AN ABSTRACT OF THE THESIS OF

Xiangqing Xu for the degree of Master of Science in Electrical and Computer Engineering presented on May 30, 1995. Title: A Pipelined Baseband Digital Predistortion Architecture for RF Amplifier Linearization

Redacted for Privacy

Abstract approved: _____

James H. Herzog

This thesis describes the design of a baseband digital predistortion architecture for radio frequency (RF) amplifier linearization. Existing amplifier linearization techniques are first examined. The nonlinearity of an amplifier can be characterized as amplitude-to-amplitude distortion and amplitude-to-phase distortion. The distortion can be corrected by providing a predistorted driver signal (both in amplitude and phase) to the amplifier. This counteracts the nonlinearity of the amplifier so that the overall system is more linear. A pipelined predistortion architecture is designed to allow fast processing speed, and is capable of providing linearization for a bandwidth of 25 Mhz. The simulation results show significant improvement in amplifier's performance using predistortion technique. The system bandwidth versus cost is examined by using commercially available components.

A Pipelined Baseband Digital Predistortion Architecture for RF Amplifier Linearization

by

Xiangqing Xu

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Completed May 30, 1995
Commencement June 1996

Master of Science thesis of Xiangqing Xu presented on May 30, 1995

APPROVED:

Redacted for Privacy

Major Professor, representing Electrical and Computer Engineering

Redacted for Privacy

Chair of Department of Electrical and Computer Engineering

Redacted for Privacy

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for Privacy

Xiangqing Xu

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to professor James Herzog for his understanding, help, patience, advice and encouragement. I am grateful to professor Bella Bose, professor V.K. Trapathi and professor R.J. Schultz for agreeing to be on my graduate committee and taking time out of their busy schedules to be present at my defense.

Special thanks to Mr. Robert S. Larkin, chief engineer (retired) of Celwave for his help, advice and concern on this project. I also thank Mrs. Martina Testa, engineering manager at Celwave for her understanding and support throughout my research. Thanks also to Qiang Wu, Hans-Albert Stumm, David Wills, Phil Mages and all of my colleagues for their friendship and support.

I thank Celwave for their financial support during my study.

I thank Rita Walls, graduate secretary of the Electrical and Computer Engineering department, for her help and support.

Finally, I want to give my deepest gratitude to my husband, Liping and my daughter, Yin. Without their love and encouragement, this would surely not have been possible.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF TABLES

A Pipelined Baseband Digital Predistortion Architecture for RF Amplifier Linearization

Chapter 1. Introduction

Cellular telephone communication is becoming increasingly important in the world. Power amplifiers are essential elements in the signal processing path of Cellular system. Nonlinearity of an amplifier creates an interference problem by generating signals at undesired frequencies. The overall goal of this work is to develop a technique to linearize power amplifiers in cellular communication systems.

1.1 Radio frequency (RF) amplifiers in Cellular communication systems

Signal amplification is important in Cellular communication systems. Depending on the application, radio frequency (RF) amplifiers can be classified as low noise and low power receiver amplifiers, or high power transmitter amplifiers.

The receiver amplifier emphasizes lower noise figures and is able to handle weak signals. High dynamic range and high sensitivity allow the amplifier to deal with different amplitude signals that are received from the antenna. Receiver amplifiers are normally small signal class A amplifiers (definition of "class" is given in appendix A).

Transmitter amplifiers are normally power amplifiers. The noise figure is of minor interest since the transmitted signal is usually a known signal and the level can be controlled. For most applications, the efficiency is also a concern,

especially in portable communication devices. For the purpose of obtaining high power efficiency, power amplifiers are normally class-AB, class-C or Class-A operating near the saturation condition. These types of amplifiers generally experience a degree of nonlinearity in their operating range.

In a multi-channel amplifier, when multiple RF signals are amplified simultaneously, intermodulation distortion is produced in the output due to the nonlinearity of the amplifier. These distortions can interfere with the signal being amplified.

In a single-channel amplifier, when non-constant envelope signals pass through the amplifier, the nonlinear distortion causes emissions and adjacent channel interference. Because of the limited RF spectrum, restrictions on emissions and levels of acceptable adjacent channel interference are stringent. RF linear amplification is required in mobile radio systems.

1.2 Frequency spectrum of Cellular telephone system

In the US, the RF spectrums are set by FCC regulation. In cellular telephone communications, an RF frequency band is split into smaller fixed frequency channels. Multiple signals, each with a different carrier frequency, share the same transmission medium. This is called frequency multiplexing.

In the current US Cellular band, there are two 25 MHz frequency spectrums allocated for cellular phones, the frequency 824 MHz to 849 MHz is for mobile transmit and 869 MHz to 894 MHz is for base station transmit. Each of these 25 MHz band is split into 832 channels which are 30 kHz wide (21).

Amplifiers used in the Cellular system can cover a single channel or the whole band. Depending on the architecture, sometimes an amplifier shifts the

25 MHz spectrum from the high frequency band to the 0-25MHz frequency band. The signal is then processed. This lowest frequency band is often referred to as a baseband. Modulation and demodulation are used to shift signal to different regions of the spectrum.

1.3 RF amplifier nonlinear characteristic

In communication systems, an amplifier can be viewed as a two-port network. The output y(t) can be expressed as a function of the input x(t). That is, $y(t)= H[x(t)]$.

If the output y(t) is a scaled version of the input x(t) and superposition applies, then the amplifier is a linear amplifier and the amplification is distortionless amplification. Figure 1.1 and Figure 1.2 show the amplitude and phase response of a linear amplifier. In the real world, there is no amplifier that has this type of ideal transfer characteristic. A typical RF amplifier's amplitude and phase characteristic is shown in Figure 1.3 and Figure 1.4 respectively. The transfer function is nonlinear.

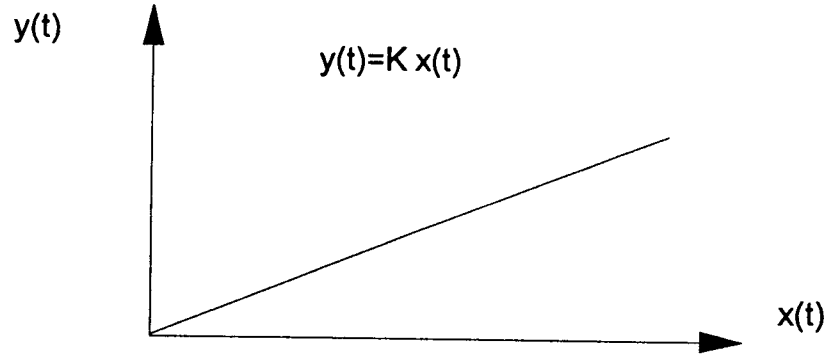

Figure 1.1 Transfer characteristic of a linear amplifier (amplitude response)

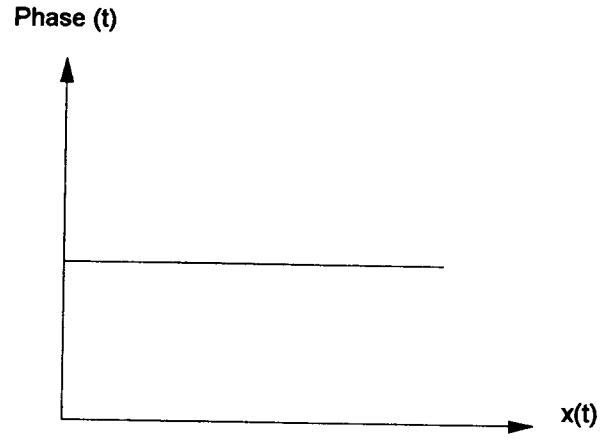

Figure 1.2 Transfer characteristic of a linear amplifier (phase response)

Signal distortion resulting from passing a signal through an amplifier with a non-linear transfer characteristic is called "nonlinear distortion". Nonlinear distortion puts limits on the dynamic range of the amplifier. The dynamic range of an amplifier can be defined by its noise figure or its sensitivity as the lower limit and by its acceptable level of signal distortion as its upper limit. Normally, nonlinear distortion occurs when the signal level is high and the amplifier is driven into the saturation region. This type of distortion can be characterized as amplitude to amplitude distortion (AM-AM) and amplitude to phase distortion (AM-PM).

