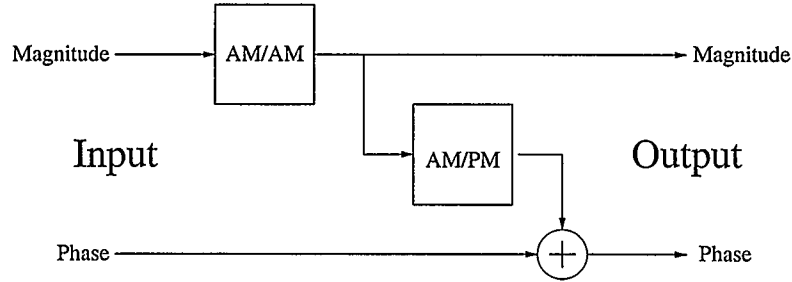


Figure 6.4: Nonlinear Predistortion

The characteristics of the networks are transferred from the nonlinear training block, which is described in 6.2.3. The networks receive their weights and biases periodically from the nonlinear training networks. This periodic update allows the networks to continuously operate on the input signal while new weights and biases are being calculated.

### 6.2.2 Linear Predistortion

The linear section is composed of a complex FIR filter. The input,  $v(t)$ , is the output of the nonlinear predistorter, and the output  $y(t)$  becomes the input to the amplifier. Like the nonlinear predistortion block, the tap coefficients are adjusted periodically by a training block, in this case the linear training block described in Section 6.2.4. The goal of the linear predistortion block is to compensate for the memory effects of the amplifier.

### 6.2.3 Nonlinear Training Block

The nonlinear training networks are responsible for modeling the inverse of the memoryless nonlinearity in the amplifier, and updating the weights and biases of the nonlinear predistortion networks. The nonlinear system in the amplifier can be defined as

$$\tilde{x}(t) = \Psi_{NL}(\tilde{v}(t)) \quad (6.1)$$

where  $\Psi_{NL}$  is the transfer function of the nonlinear component in the amplifier. The inverse characteristic is the relationship

$$\tilde{v}(t) = \Psi_{NL}^{-1}(\tilde{x}(t)) \quad (6.2)$$

Once the system has reached equilibrium the output of the amplifier can be expressed as  $G\tilde{x}(t)$ , a very mildly distorted version of the input with a linear gain. The feedback attenuation removes the gain leaving only a distorted version of the input. The signal  $\tilde{v}(t)$  will be approximately equal to the nonlinearly compensated signal  $v(t)$ . Thus the nonlinear block uses the signals  $\tilde{x}(t)$  and  $v(t)$  to learn the inverse of the nonlinearity in the amplifier.

The inputs to this block are queued and the training algorithms can be performed offline while the nonlinear predistortion networks continue to operate with the last set of weights and biases. The delay  $d$  is critical to align the two input signals. If the delay is misaligned, the relationship between the two signals will become overly complicated and the neural networks will have a

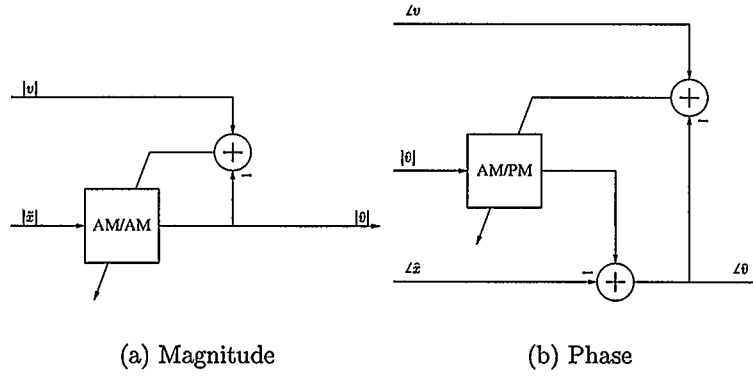


Figure 6.5: Nonlinear Training Block

more difficult time learning. Gradient descent with adaptive learning rate and momentum method defined in Section 3.3.2, is used to train the networks. The output of the networks in training is  $\hat{v}(t)$ , which is an estimate of the internal signal within the amplifier  $\tilde{v}(t)$ , which in turn is a mildly distorted version of the input to the linear predistorter  $v(t)$ . Thus the difference between  $\hat{v}(t)$  and the actual input to the linear predistorter  $v(t)$ , is used as the error term in the backpropagation. The internal layout for the block is given in Figure 6.5. The tuning values are used as error terms in the backpropagation algorithm described in Section 3.3.1.

#### 6.2.4 Linear Training Block

The linear training block is used to isolate the linear distortion of the amplifier. The linear distortion within the amplifier is defined as the relationship

$$y(t) = \Psi_{LIN}(\tilde{v}(t)) \quad (6.3)$$

where  $\Psi_{LIN}$  is the linear distortion component in the amplifier, and the inverse is

$$\tilde{v}(t) = \Psi_{LIN}^{-1}(y(t)) \quad (6.4)$$

Since the nonlinear training block has generated  $\hat{v}(t)$ , an estimate of the linearly distorted signal  $\tilde{v}(t)$ , the linear training block can use this and  $y(t)$  to estimate the linear distortion characteristic.

Like the nonlinear training block the linear trainer queues the input in order to perform the training offline. A LMS algorithm is used to train a set of FIR filter tap weights to minimize the error between  $y(t)$  and the estimate of the amplifier input  $\hat{y}(t)$ . Once a new set of filter coefficients have been determined, they are passed to the linear predistorter and the next repetition begins with the data that was queued during the offline calculations.

## Chapter 7

### Simulation Results

With the system architecture proposed in Chapter 6, a series of simulations were constructed to test the setup. To facilitate the tests, a model of the amplifier setup was constructed as well. The first section introduces the specifications of the amplifiers used in the tests. The second section determines the proper network sizes for the amplifier model, and the nonlinear predistortion networks. With this the nonlinear predistortion is tested. Then the linear system with memory is introduced to the model of the amplifier to complete the Hammerstein system. Finally, predistortion with memory is demonstrated showing an improvement in intermodulation products with increasing signal bandwidth.

#### 7.1 Device Under Test

The device used during subsequent tests is an amplifier setup composed of a Motorola MHW-1916 driving a Motorola MRF-282SR1. Since the two amplifiers will be treated as a single device under test, individual specifications are not necessary. The system as a whole exhibits 50.4 dB of gain, with a 1 dB compression point of -10.5 dBm input power. The maximum output power is 39.8 dBm or 11.2 Watts of power. For all experiments, the input power to

the amplifier was adjusted such that the peaks of the input signal were 4dB beyond the 1dB saturation point.

## 7.2 Network Size

The size of the neural network is a prime concern. A large network can be overly complex and require a large amount of computing power. Also, large networks can over-learn an input to output relationship, memorizing the relation instead of generalizing it [4].

As described in Section 6.1, the memoryless nonlinear distortion is assumed to be static with respect to frequency. As such, a signal with 1.25MHz of bandwidth was used for this test, since it will demonstrate virtually no linear distortions or memory effects.

To find the optimal size, a network with a given number of neurons is first trained on a known input to output data set, and then it is tested on a different input signal. QAM-16 signals were used to test the networks. Figure 7.1 plots the mean squared error of the model output against the number of nodes in the neural network for both the amplitude and phase transfer functions. For the amplitude transfer characteristic, the error generally decreases until 29 neurons. However, the drop in error is only one thousandth of a percent and drastically increases the computational requirements. The error in phase is fairly constant after five neurons. As such only five neurons each for amplitude and phase are all that is required.

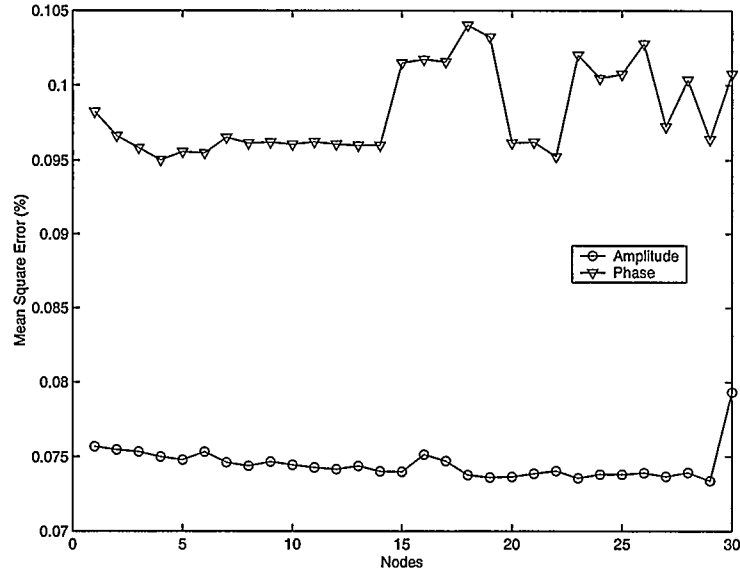


Figure 7.1: Network Size Vs. Modeling Accuracy

This measures the number of neurons required for modeling the amplifier, however the inverse of the amplifier needs to be modeled for the predistortion method. The inverse phase transfer function is the negative of the non-inverted function and will have the same complexity, however the amplitude transfer function is quite different. Figure 7.2 shows the mean squared error of the inverse amplitude transfer function. Comparing this plot to the previous plot shows that the inverse amplitude transfer function is more complex to model. While the forward transfer characteristic saturates to a constant value, the inverse characteristic has a vertical asymptote where it will approach infinity. This is reflected in the error curve in Figure 7.2 as the overall error is far higher than the previous case. However like the previous case it plateaus after

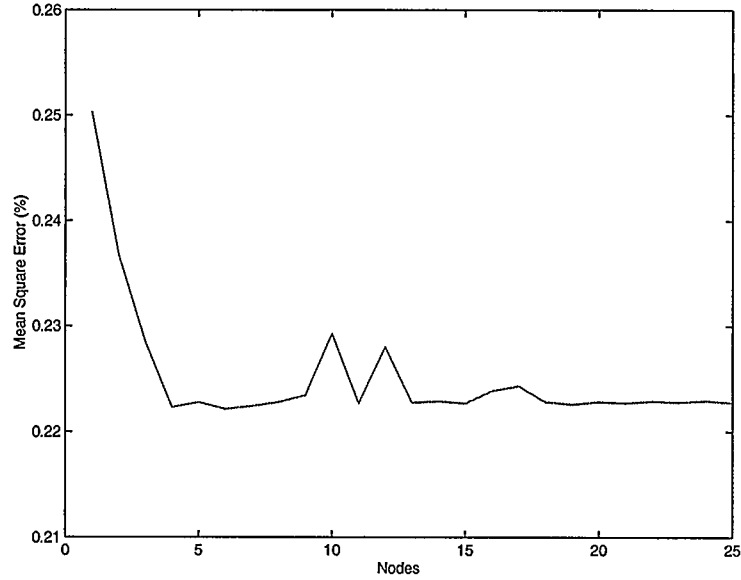


Figure 7.2: Network Size Vs. Amplitude Distortion Modeling Accuracy

only four nodes. Again only four or five neurons are required to achieve the minimum modeling error.

### 7.3 Nonlinear Predistortion Performance

With a model of both the amplifier and the inverse of the amplifier, we can simulate predistortion with the simple memoryless nonlinear model. Figure 7.3 shows the output spectrum of a predistorted waveform and an uncompensated waveform. The simulation shows that for this particular dataset 30 dB of intermodulation product reduction is possible.

Figure 7.4a shows the constellation of an QAM-16 waveform that is undis-



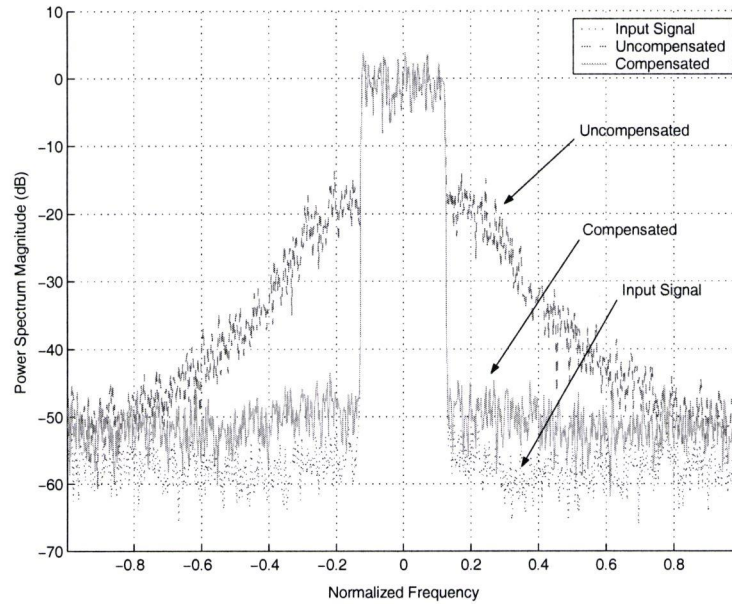


Figure 7.3: Nonlinear Predistortion Performance

torted. Figure 7.4b is an uncompensated waveform that was distorted by the amplifier model developed in Section 7.2. Notice that the corners of the waveform have been pushed inward, as the outer parts of the signal reach into saturation. This constellation warping will result in more difficulty determining what symbol was received, and a higher bit error rate. Figure 7.4c is a waveform that has been predistorted, which extends the outer corners to compensate in advance for saturation. Finally, Figure 7.4d is the amplified predistorted waveform. The compensated waveform results in a waveform that is only slightly distorted when compared to the original waveform.