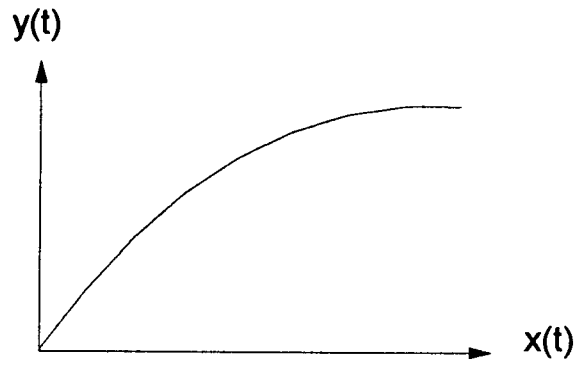

Figure 1.3 Transfer characteristic of nonlinear amplifier (amplitude response)

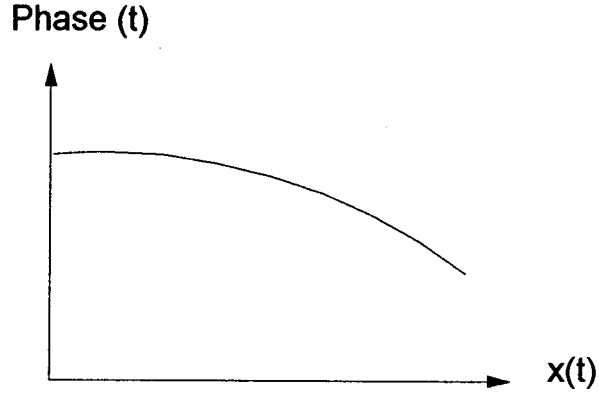

Figure 1.4 Transfer characteristic of nonlinear amplifier (phase response)

1.4 Amplifier linearization techniques

The simplest way of achieving linear amplification is to operate a class A amplifier far below saturation. The disadvantage is low power efficiency (Appendix A). In general, if power amplifiers are operated in the linear region of their transfer characteristics, they are power inefficient. If they are power efficient, they are normally not linear. The challenge is to maintain linearity and power efficiency at the same time. This is especially true if the amplifiers are being used in portable communication devices. One of the methods to solve

this problem is by using linearization techniques to make nonlinear power amplifiers more linear.

Many different techniques have been studied or proposed during recent years. Some techniques directly manipulate the high frequency RF carrier signals, such as feed forward correction and RF predistortion. Other techniques work with the baseband signal, then shifting the signal to high frequency RF carrier signal. Cartesian feedback and digital predistortion are examples of such techniques.

1.4.1 Feed forward

Feed forward amplifiers (11,14,15) apply the nonlinearity cancellation directly to the RF carrier signal. A simplified block diagram of feed forward amplifier is shown in Figure 1.5.

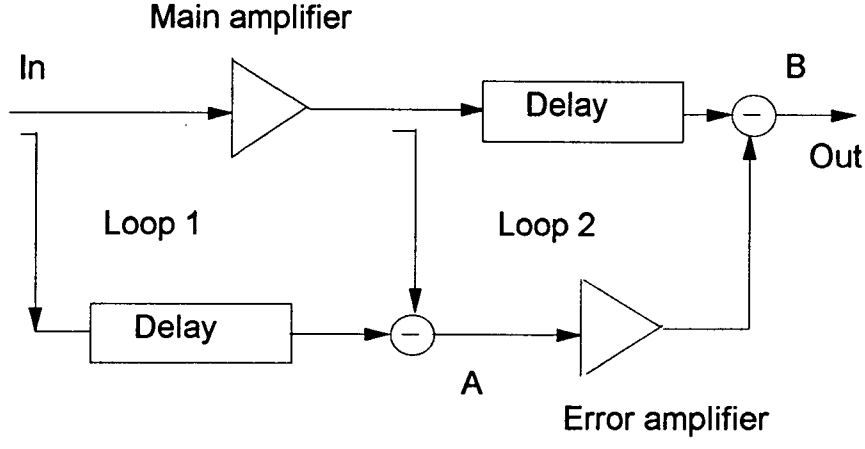

Figure 1.5 Block diagram of a feed forward amplifier

A feed forward amplifier consists of two loops. In loop 1, the input signal feeds into two paths. One goes through the main amplifier and the other to a linear circuit which delays the signal the same amount as the delay through the main amplifier. The original signal is distorted at the output of the main amplifier because of the nonlinearity of the amplifier. Part of the distorted signal is then subtracted from the delayed original signal at point A. These two signals are controlled so that they are of equal amplitude and 180° out of phase. Ideally the resulting error signal only contains the distortion information from the main amplifier. This signal is then amplified to the required level through the error amplifier to cancel out the distortion in the main path and form a non distorted amplified output signal at point B. Again, as in loop 1, two signals at point B must be equal in amplitude and 180° out of phase in order to cancel out the distortion.

The feed forward process is an open loop process. The bandwidth can be from ten's of MHz to hundred's of MHz (15). Therefore it is suitable for application on a wide-band system. In order to efficiently cancel out the distortion, the amplitude and phase of the signals must closely match in the amplifier and the delay paths. also the error amplifier must not generate any distortion of its own. The balances need to be within a few tenths of a dB and a few degrees to meet required linearity while optimizing maximum available output power.

1.4.2 Cartesian feedback

The Cartesian feedback method applies linearization techniques to the baseband signal. The baseband signal is transmitted by modulating a carrier

signal. Modulation is the process of encoding the low frequency baseband signal onto a carrier signal. The frequency of the carrier is compatible with the transmission medium being used.

The Cartesian feedback technique (3,12) was primarily designed to be applied in transmitter amplifiers where the baseband signal is available. It adjusts the signal according to the feedback from the output of the amplifier in the baseband, then shifts up to the RF carrier frequency for transmission.

One common method to represent a digital baseband signal in a communication systems is to use a vector in a polar plot (21,23), as shown in Figure 1.6. The length of the vector, ρ, represents the magnitude of the signal while the angle, θ, represents the phase of the signal. This vector can be further decomposed into two rectangular components. The two signals are referred to as the in-phase I and quadrature-phase Q and:

$\theta = \tan^{-1} Q/I$

$\rho = \sqrt{(I^2 + Q^2)}$

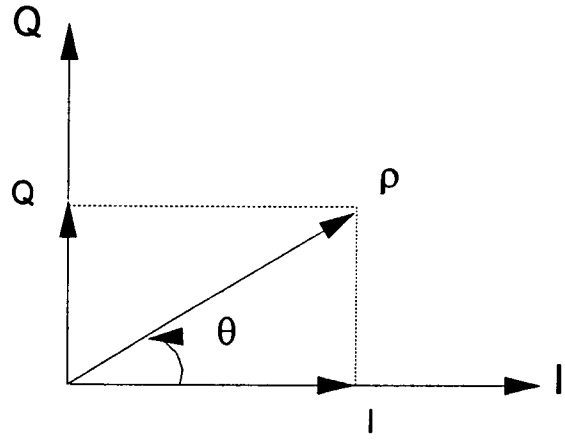

Figure 1.6 I/Q plane

With proper lengths selected for I and Q, any point can be generated in the plane.

I and Q are modulated onto an RF carrier frequency by an I/Q modulator before final amplification and transmission. An I/Q modulator generates a sine and a cosine channel for I and Q data. Figure 1.7 shows the I/Q modulator and demodulator. The 90° phase shifter is used to generate the sine and cosine channels. Assume the I and Q signals are I(t) and Q(t) respectively, and the carrier signal is $\cos\omega_c t$, then the modulated RF signal x(t) can be described as:

$$x(t) = I(t)\cos\omega_c t + Q(t)\sin\omega_c t$$

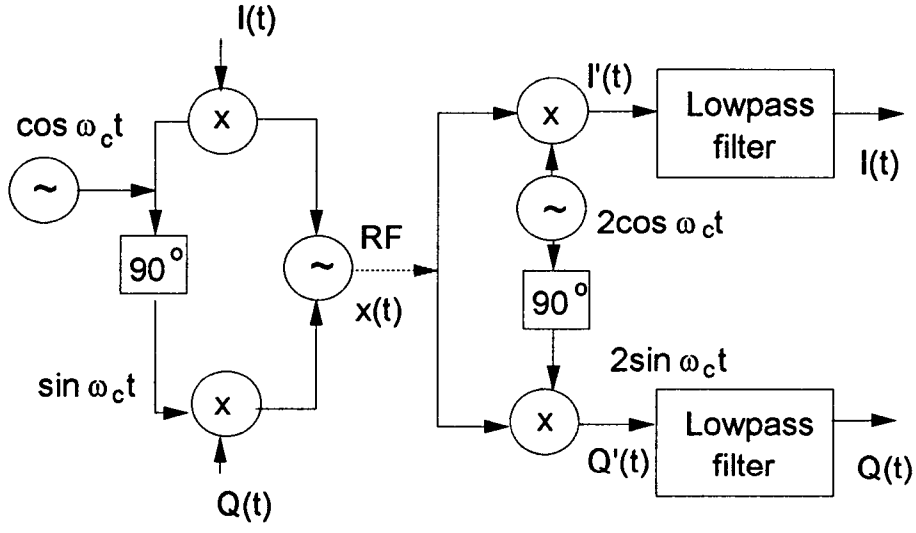

Figure 1.7 I/Q modulator and demodulator

The process that removes the carrier and separates the signal into I and Q components is called I/Q demodulation. The I channel signal is obtained by multiplying x(t) with $2\cos\omega_c t$. It can be shown that:

$$I'(t) = 2(I(t)\cos\omega_c t + Q(t)\sin\omega_c t)\cos\omega_c t$$

$$= I(t) + I(t)\cos 2\omega_c t + Q(t)\sin 2\omega_c t$$

The last two terms are suppressed by the low pass filter, yielding the desired output I(t). Similarly, by multiplying x(t) with $2\sin\omega_c t$, yielding:

$$Q'(t) = 2(I(t)\cos\omega_c t + Q(t)\sin\omega_c t)\sin\omega_c t$$

$$= Q(t) + I(t)\sin 2\omega_c t - Q(t)\cos 2\omega_c t$$

Q(t) is obtained from the resulting signal through the low pass filter.