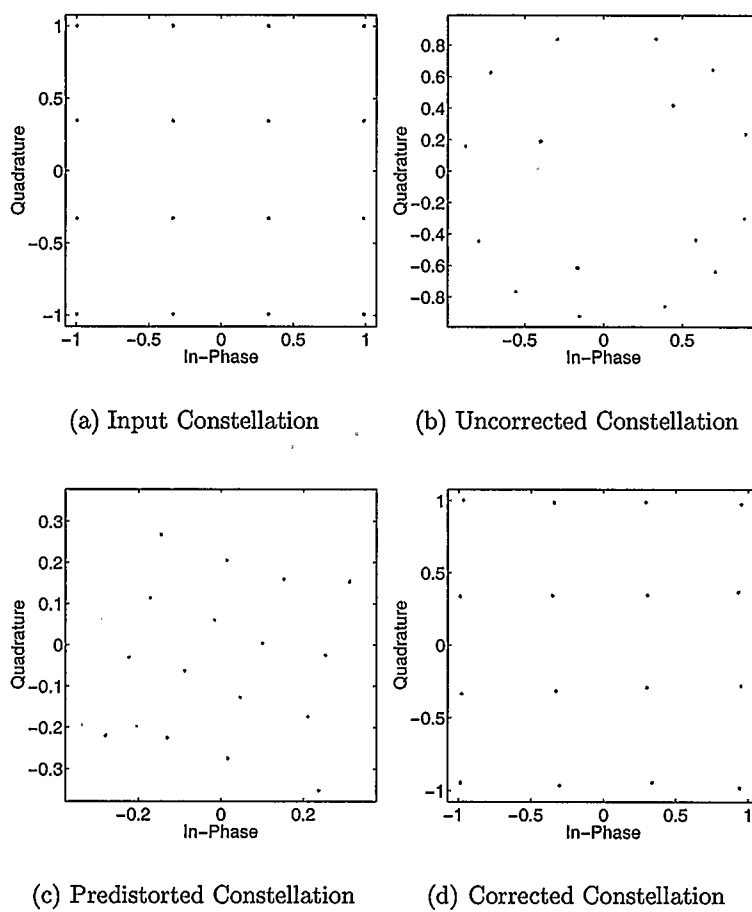


Figure 7.4: Distortion of QAM-16 Constellation

## 7.4 Nonlinear Model Performance

To test the performance of the nonlinear amplifier model against the actual amplifier, the QAM-16 test waveform was generated at several frequencies, and the corresponding input to the amplifier and its output were measured. Using data of signals at various bandwidths, the prediction from the nonlinear model was compared to the actual output. The result is shown in Figure 7.5. For low bandwidth applications the model performs very well, however as the bandwidth of the signal increases, the error from the model grows significantly. This error can be attributed to the linear distortions and the memory effects of the amplifier, which will be addressed in the next section.

## 7.5 Model Performance with Linear Distortion

Next the treatment of linear distortion needs to be introduced. Linear distortion become more of an issue as the bandwidth of the input signal increases. To simulate this, a linear predistortion block was added to the model which shapes the spectrum before it reaches the nonlinearity. The result of the two cascaded systems is a nonlinear system with memory. To determine the linear portion of the amplifier, the signal  $\tilde{v}(t)$  in Figure 6.3 has to be determined. If the nonlinearity is assumed to be static, independent of frequency, the inverse nonlinearity determined in the previous sections can be applied to the output waveform to obtain an estimate of  $\tilde{v}(t)$ . Figure 7.5 compares the predicted

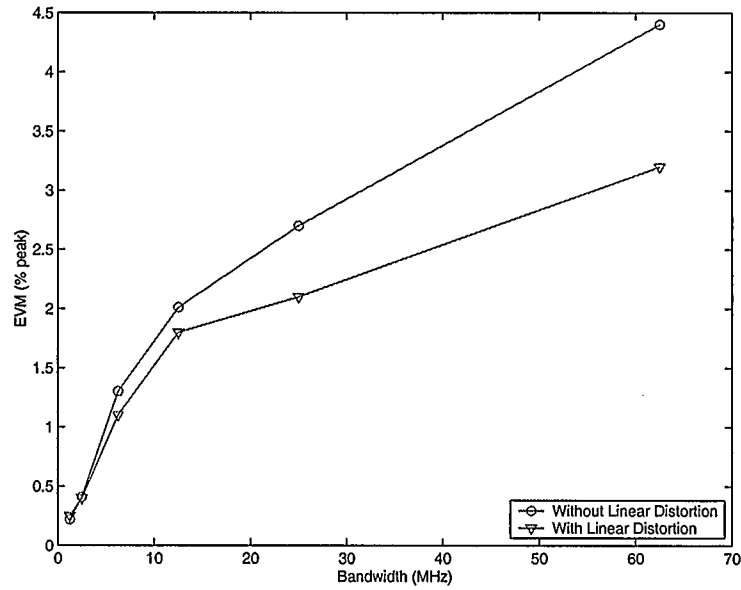


Figure 7.5: Model Performance

output of the model with a measured output, at the various signal bandwidths from the previous section. For low signal bandwidths, both models perform roughly the same, however, as the signal bandwidth increases the model with linear filtering is more accurate at predicting the output of the amplifier.

## 7.6 Linear and Nonlinear Predistortion Performance

With the Wiener model in place, the performance of a predistortion system based on a Hammerstein system can be evaluated. As with the test in Section 7.3, the test signal that was used is a QAM-16 waveform. Here the waveform is increased in bandwidth and compared to known values.

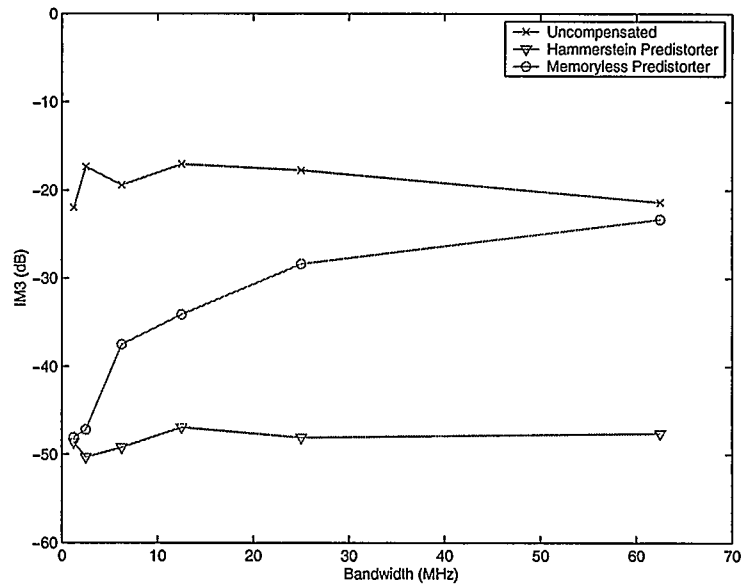


Figure 7.6: Predistortion with Memory Performance

As the results show in Figure 7.6 the performance of memoryless predistortion decreases as the bandwidth increases. This follows the expectation that memory and linear distortions will play a more important role as bandwidth increases. For low bandwidth signals the predistortion with memory performs almost exactly the same as the memoryless predistortion. However, as the signal bandwidth increases the performance of predistortion with memory remains almost constant at 30dB of improvement.