Figure 1.8 shows the block diagram of a Cartesian feedback amplifier.

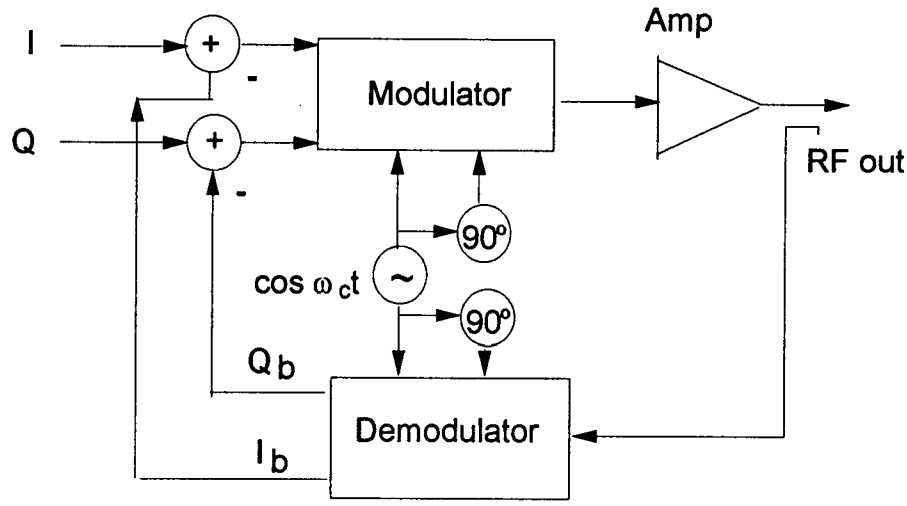

Figure 1.8 Block diagram of a Cartesian feedback amplifier

The baseband input signal is in quadrature components, I and Q, format. The feedback path takes a portion of the output signal and demodulates it into quadrature components, I_b and Q_b, to provide negative feedback. Then I_b and Q_b are subtracted from the original I and Q input drive signal to generate a loop error signal. The Cartesian feedback is a closed-loop configuration. The linearizing bandwidth (the bandwidth over which the system is capable of performing linearization) of the Cartesian feedback system is on the order of ten's of kHz. Therefore it is suitable for a narrow-band application. The closed-

loop feature can easily compensate for changes in an amplifier's nonlinearities caused by component aging, temperature drifts or other changes.

1.4.3 Predistortion

The idea of predistortion is to predistort the input signal prior to amplification. This counteracts the nonlinear characteristics of the amplifier and makes the overall system appear linear. A detailed discussion of this technique is given in Chapter Two.

1.5 Thesis objectives

Because linear amplification is essential and desirable for cellular radio systems, linear amplification techniques have been an important study topic in recent years. Researchers are studying various types of techniques. Each method has its advantages and disadvantages. Some may be better suited for one application than another, but there is no single best solution to the problem. Research is going in many directions. Digital predistortion is one of them.

The purpose of this work is to study various ways of realizing a digital predistortion technique in RF amplifiers. A practical architecture will be presented to implement a digital predistorter and explore the trade-offs between processing speed, circuit complexity, cost and system bandwidth.

Chapter 2. Literature Review on Predistortion

2.1 Predistortion concept

In predistortion amplifier systems, a module called a predistorter is placed between the input signal and the power amplifier. The predistorter is a nonlinear module providing an appropriately distorted signal to the amplifier to compensate for its nonlinearity (Figure 2.1). Like the feed forward system, the correction to the nonlinearity in the predistortion amplifier is an open-loop compensation. There are various ways to implement the predistorter. Most methods can be characterized by two major categories, namely, analog RF predistorters and baseband digital predistorters.

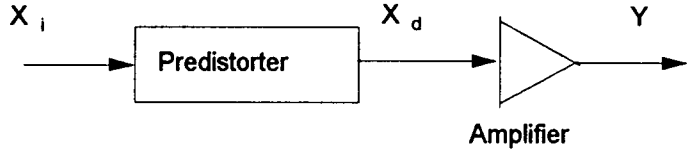

Figure 2.1 Block diagram of a predistortion amplifier system

2.2 Analog RF predistorter

The analog RF predistorter places a network of active and passive components which are designed to compensate for the amplifier characteristics(9) between the signal and the input of the amplifier. Diodes are often used in the network to provide the nonlinear compensation. The correction can be fixed compensation, meaning that once the network is implemented, the correction will not change, or adaptive in which the correction

can be continually adjusted with the change of the amplifier characteristics. Because of the open loop feature, RF predistortion is suitable for wide-band application. Due to the difficulty of designing a circuit that can perfectly match the amplifier, the cancellation is not complete.

2.3 Fixed baseband digital predistortion

Unlike RF predistortion, the signal predistortion in this technique takes place in the baseband. In fixed baseband digital predistortion, the signal is first predistorted, then modulated onto a carrier RF frequency and fed into the amplifier. Figure 2.2 gives a block diagram.

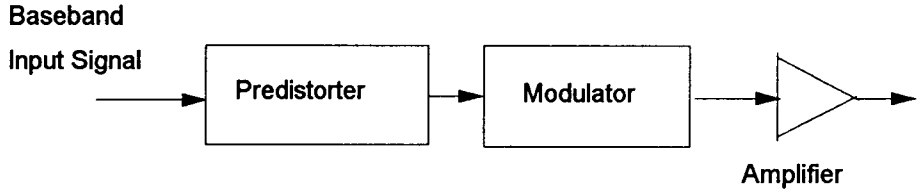

Figure 2.2 Block diagram of fixed baseband predistortion system

The advantage of predistorting the signal in the baseband is that digital techniques can be applied. Fixed digital predistortion is simple but will not compensate for changes in the amplifier due to aging, temperature or other reasons. For different amplifiers, the signal may need to be predistorted in a different way. This type of predistortion is not used much in practice.

2.4 Adaptive baseband predistortion

This implementation predistorts the baseband signal as in fixed predistortion but the predistortion circuit is adapted by an error signal from the output of the amplifier. The feedback is used only for compensating the drifts of the amplifier as a result of changing temperatures, aging and so on. The feed back is not used for real time modification of predistortion signals. Adaptive baseband predistortion is still an open-loop system in the sense of nonlinearity cancellation. The general block diagram of an adaptive predistortion system is shown in Figure 2.3

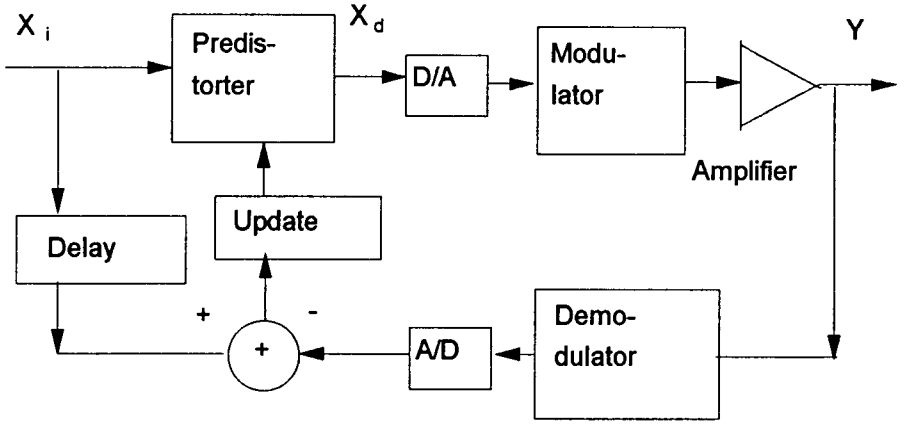

Figure 2.3 Block diagram of an adaptive predistortion system

The baseband digital modulated input signal, X_i, is predistorted to produce X_d, converted to analog form through a digital to analog (D/A) converter, then amplified through a power amplifier. The output signal is demodulated and fed back to compare with the appropriately delayed original signal. This generates an error signal to be used for adjusting the predistortion

circuit to adapt to changes in amplifier characteristics. The main problems in an adaptive predistortion system are generating the predistortion signal, monitoring the performance of the system and continuously adapting the predistorter as the characteristics of the amplifier change.