## Chapter 8

### Experimental Results

With the system architecture in place and the elements simulated, this chapter explores the implementation issues and experimental results from a test bench. The first section will describe the implementation issues, while the second section will discuss the test-bench setup and verify its operation. Finally the third section will introduce and discuss the experimental results obtained.

#### 8.1 Implementation Issues

Implementing a theoretical system into a physical form introduces many limitations and problems. Since the predistortion is done in a digital signal processor, the inputs and outputs are subject to a maximum voltage. Voltages above the maximum will be clipped. As with feedback linearization, the key to adjusting signal levels is in the magnitude of the feedback attenuation.

##### 8.1.1 Feedback Attenuation

The output of high power amplifiers can commonly exceed 10 watts, which is far too much for the feedback ADCs to handle directly. As such the feedback path must have enough attenuation to bring the signal down to a power that the ADCs can accommodate. However the overall gain of the system plays

a key role in the system. In (5.4) it is assumed that the system  $\Psi$  will have a linear gain of  $G$ . However since the output of an amplifier is orders of magnitude higher than the input, it is more convenient to scale down the output and assume  $G = 1$ . This is accomplished by tuning the feedback attenuation.

For the system to operate at its optimal point, the feedback attenuation must also be tuned properly. Figure 8.1 illustrates the effects of too much or too little feedback attenuation. If the attenuation is excessive, the feedback dynamic range will be less than the maximum, under-utilizing the ADC. When this system is inverted in the amplitude predistortion network, the input signal could exceed what has been learned, entering an unknown region of the neural network's operation. The output at this point is unpredictable, and most likely undesirable. If there is not enough feedback attenuation, the signal will become clipped. In turn, the inverted system will generate signals that will under-utilize the DACs. Only with careful tuning will the neural network generate the optimal output signal utilizing the full potential of the DACs, while not feeding the network an input which has not been learned.

The feedback attenuation will also have a direct impact on the overall gain of the system. Since the attenuation is set to make  $G = 1$ , the actual gain of the system will be the inverse of the attenuation.

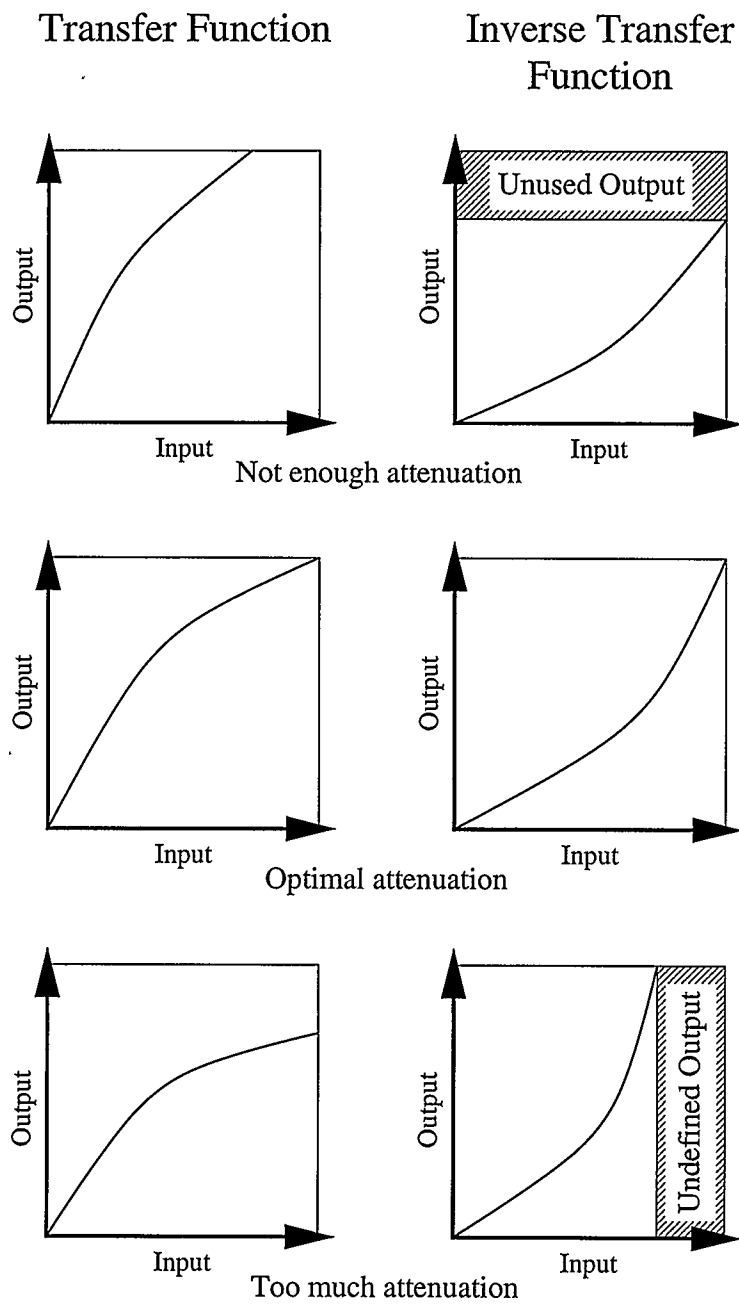


Figure 8.1: Effects of Feedback Attenuation



## 8.2 Experimental Setup

Simply simulating a predistortion system is not adequate to evaluate its performance. The simulation will only be as precise as the model used, and once the model is known the compensation to reverse the distortion is also known.

To facilitate the testing of a physical system, an experimental test-bench was constructed. The first section will describe the makeup of the test-bench, while the second section will describe the limitations of it. Finally the test-bench will be verified with by a couple of measurements to validate that it is working.

### 8.2.1 Experimental System

The experimental system block diagram is shown in Figure 8.2. The base-band predistortion is performed by Matlab on the host computer. The test signal is passed digitally to the arbitrary waveform generator via the GPIB bus connecting all the instruments and the host. Using the GPIB bus to control the vector signal generator, the signal is modulated to a desired frequency and power. The output of the vector signal generator is then passed through the device under test, demodulated, and sampled by the oscilloscope. From there the Matlab host uses the GPIB bus to capture the waveform from the oscilloscope for use in training.