With the development of digital signal processing (DSP) and microcontroller techniques, digital adaptive predistortion has become very popular. Several adaptive predistortion systems that make use of microcontrollers, memory look-up tables(LUT) and DSP techniques have been reported in recent years (8,1,2,5,6). One of the most general and powerful methods was reported by Nagata (8) in 1989. In his implementation, the input signal, X_i, is represented in quadrature format. The signals, I and Q, are used as addresses of a two-dimensional memory LUT to generate a predistorted signal, X_d. The output, X_d, from the LUT is also in quadrature format, I_d and Q_d. The mapping of the two is from the I/Q plane to itself. This implementation is not restricted by the type of modulation used. On the other hand, since I and Q carry both amplitude and phase information, the amplifier's amplitude and phase characteristics can be compensated by X_d. The number of quantization bits plays an important role in the performance of the system. The more bits, the more precisely the signal is represented in the plane. This provides better cancellation of the amplifier nonlinearity. But the more bits, the larger the memory look up table will be. Nagata's study showed that 10 quantization bits were sufficient to achieve reasonable nonlinearity cancellation. The size of the two-dimensional look up table is 20M bits (2^20x20 bit) (Figure 2.4). The table is updated when the nonlinearity is changed. Since the time required to update the LUT is proportional to the size of the memory, each iteration in the adaptive process requires a significant length of time.

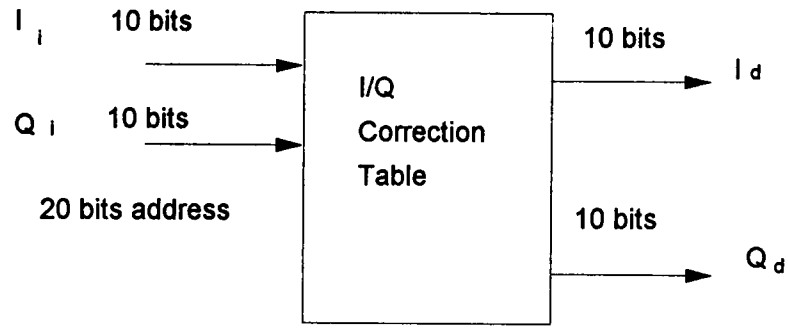

Figure 2.4 Nagata's two dimensional look up table

In 1990, Cavers (1) used a one-dimensional table to reduce the memory size. The idea is based on that an amplifier's AM-AM and AM-PM distortion are input signal level dependent (19,1,2). This implies that the magnitude of the input signal can be used to generate correction signals for both amplitude and phase. A complex gain which contains both amplitude and phase information is first defined, the coefficients of this complex gain are stored into a memory look-up table, then the squared magnitude of the input signal is used as a table index to obtain the coefficients. These coefficients are used to calculate predistorted signal X_d (Figure 2.5). Since calculations are required to obtain X_d, the table size is reduced in the expense of computation load. Both the squared magnitudes of X_i and the multiplication of X_i by the table values have to be performed between data samples.

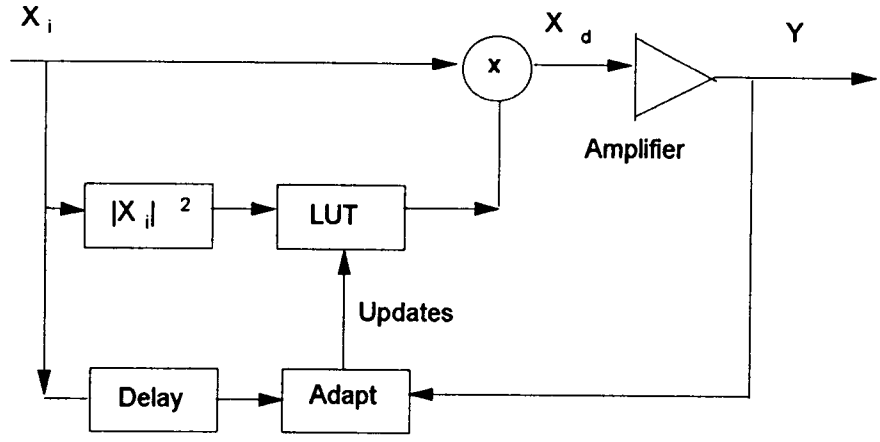

Figure 2.5 Cavers (1) predistortion system

Based on the same factor that the distortion in the power amplifier is essentially caused by amplitude variations, M. Faulkner (2) used the input amplitude as the index of an amplitude and a phase correction table. The look-up tables are much smaller than Nagata's two-dimensional table and the update to the table is much faster. This approach requires processing time to perform rectangular (I and Q) to polar (magnitude and phase) conversion.

Stapleton and Cavers (5) took another approach to implement the predistorter. They modeled the predistorter and power amplifier output as two truncated complex power series up to the fifth order. The coefficients of the power series are complex so that they can accurately represent the AM-AM and AM-PM distortion in a power amplifier. By properly choosing the coefficients of the predistorter, the distortion of the amplifier can be minimized. The adaptation is accomplished by iterative adjustment of the coefficients of the power series to minimize the distortion in the output of the amplifier. Later, Stapleton and Costesua (6) presented a detailed discussions of how to measure the distortion and use a microcontroller to adjust the coefficients.

Nested multiplication is used to implement the power series polynomial. Clearly the computation load is high in this implementation.

Digital predistortion allows more freedom in handling the cancellation of the nonlinearity since the cancellation is performed through software by using DSP techniques. The restriction is the bandwidth. Because of processing speed limitation, all the systems reported in the literature are limited to narrow-band systems. The linearizing bandwidth is similar to that of the Cartesian feedback system which is in the range of 10 to 30 KHz.

Chapter 3. Architecture Design

3.1 The design goal

The goal of this work is to design a baseband digital predistorter that can be used as a building block in an adaptive predistortion amplifier system. The work is focused on exploring the methods that can be used and the performance that can be achieved rather than meeting some particular specification.

The digital predistortion techniques discussed in Chapter 2 are all implemented in narrow-band systems. Nagata's implementation is simple but requires the use of a large memory table which would require extensive time to update. Caver, Faulkner and Stapleton's implementation require heavy computation to obtain the correction signal. These methods work well for a single channel (a few ten's of kHz) in the mobile communication band. When the bandwidth increases to the order of MHz, these techniques will not work due to the limitation on the processing speed. A new architecture is proposed in this thesis to implement a digital predistorter in wider bandwidth systems.

The predistorter should have the following features:

1. General purpose. The architecture should not be restricted to a particular amplifier or modulation type.

2. Wide bandwidth. In most mobile communication systems, the transmit and receive bands comprise a 20 MHz to 30 MHz frequency spectrum. It is preferable to have an amplifier that can cover the whole frequency band rather than a single channel or a few adjacent channels.

3. Simple interface. The predistorter should have a simple interface and be controlled by a microcontroller or personal computer.

4. Cost effectiveness. The circuit should be simple and realizable by using commercially available components. Eventually the circuit could be integrated into one or more ASICs.

3.2 Architecture design consideration

All the baseband predistortion systems reported so far use DSP microprocessors. Because of the limitation in processing speed, it is not possible to implement a predistorter with 20 MHz bandwidth by using the existing DSP implementation methods described in Chapter 2. For example, suppose a DSP processor with a 50 MHz clock rate and 12.5 MHz instruction rate is used. A look-up table structure is utilized in the predistorter. Between two data samples, the DSP processor must read data from the table and process this data (addition, subtraction or multiplication depending on the algorithm). Assuming 10 instructions are required to process a data unit, then the fastest data sample rate is 12.5 MHz/10 = 1.25 MHz. According to the Nyquist sample theorem, the sample rate must be at least twice the maximum frequency component of the real time spectrum in order to avoid aliasing. With a sample rate of 1.25 MHz the actual linearizing bandwidth is limited to 0.625 MHz maximum. That means that a DSP implementation will not be suitable for use in a wide-band predistorter.

To achieve 25 MHz of signal bandwidth, the real time processing speed must not exceed 20 nanoseconds. Direct hardware implementation of the predistorter with a pipeline structure and as little data manipulation as possible seems to be a feasible approach. The requirement for less data manipulation

also suggests the use of a look-up table rather than computation to generate predistortion signals.

As discussed in Chapter 2, predistortion correction is an open-loop operation. The correction must be performed in real time. But the adaptation is only for tracking the slow changes in the amplifier characteristic, such as the drifts with temperature and component aging. A general purpose microprocessor or personal computer can be used to handle the adaptation portion of the system.

A block diagram of the proposed system is shown in Figure 3.1. The RF band-limited input signal, X_i, is first demodulated via a quadrature demodulator.

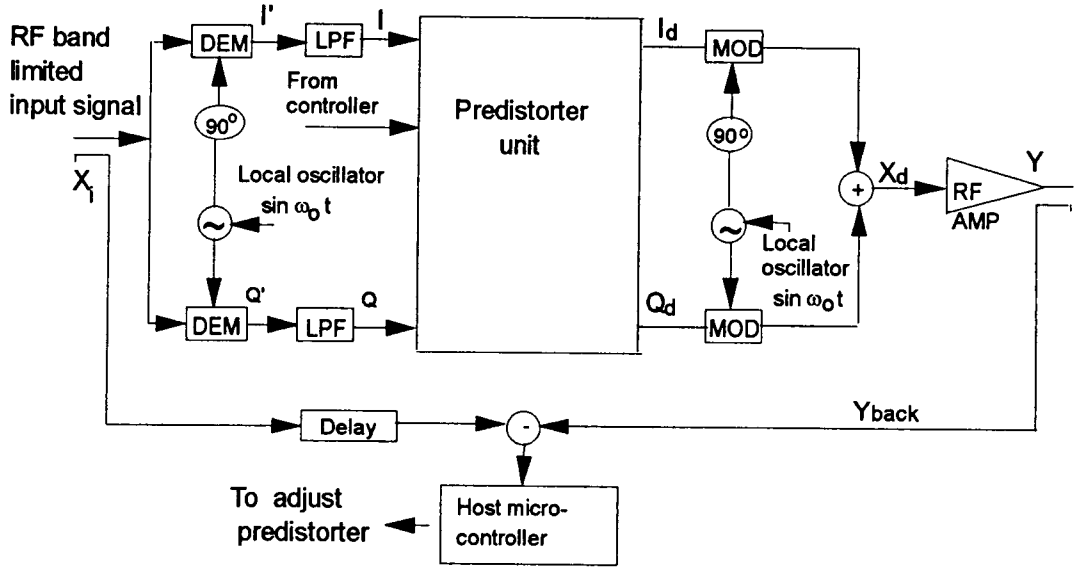

Figure 3.1 System block diagram

The quadrature demodulation process shifts the high frequency input signal to low frequency and represents the signal with two components, in phase component, I' and quadrature component, Q'.

For example, let the input RF signal be $X_i=A\sin\omega_1 t$, and the local oscillator frequency in the demodulation process be ω_o, then the outputs of the quadrature demodulation circuit are:

$$I' = X_i \cos \omega_o t$$

$$= A \sin \omega_1 t \cos \omega_o t = (A/2) [\sin (\omega_1+\omega_o)t + \sin (\omega_1 -\omega_o)t]$$

$$Q' = X_i \sin \omega_o t$$

$$= A \sin \omega_1 t \sin \omega_o t = (A/2) [\cos (\omega_1+\omega_o)t - \cos (\omega_1 -\omega_o)t]$$

The input and output spectrum are shown in Figure 3.2.

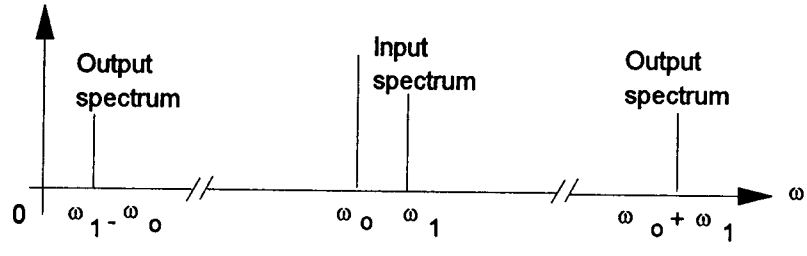

Figure 3.2 Spectrum of demodulation process

The frequency ω_o can be selected so that the signal components at frequency $(\omega_1 -\omega_o)$ which contains all the information of X_i are in the desired low frequency region, such as in a baseband. The signal components at frequency $(\omega_1 +\omega_o)$ are at approximately twice the frequency of ω_1 and thus can be suppressed by a low pass filter. The low pass filtered baseband signals, I and Q now are:

$$I = (A/2)\sin(\omega_1 -\omega_o)t$$

$Q = (A/2)\cos(\omega_1 - \omega_0)t$

Since this process shifts the signal from high carrier frequency to the low baseband frequency, it is also referred to as a frequency down-conversion process.

For an amplifier which covers one of the 25 MHz cellular bands, the local oscillator frequency ω_0 can be selected at the edge of the band, so that the whole band would be folded back into a low frequency region with the same bandwidth. Figure 3.3 shows that ω_0 is selected at the low edge of the working band, where ω_1 is the highest frequency in the band and $\omega_1 - \omega_0$ is the bandwidth.

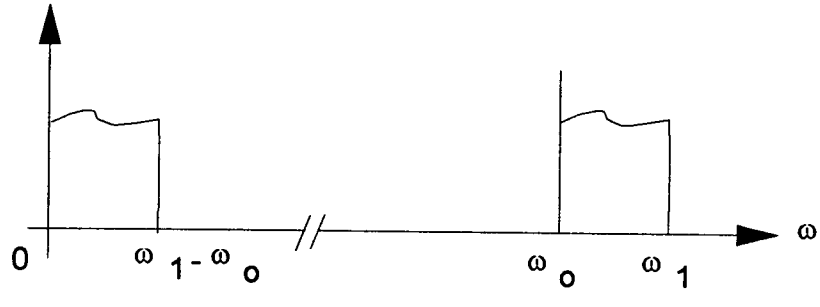

Figure 3.3 Spectrum of shifted frequency band to baseband $\omega_1 - \omega_2$

By choosing an I and Q implementation, the system can handle different types of signals because any signal can be represented by its I and Q components. The I and Q channels are symmetrical which simplifies the hardware.

After being demodulated, the I and Q signals are then fed to the digital predistorter unit. The predistorter unit predistorts the signal to I_d and Q_d according to the amplifier's nonlinearity. It then brings the signal back to the

RF carrier frequency, X_d, by a modulation process. X_d drives the amplifier. X_d is given as:

$$X_d = I_d \cos\omega_0 t + Q_d \sin\omega_0 t$$

Notice that the local oscillator frequency ω_0 here is the same as ω_0 in the demodulation process. Since the modulation process shifts the signal from baseband frequency to the high carrier frequency, it is also referred to as frequency up-conversion process.

After amplification, part of the output signal, Y_{back}, is fed back and compared with the delayed input signal to produce an error signal. The microcontroller uses this signal to adjust tables in the predistorter to minimize the error.

The predistorter unit consists of analog-to-digital (A/D) converters, data processing hardware and digital to analog (D/A) converters. According to the Nyquist sample theorem, for a 25 MHz bandwidth signal, the sample rate must be at least 50 MHz. This means that the predistorter unit needs to be able to process each data in a 20 nanosecond (ns) time period. To ensure fast processing speed, a pipeline structure is used in the predistorter unit. Two one-dimensional look-up tables are used in the unit to produce predistorted signals. The unit is controlled by a microcontroller.

A previous study by Nagata (8) has indicated that direct mapping of the predistorted signals, I_d and Q_d, from I and Q requires the use of large two-dimensional look-up table. This slows down the adaptation processes. Alternatively, the amplitude of the signal can be used to generate correction signals for both amplitude to amplitude (AM-AM) and amplitude to phase (AM-PM) distortions since both distortions are input signal level dependent. This implies that the input signal amplitude can be used as an address in the

correction table. A two-dimensional table can be reduced to a one-dimensional table. This is the approach taken in this work.

The correction look-up tables in the predistorter unit are functions of the input amplitude. A rectangular to polar (R/P) conversion is required to convert I and Q format into amplitude(ρ) and phase(θ) format. The simplest and fastest way to perform this conversion is the use of a look-up table. This conversion table is a two-dimensional table and the size may be large. The difference between a conversion table and the correction table is that no update is required in the conversion table. The size of the conversion table does not affect system performance.

3.3 The predistorter unit

3.3.1 Block diagram

Figure 3.4 shows the block diagram of the predistortion unit. The input data of the predistorter unit are the I channel and Q channel analog baseband data. They are provided by the demodulation process. I and Q are converted into digital data by a pair of A/D converters. The outputs of these two A/D converters are used as addresses of a rectangular to polar conversion look-up table. The outputs of the look-up table are amplitude, ρ, and phase, θ, which are the polar format of the input signal.

$$\rho = \sqrt{(I^2 + Q^2)}$$

$$\theta = \tan^{-1}(Q/I)$$

The amplitude signal, ρ, is used as an address for the AM-AM and AM-PM correction tables. The output of the amplitude correction table is the

predistorted amplitude signal, ρ_d. This signal is directly used as the amplitude of the drive signal for the amplifier. The phase correction signal, $\Delta\theta_i$, which is the output of the AM-PM correction table, is added to the original phase signal to produce a new phase of the drive signal, θ_d. ρ_d and θ_d are then converted back to the I and Q format through a polar to rectangular conversion look-up table. The outputs of this table are:

$I_d = \rho_d \cos\theta_d$

$Q_d = \rho_d \sin\theta_d$

I_d and Q_d are then converted back to analog format by a pair of D/A converters and up-converted to the RF carrier frequency to drive the amplifier.