Once the waveform has been captured off the oscilloscope, Matlab will train the linear and nonlinear sections of the predistortion, and upload the

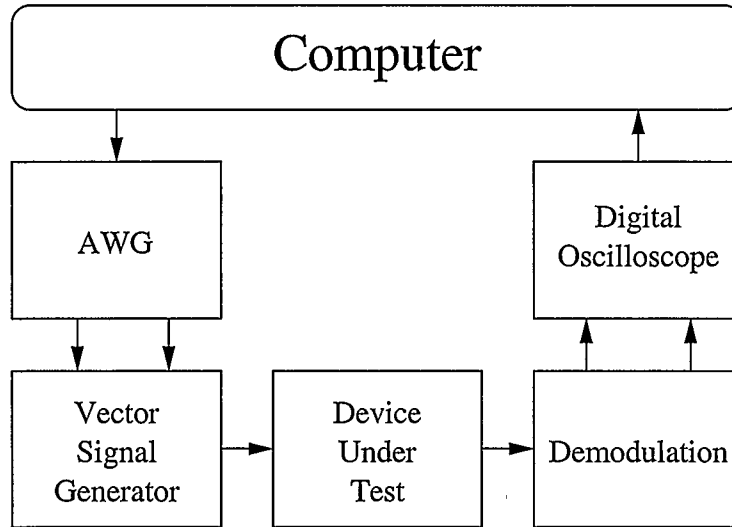


Figure 8.2: Experimental Setup

new coefficients to the predistortion blocks. The predistortion blocks will then calculate a new test signal, upload it to the arbitrary waveform generator and the cycle will begin again. Matlab scripting allows the system to be fully automated and can perform an arbitrary number of training cycles without human intervention.

### 8.2.2 Setup Limitations

While the test-bench provides a flexible platform to implement various setups, it does have some limitations. Two major limitations are the non-synchronized sampling clocks and path imbalances.

### **Non-synchronized Sampling Clocks**

Since an arbitrary waveform generator and an oscilloscope were used to generate and capture the signals, each has its own sampling clock. Unfortunately it was not possible to synchronize the clocks. A common 10MHz reference was used to ensure that both clocks are operating at the same frequency, however there is a random phase error generated each time the system is started.

Since the input signal needed to be delayed to line up with the output signal, the correction for the random phase was included with the delay used to lineup the input and output signals. Since the delay was done digitally, it was made adaptive so that the input correlated maximally with the output. Although the output waveform was distorted by the amplifier, the waveforms still exhibit a strong correlation when they are lined up.

### **Path Imbalance**

Once the signal is demodulated, the in-phase and quadrature arms are subject to individual imbalances. The imbalances come in the form of bias and gain. The specifications for the Lecroy LC684DXL oscilloscope state that the measurement accuracy is 2% of the total display. This error contributes to a measurement discrepancy in each oscilloscope channel resulting in uneven gain between channels.

The bias in each signal is most likely a result of the demodulation stage. The mixer's local oscillator to RF port isolation is specified as 30dB. With a local oscillator power of 4dBm, this means that up to -26dBm is leaking into

the RF port. Due to mismatches between other devices, some of this power will be reflected back into the mixer, resulting in a DC term. This DC term is directly added to the output signal. This is consistent with the 3 to 4 mV DC offsets typically measured in the system without any signal applied.

Once the Cartesian signal is converted into polar components, the amplitude and phase characteristics become distorted. If the signal is received as

$$S = (Ax + B) + (Cy + D)i$$

where  $A$  and  $C$  are the gains for the two channels, and  $B$  and  $D$  are the biases, the magnitude of the signal becomes

$$|S| = \sqrt{A^2x^2 + 2ABx + C^2y^2 + 2CDy + (B^2 + D^2)}$$

and the phase becomes

$$\angle S = \arctan\left(\frac{Cy + D}{Ax + B}\right)$$

This can complicate relations between signals. Measures need to be taken in order to ensure that  $A = C$  and  $B = D = 0$ . To accomplish this a test waveform was introduced into the system to calibrate these errors out. The calibration waveform was a square waveform, fed to both input ports. From this result the bias and the gain imbalances can be determined and eliminated.

### 8.2.3 Setup Verification

Several aspects of the test-bench had to be verified before experimental results could be obtained. Since the system is used to measure and correct nonlinearities, the system itself had to be linear. This was verified by connecting the vector signal generator directly to the demodulation stage, removing the device under test. With no device under test, a known waveform was fed through the system.

The spectrum measured at the spectrum analyzer is shown in Figure 8.3. The signal shows a slight indication of nonlinear effects around the signal. This is probably due to the mixers in the vector modulator, however the distortion is very small and will be ignored. To make sure that the mixers in the demodulation stage were not being over-driven into their nonlinear region, a waveform was injected into the setup and measured by the oscilloscope. Figure 8.4 shows the PSD of the signal measured by the oscilloscope. Here, no signs of nonlinearities are present.

To calibrate the local oscillator delay in the demodulation stage, a waveform was fed into the in-phase or quadrature channels. If the delay is calibrated properly, the signal will only be present in one channel or the other indicating that the output will be orthogonal.