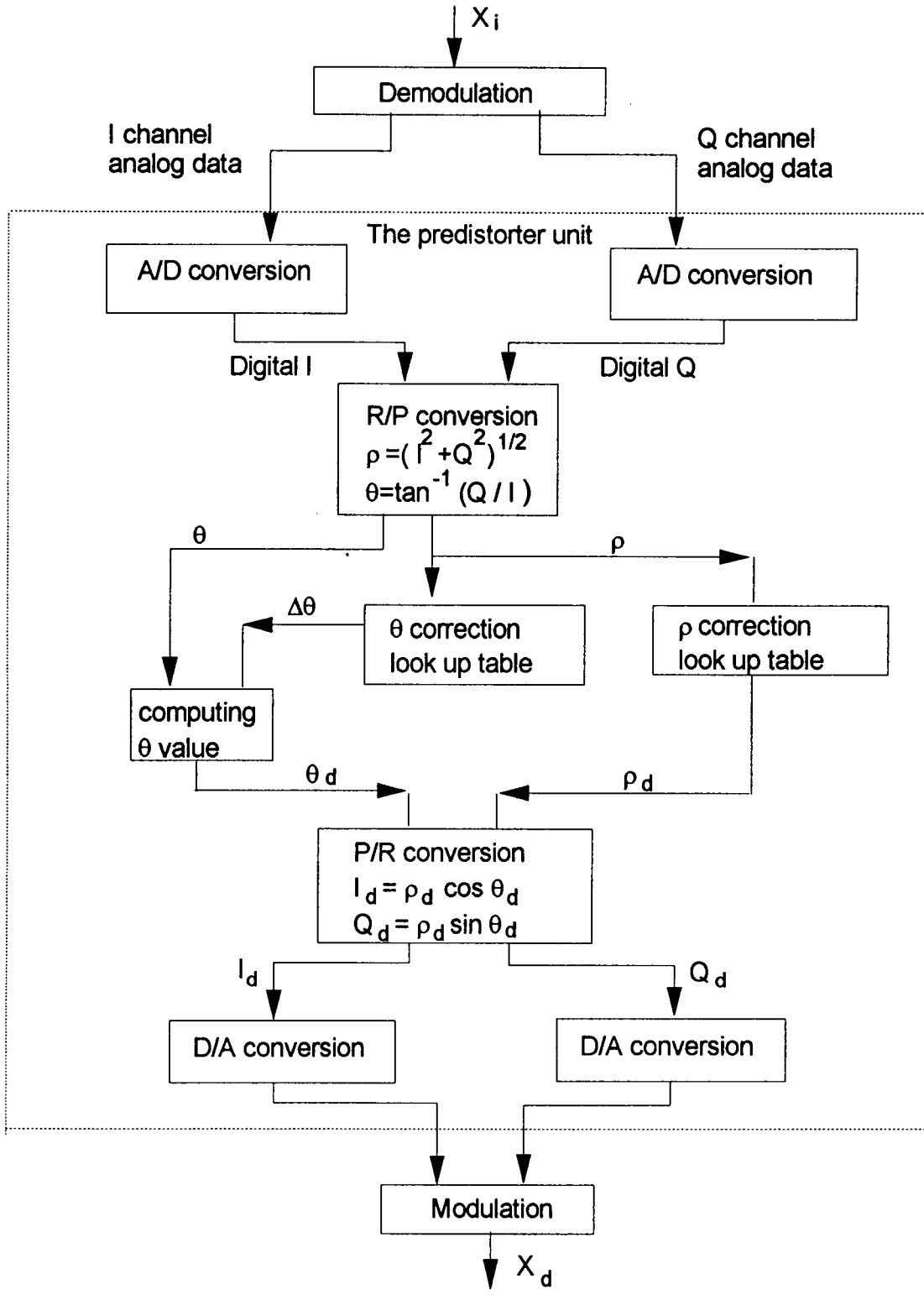

Figure 3.4 Block diagram of predistorter unit

3.3.2 Digital processing architecture

As shown in the block diagram, each data sample passes through several steps in the predistorter unit. To achieve fast data processing speed, a six-stage pipeline structure is used. Each stage of the pipeline performs one processing task per clock cycle. After the initial pipeline latency, the unit can accept and output one data value in every clock cycle. In one clock period, a pipeline segment needs to allow time for performing logic functions, register set up and signal propagation. The propagation time for a compact unit is negligible and the set up time for the fast logic register is about 2 ns. The time for performing logic functions must not exceed 18 ns for a 20 ns clock cycle. For a 20 ns clock period, the total latency of the six-stage pipeline is 120 ns. In cellular communications, the information transmitted is mainly voice signals. A time delay of 120 ns has almost no effect on the original signal and is not noticeable in human speech.

A detailed block diagram of the pipelined predistorter is shown in Figure 3.5. The inputs to the unit are I and Q, and outputs are predistorted signals, I_d and Q_d. The following section discusses the unit in detail, stage-by-stage along the pipeline. There are data registers for intermediate storage between stages.

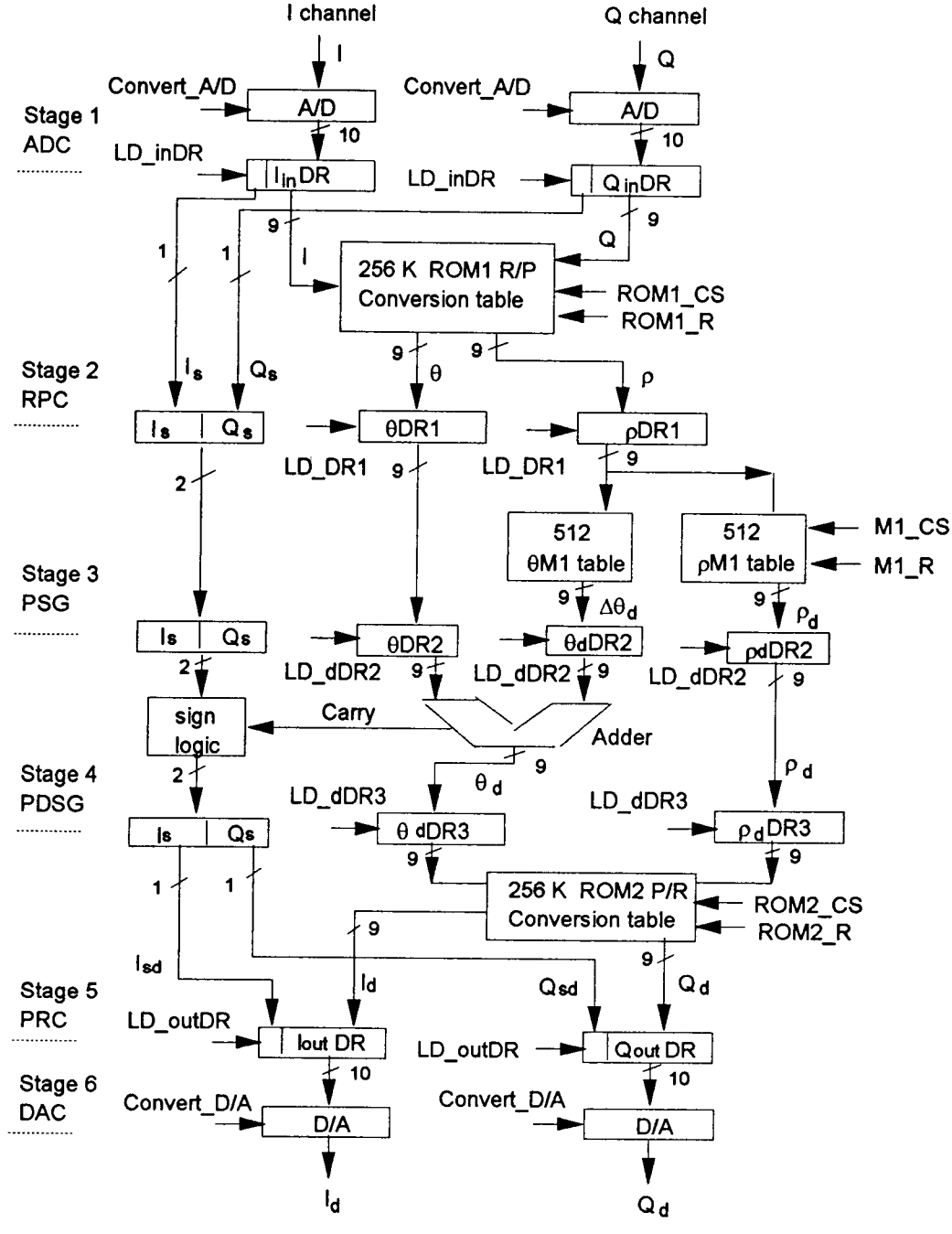

Figure 3.5 The predistorter unit

Stage one: Analog to digital conversion (ADC)

As discussed in Chapter 2, at least 10 quantization bits are required to achieve sufficient system resolution. Increasing the number of quantization bits gives the system more resolution and accuracy but raises the cost. The cost of most components are increased by resolution and speed, especially the A/D converters.

The digital signals generated by the A/D converter are stored in the I channel input data register, $I_{in}DR$, and Q channel input data register $Q_{in}DR$. $I_{in}DR$ and $Q_{in}DR$ are data sources to the next stage.