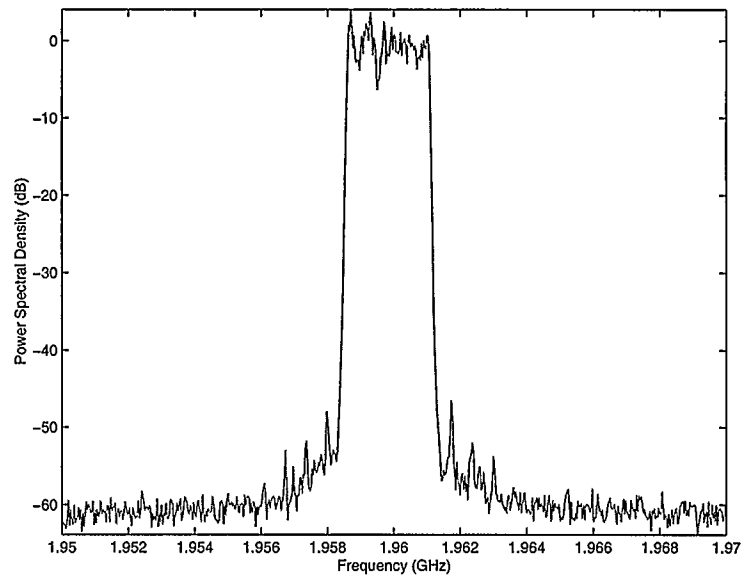


Figure 8.3: Modulated Spectrum Verification

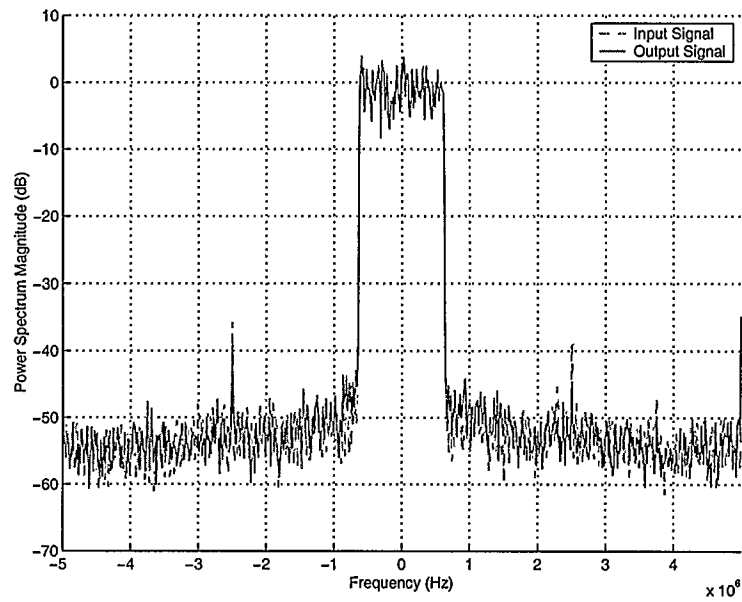


Figure 8.4: Demodulated Spectrum Verification

### 8.3 Results

The focus of the physical experiments was to test the predistortion under various conditions. First the effectiveness of the nonlinear predistortion with respect to intermodulation products is tested against various signal bandwidths. Then, with both the linear and nonlinear blocks in place, the overall performance is again gauged to see if an improvement in the inter-symbol interference can be made.

#### 8.3.1 Memoryless Predistortion Performance

Figure 8.5 shows the results of using only the nonlinear predistortion block while increasing the bandwidth. The graph measures the average power of 3rd order intermodulation products with respect to the average power of the signal. The figure shows that for lower signal bandwidths, the network can reduce the intermodulation products by over 20dB. However the cancellation steadily decreases as the bandwidth increases.

Figure 8.6 shows the overall decrease in 3rd order intermodulation products with respect to bandwidth. For 2MHz of bandwidth, the performance drops to less than 10dB of correction, and by 20MHz only 5dB of correction.

After approximately 50MHz of bandwidth, the intermodulation products begin to decrease, even without any compensation. This can be directly attributed to the frequency response of the amplifier. The 3dB bandwidth of the amplifier is 50.1MHz. As the bandwidth approaches this point the signal

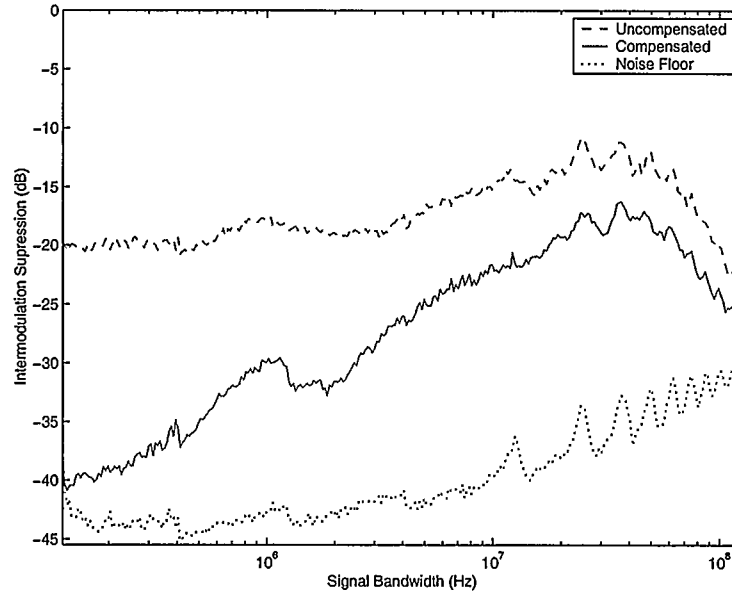


Figure 8.5: Nonlinear Cancellation Vs. Bandwidth

is being drastically attenuated, and the resultant intermodulation products will also decrease. The decrease in intermodulation products is also accompanied with an increase in linear distortions. The linear distortions will cause severe inter-symbol interference leading to elevated bit error rates, or as the bandwidth increases, total loss of the data contained in the signal.

### 8.3.2 Predistortion with Memory

To test the predistortion algorithm with memory the same test was repeated for a number of different bandwidths. The results are shown in Figure 8.7. Again the graph measures the average power of 3rd order intermodulation products with respect to the average power of the signal.



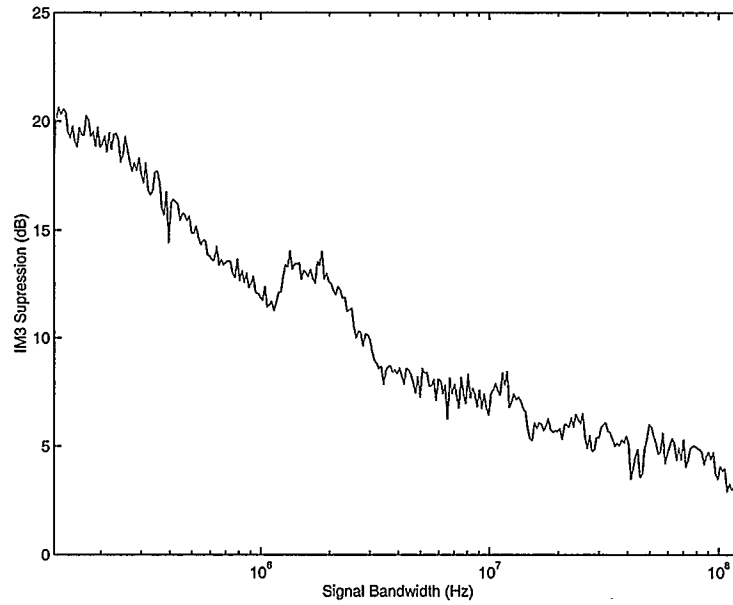


Figure 8.6: IM3 Suppression Vs. Bandwidth

Performing predistortion with memory compensation only improves the performance 1-2dB over memoryless predistortion. There are several possible reasons for the poor performance when compared to the simulation, including oversimplified models and distortion from the test-bench.