The sample rate requirement for A/D is at least 50 MHz for a 25 MHz bandwidth RF signal.

Stage two: Rectangular to polar conversion (RPC)

This stage performs a rectangular to polar (R/P) conversion to obtain the amplitude ρ and phase θ from I and Q by using a two-dimensional look-up table.

The input and output relationships of this conversion table are:

$$\rho = \sqrt{(I^2 + Q^2)}$$

$$\theta = \tan^{-1} (Q/I)$$

Where I and Q are the input data, and ρ and θ are the output data from the table. ρ and θ are only dependent on I and Q and independent of other system operation parameters. Since ρ, θ and I, Q are mathematically defined, the content of the table is fixed. The R/P conversion table can be implemented in ROM memory.

The size of the R/P conversion table is proportional to the number of bits in the I and Q signals. A 10-bit I and 10-bit Q will result in a 20 M-bit table. The table size can be reduced by taking advantage of the symmetry of the I/Q

plane, as shown in Figure 3.6. By keeping track of the quadrant the signals are located in, the ρ and θ in any location of the plane can be calculated from the ρ and θ of the first quadrant. This decreases the size of the table by three-fourths. However, some hardware is necessary to keep track of the quadrant.

If the sign-magnitude form is selected as the digital data format, the sign bit can be used to keep track of the phase while the magnitude bits can be used as the address to access the conversion table. The table is pre-computed according to the range of I and Q. The sign bits from the I and Q channels are not used in this stage and are forwarded to the next stage.

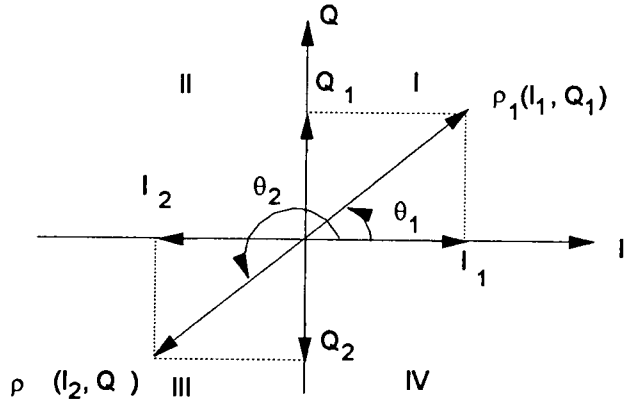

Figure 3.6 Symmetry of the I/Q plane

Two data registers, ρDR1 and θDR1, are used to hold the output data from the conversion table.

Stage three: Predistorted signal generation (PSG)

This stage generates two predistorted signals, ρ_d and $\Delta\theta_d$, using the AM-AM and AM-PM correction tables.

The contents of these two tables are pre-generated according to the amplifier nonlinearity and loaded during initialization. When the amplifier nonlinearity changes, the tables must be updated. These tables are stored in RAM memory. The size of each look-up table is 512 words for a 9 bit address. These two tables are critical to the system.

The phase(θ) of the original signal is not modified in stage four, it is passed to the next stage from θDR1$\rightarrow\theta$DR2. The sign signals are also directly passed through.

Stage four: Phase predistorted signal generation (PDSG)

An adder is used to combine the predistorted phase signal, $\Delta\theta_d$, from the AM-PM table with the original phase, θ, and produce a phase offset signal, θ_d, for the amplifier. The fast processing time requires the adder be a high performance adder with a carry look ahead circuit.

As described in Stage 2, the phase information in the data register θDR2 only reflects the phase in the first quadrant of the I/Q plane. The real phase can be formed by the sign bits of I and Q. For example, if I and Q are both greater than 0, the signal is located in the first quadrant and no correction is required; if I>0 and Q<0, the signal is located in the fourth quadrant (marked as IV in Figure 3.6) and the real phase is equal to 2π-(phase in θDR2).

Actually, there is no need to compute the absolute phase, but only to generate the correct sign bits for the I_d and Q_d signals. This can be done by monitoring the carry signal from the adder. If a carry is produced from the adder, which means the predistorted signal is shifted to the next quadrant, then the original signs of I_s and Q_s must be corrected. Logic is used to produce the correct signs, I_{sd} and Q_{sd}, according to I_s, Q_d and the carry. The relationship is shown in Figure 3.7.

In Figure 3.7, I_s and Q_s are the I-channel and Q-channel sign bits respectively. C is a carry produced by the adder. I_{sd} and Q_{sd} are predistorted I-channel and Q-channel sign bits respectively.

The output data from this stage are ρ_d and θ_d and they are stored in data registers, ρ_dDR3 and θ_dDR3.

Quadrant	I_s	Q_s	C	I_{sd}	Q_{sd}
I	0	0	0	0	0
II	1	0	0	1	0
III	1	1	0	1	1
IV	0	1	0	0	1
I	0	0	1	1	0
II	1	0	1	1	1
III	1	1	1	0	1
IV	0	1	1	0	0

Logic relations: $I_{sd} = I_s \cdot \overline{C} + C \cdot \overline{Q}_s$

$Q_{sd} = Q_s \cdot \overline{C} + C \cdot I_s$

Figure 3.7 Sign generating logic

Stage 5: Polar to rectangular conversion (PRC)

A ROM table is used to perform the polar to rectangular conversion. For the same reason as in Stage 2, sign bits are not used in memory access, but appended to the output data. The table is the same size as the table in Stage 2. The outputs of the table are:

$I_d = \rho\cos\theta$

$Q_d = \rho\sin\theta$

At this point, the digital predistorted signals, I_d and Q_d, have been generated, and they are ready to be converted back to analog signals.

Stage 6: Digital to analog conversion (DAC)

The predistorted digital signals, I_d and Q_d, are converted to analog signals by a pair of D/A converters. The analog signals then are up-converted and summed to form an RF drive signal, X_d, for the amplifier.

3.3.3 Updating the memory table

To ensure real time performance and provide adaptability, two sets of the memory tables (Table-set1 and Table-set2) are employed. One set can be used in real time while the other set can be used in the update process.

The interconnection of Table-set1 (θM1, ρM1) and Table-set2 (θM2, ρM2) is shown in Figure 3.8.

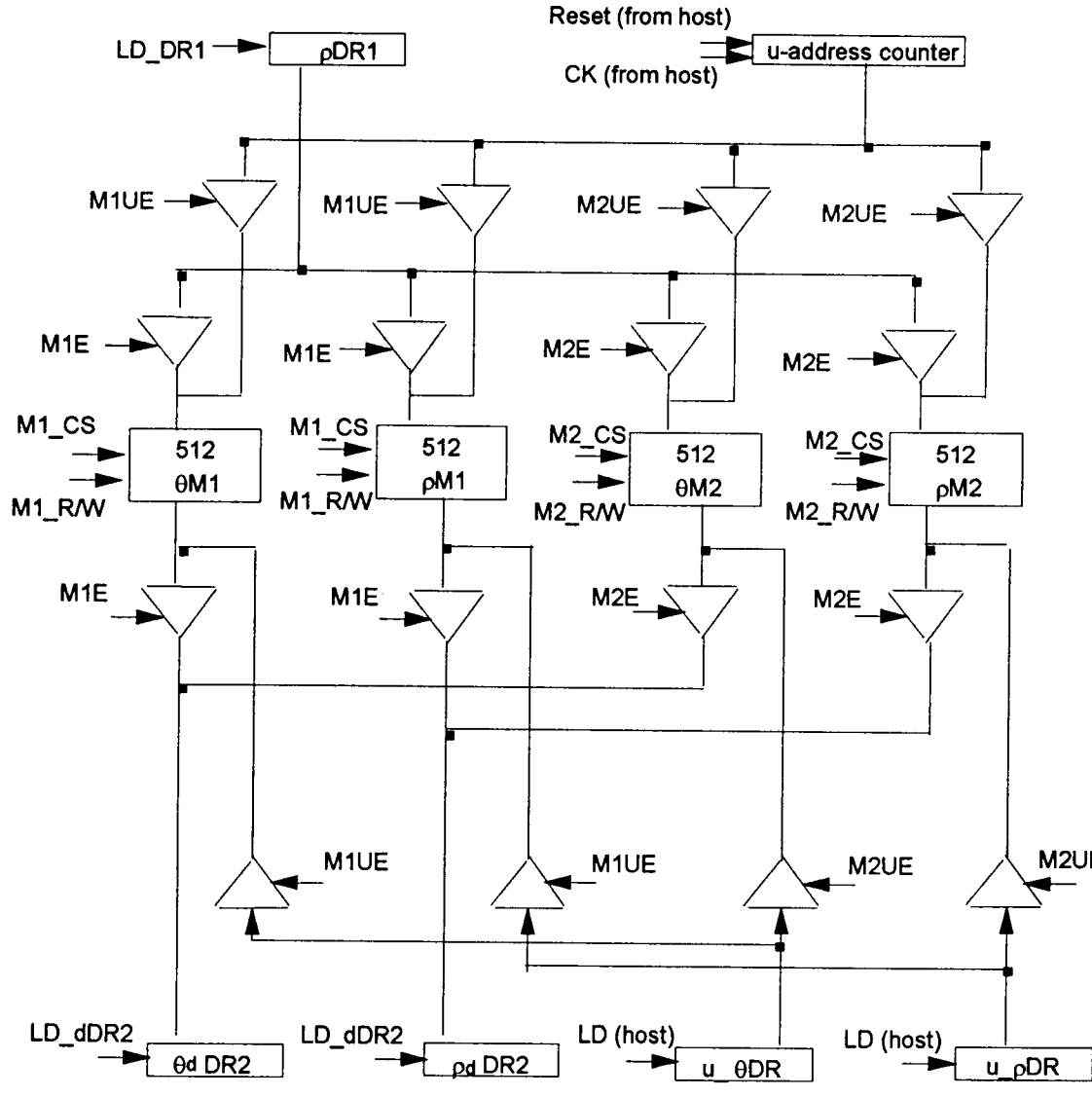

Figure 3.8 Memory look up table (Table-set1 and Table-set2)