The most obvious reason for the performance discrepancy from simulation is the oversimplification of the amplifier model. By modeling the system as a Wiener system the characteristics of the amplifier might not be adequately represented to make a prediction of what the distortion will be. Several papers have proposed using a Wiener model of the amplifier and a Hammerstein system for the predistortion, including a system with a Volterra nonlinearity

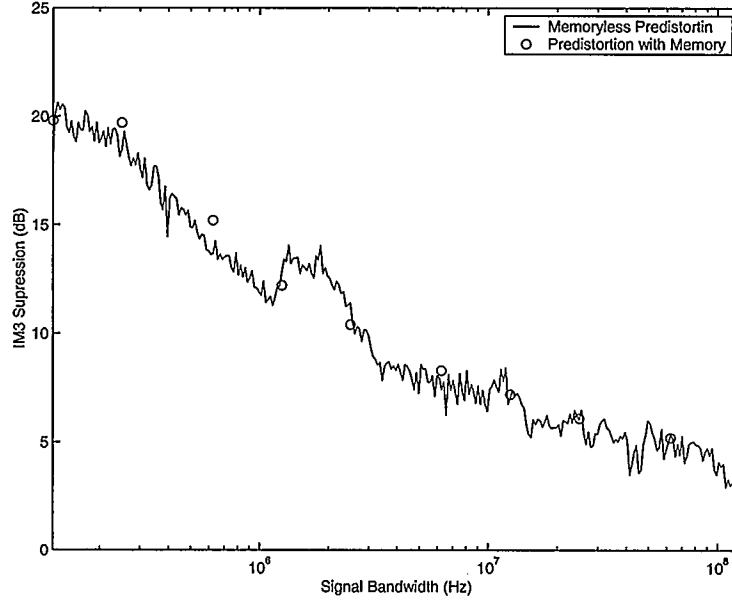


Figure 8.7: Predistortion With Memory

[34], and another system with a polynomial memoryless nonlinearity [35]. To the author's knowledge, none has ever been tested on the bench.

The test-bench could also dramatically affect results. The goal of predistortion is to linearize the transmitted waveform. However the feedback loop observes a demodulated version of the transmitted waveform. Several papers have been published on the errors introduced by quadrature modulators and demodulators, and the performance penalty of those errors [36, 37]. If the demodulation stage introduces any artifacts into the feedback signal, the predistortion will attempt to correct for them. However, these artifacts are not actually present in the transmitted waveform, and the extra correction will in

turn introduce distortion into the transmitted waveform.

## Chapter 9

### Conclusions and Future Work

This chapter concludes the thesis with some closing remarks and some proposed future areas of exploration. An overall summary of the thesis and advancements made, is provided. Then the thesis is finished off by discussing some unanswered questions that this work has raised.

#### 9.1 Thesis Summary

In Chapter 1 wireless technology was identified as a major growth area in our society. With more people than ever being connected to each other by the Internet and other wireless links, current systems are being pushed to their limits. The base station amplifier was identified as the main source of distortion in the RF front end.

The types of distortion, and their symptoms, were the focus of Chapter 2. Various categories of distortion were discussed along with their effects on a signal. The last part of this chapter focused on the metrics used to quantify the distortion.

Since this thesis used a neural network to model nonlinear characteristics of the amplifier, a brief introduction to neural networks was provided in Chapter 3. The chapter began by introducing what neural networks are and how they

are structured. The operation of the network and how they are trained close out the chapter.

Chapter 4 introduces the concept of linearization, and some of the techniques used. Feed-back, feed-forward, LINC, and EER is introduced, followed by a discussion of the advantages and disadvantages of each.

The predistortion technique was left to Chapter 5 in order to allow for a more in depth discussion. A background into predistortion is provided including previous works and a mathematical derivation. The design decisions of predistortion are then discussed with the advantages and disadvantages of each. Finally the chapter finishes with the limitations inherent in predistortion and the overall advantages and disadvantages of this technique.

With all the building blocks of the previous sections in place, the proposed system architecture is introduced in Chapter 6. The chapter begins with a brief explanation of a Wiener model and its inverse, the Hammerstein system. From there, the overall system architecture is discussed along with details on each individual block in the system.

In Chapter 7 the system architecture was implemented in Matlab to evaluate its performance in simulation. The chapter began by testing a the accuracy of a memoryless nonlinear model and how well it performed in simulated predistortion. Then the model is extended to a Wiener model by adding a linear distortion with memory. This new model is again to demonstrate the improvement in modeling accuracy and in simulated predistortion. For each test the

results are discussed.

Finally Chapter 8 put the system to test on a physical amplifier. This chapter started by introducing the actual test-bench, discussing its limitations and verifying its measurements. With the test-bench fully operational memoryless predistortion was tested with signals of various bandwidth. As the bandwidth increased the performance of memoryless predistortion dropped dramatically, highlighting the need for predistortion with memory. The last test implemented the system for predistortion with memory as introduced in Chapter 6. The improvement over the memoryless nonlinear system was approximately 1 - 2 dB. The results are then compared to the simulation results of Chapter 7, and the differences discussed.

## 9.2 Achievements

There were several goals laid out at the beginning of this thesis, including the applicability of neural networks to amplifier modeling, testing the effects of broadband signals on memoryless predistortion and implementing a predistorter with memory and testing it on a physical amplifier.

Neural networks have proven to be a very effective tool to model nonlinear systems. With only the input and output signals, they are able to model a system with remarkable accuracy. The simulation results in Chapter 7 show that they are capable of modeling the memoryless nonlinearity of the amplifier to a high degree of accuracy.

Using the memoryless nonlinear model provided by the neural network, the effects of increasing signal bandwidth were then tested. The results show that predistortion is very sensitive to increasing bandwidth and highlights the need for predistortion methods to account for memory effects.

Extending the memoryless nonlinear model into a Hammerstein system was then tested on a physical amplifier. To the authors knowledge this is the first time that a Hammerstein predistorter has been tested beyond simulation. While the results are not as impressive as the simulation, they do show some improvement, and highlight the shortcomings of simulation. If every aspect of the model is known, a theoretical system can be devised that will undo each aspect perfectly. This reduces the role of simulation to system verification, and makes it invalid for testing system performance.

### 9.3 Future Work

Through the course of this thesis, several new questions have been brought up that warrant further discussion. The non-ideal effects of the modulation and demodulation stages, the validity of the Wiener model and the effectiveness of this technique on other waveforms are all possible subjects of future work.

As stated in Section 8.3.2, any non-ideal effects of the modulation stages and demodulation stages can decrease the effectiveness of predistortion. The extent of these effects on the proposed methods is not known. Since the feedback signal is assumed to be a linearly scaled representation of the trans-

mitted signal, the possibility exists for the predistorter to correct distortions that don't exist in the transmitted signal. The feedback path needs to be characterized in order to calibrate out any distortion introduced. This thesis used a calibration signal before each trial to determine gain imbalances in the in-phase and quadrature, and any biases introduced. These imbalances and biases were then removed from the feedback path. This works under the assumption that the arbitrary waveform generator and the vector modulator are nearly ideal. As both of these pieces of equipment are high quality test equipment, for the purposes of this thesis it is probably a valid assumption. However this assumption should be investigated.

Also in Section 8.3.2, the validity of the Wiener model assumption was questioned. If the Wiener model is insufficient to model the amplifier, then a Hammerstein predistorter will not fully compensate for the amplifier distortion. For amplifier modeling to be of benefit to a predistortion system, the model must be able to be inverted. More work into modeling amplifiers with memory needs to be performed, with the focus on models that can be inverted.

Finally this thesis assumed that a QAM-16 signal would produce results representative of any signal. This assumption is true for linear systems, but may not be true for nonlinear systems. More testing with higher peak to average ratio signals, such as multiuser CDMA systems, needs to be performed.

Even with these outstanding questions, the technique in this thesis shows promise. Further enhancements in the modulation and demodulation stages,



along with better amplifier modeling will improve the performance beyond that of memoryless predistortion.

## Bibliography

- [1] P. H. Opere, "Linearization of an RF power amplifier with memory," Master's thesis, University of Calgary, June 2001.
- [2] A. Saleh, "Frequency independent and frequency dependent nonlinear models of twt amplifiers," *IEEE Transactions on Communications*, vol. 29, pp. 1715 – 1720, November 1981.
- [3] T. Ha, *Solid State Microwave Amplifier Design*. John Wiley & Sons, Inc., 1981.
- [4] D. Rumelhart and J. McClelland, *Parallel Distributed Processing*, vol. 1. M.I.T. Press, 1986.
- [5] D. Rumelhart and J. McClelland, *Parallel Distributed Processing*, vol. 2. M.I.T. Press, 1986.
- [6] F. Wang and Q. Zhang, "Knowledge based neural models for microwave design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 45, December 1997.
- [7] S. Bila et al., "Accurate wavelet neural network based models for electromagnetic optimization of microwave circuits," *International Journal of RF and Microwave CAE, Special Issue on Applications of ANN to RF and Microwave Design*, vol. 9, pp. 297 – 306, 1999.

- [8] Y. Harkouss et al., "Modeling microwave devices and circuits for telecommunications system design," *IEEE International Conference on Neural Networks*, pp. 128 – 133, May 1998.
- [9] J. Aweya, Q. Zhang, and D. Montuno, "A direct adaptive neural controller for flow control in computer networks," *IEEE International Conference on Neural Networks*, pp. 140 – 145, May 1998.
- [10] N. Benvenuto and F. Piazza, "On the complex backpropagation algorithm," *IEEE Transactions on Signal Processing*, vol. 40, pp. 967 – 969, April 1992.
- [11] H. Black, "Wave translation system." U.S. Patent 1,686,792, December 21 1937.
- [12] J. McRory, "An RF feedback amplifier for low intermodulation distortion performance," Master's thesis, University of Calgary, April 1993.
- [13] S.C.Cripps, *RF Power Amplifiers for Wireless Communications*. Artech House, 1999.
- [14] A. Javed, P.Goud, and B. Syrett, "Analysis of a microwave feedforward amplifier using Volterra series representation," *IEEE Transactions on Communications*, pp. 355–360, March 1977.
- [15] S. Grant and P. Goud, "A DSP controlled adaptive feedforward amplifier linearizer," *IEEE Conference on Universal Personal Communication*,

pp. 788–791, September 1999.

- [16] A. Smith and J. Cavers, “A wideband architecture for adaptive feedforward linearization,” *Proceedings of the 1998 IEEE Vehicular Conference*, pp. 2488–2492, 1998.
- [17] D. Cox, “Linear amplification with nonlinear components,” *IEEE Transactions on Communications*, vol. 22, pp. 1942–1945, December 1974.
- [18] C. Conradi, “LINC transmitter linearization techniques,” Master’s thesis, University of Calgary, January 2000.
- [19] F.H. Raab et al, “High efficiency L-band kahn technique transmitter,” *Proceedings of the MTT-S*, 1998.
- [20] Davis et al., “Adaptive predistortion technique for linearizing a power amplifier for digital data systems.” U.S. Patent 4,291,277, September 1981.
- [21] A. Saleh and J. Salz, “Adaptive linearization of power amplifiers in digital radio systems,” *Bell System Technical Journal*, vol. 62, pp. 1019 – 1033, April 1983.
- [22] J. Cavers, “Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements,” *IEEE Transactions on Vehicular Technology*, pp. 374 – 382, November 1990.

- [23] S. Stapleton and J. Cavers, "A new technique for adaptation of linearizing predistorters," *IEEE Proceedings on Communications*, May 1991.
- [24] R. Vaughan, N. Scott, and D. White, "The theory of bandpass sampling," *IEEE Transactions on Signal Processing*, vol. 39, pp. 1973 – 1984, September 1991.
- [25] T. Nojima and T. Konno, "Cuber predistortion linearizer for relay equipment in 800MHz band land mobile telephone system," *IEEE Transactions on Vehicular Technology*, vol. 34, pp. 169 – 177, November 1985.
- [26] N. Imai, T. Nojima, and T. Murase, "Novel linearizer using balanced circulators and its application to multilevel digital radio systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 37, pp. 1237 – 1243, August 1989.
- [27] C. G. Rey, "Predistorter linearizes CDMA power amplifiers," *Microwaves & RF*, pp. 114 – 123, October 1998.
- [28] S. Stapleton and F. Costescu, "An adaptive predistorter for a power amplifier based on adjacent channel emissions," *IEEE Transactions on Communications*, vol. 41, pp. 49 – 56, February 1992.
- [29] M. Ghaderi, S. Kumar, and E. Dodds, "Adaptive predistortion linearizer using polynomial functions," *IEEE Proceedings on Communications*, vol. 141, pp. 41 – 55, April 1994.

- [30] N. Benvenuto, F. Piazza, and A. Uncini, "A neural network approach to data predistortion with memory in digital radio systems," *Proceedings of IEEE International Conference on Communications*, pp. 232 – 236, 1993.
- [31] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electronic Letters*, vol. 37, pp. 1417 – 1418, November 2001.
- [32] J. Bendat, *Nonlinear Systems Techniques and Applications*. John Wiley & Sons, Inc., 1998.
- [33] C. Clark, G. Chrisikos, M. Muha, A. Moulthrop, and C. Silva, "Time-domain envelope measurement technique with application to wideband power amplifier modeling," *IEEE Transactions on Microwave Theory*, vol. 46, pp. 2531 – 2540, December 1998.
- [34] C. Eun and E. Powers, "A predistorter design for a memoryless nonlinearity preceded by a dynamic linear system," *Proceedings of GLOBECOM*, pp. 152 – 156, November 1995.
- [35] L. Ding, R. Raich, and G. Zhou, "A hammerstein predistorter linearization design based on the indirect learning architecture," *IEEE Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp. 2689 – 2692, 2002.
- [36] J. Cavers, "The effects of quadrature modulator and demodulator errors

on adaptive digital predistorters,” *Proceedings of IEEE conference on Vehicular Technology*, vol. 2, May 1996.

- [37] J. Cavers, “A fast method for adaptation of quadrature modulators and demodulators in amplifier linearization circuits,” *Proceedings of IEEE conference on Vehicular Technology*, vol. 2, pp. 1307 – 1311, May 1996.