Assume that memory Table-set1, θM1 and ρM1, is used for the real time processing, and Table-set2, θM2 and ρM2, is used for updating. The data paths are created by setting the correct control signals in the tri-state devices. The control signals, M1E (Table-set1 enable), enables the content of ρDR1 to provide an address to the θM1 and ρM1 tables. It also creates paths for the predistorted signals, ρ_d and $\Delta\theta_d$, (output from the tables), to reach data

registers ρ_dDR2 and θ_dDR2. The control signal, M2UE, connects data from the update data register, u-ρDR and u-θDR, to Table-set2. The address to M2 is generated by the update address counter, u-address counter. The update data register, update address counter and update data register are controlled by the host computer.

The host computer updates the table by sending a signal to initialize the address counter. It sends data to the update data registers. It than loads the data to the memory table. The host then advances the counter and sends another data unit. It repeats the steps until all the tables are loaded. The update to the table is sequential from location 0 to 511. When M2 is fully updated, the host computer sends a command to switch the memory units. M2 will now work in real time and M1 is ready to be updated.

With this arrangement, memory update is performed off line and does not affect the processing speed of the system. At the moment the table is switched, new corrections are applied. This results in some data using old corrections and some using new corrections. This is not a problem, because the update is mainly for slow drift of the amplifier's nonlinearity. New correction data are only slight changes of the old ones.

3.3.4 Micro-operation diagram

The predistortion unit has three operation modes, an initialization mode and two data processing modes with either Table-set1 or Table-set2 operating in real time.

The host computer sends commands to the unit to trigger changes between modes. The command is stored in a command register to generate

control signals for the unit. The control commands are encoded into 3-bit format as follow:

Command code	Name	Operation
1 0 0	START	initialization
0 0 1	M1	begin processing using Table-set1
0 1 0	M2	begin processing using Table-set2

The diagram of the interchanges between modes is shown in Figure 3.9.

When power on, the unit waits for the START command from the host computer. When START is received, the unit enters the initialization mode.

In the initialization mode, the unit resets all the data registers except the update data register. It then puts memory Table-set1 into the ready-to-receive (RTR) data state. The unit then waits until the host computer finishes loading the table into Table-set1.

After loading the table, the host computer sends a command to start data processing.

In the data processing modes, the unit receives one data unit from the A/D converter in each clock cycle. Also in each clock cycle, a data unit is processed by one pipeline stage. The first data requires 6 clock cycles to process. After that, there will be one processed data unit output per cycle. The unit either uses Table-set1 as the real time working table and Table-set2 as the update table or vice versa.

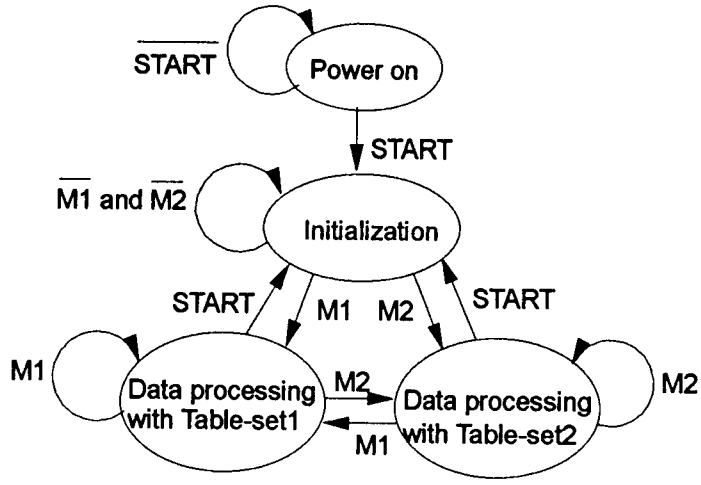

Figure 3.9 Interchanges between modes

The micro-operation diagram is shown in Figure 3.10. Each of the rectangular boxes indicates the action in each clock cycle. The upper half of the box shows the register to register transfer action and the lower half of the box indicates the required control signal to carry out the action. The control commands to switch modes are also shown.

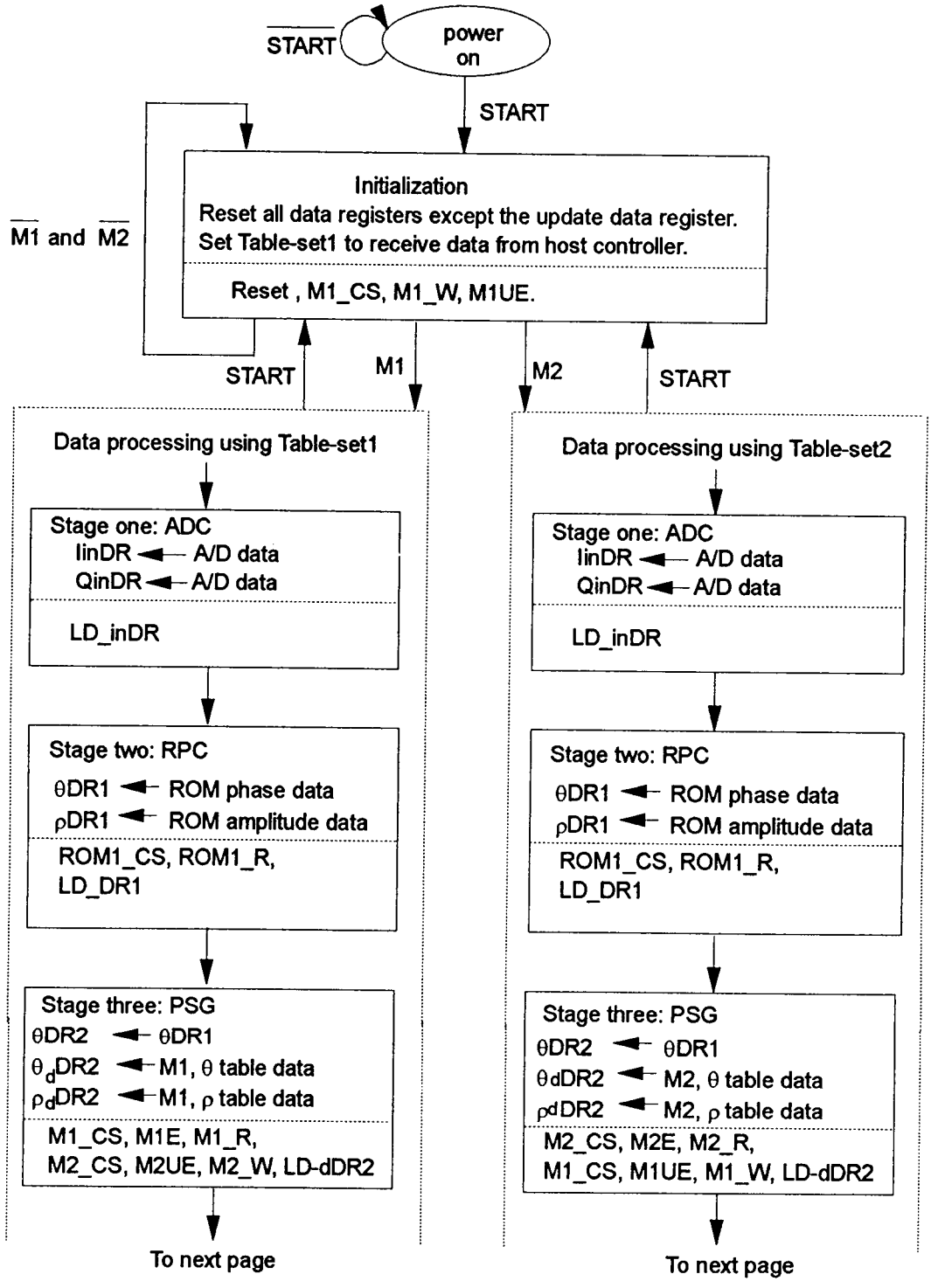

Figure 3.10 Micro-operation diagram

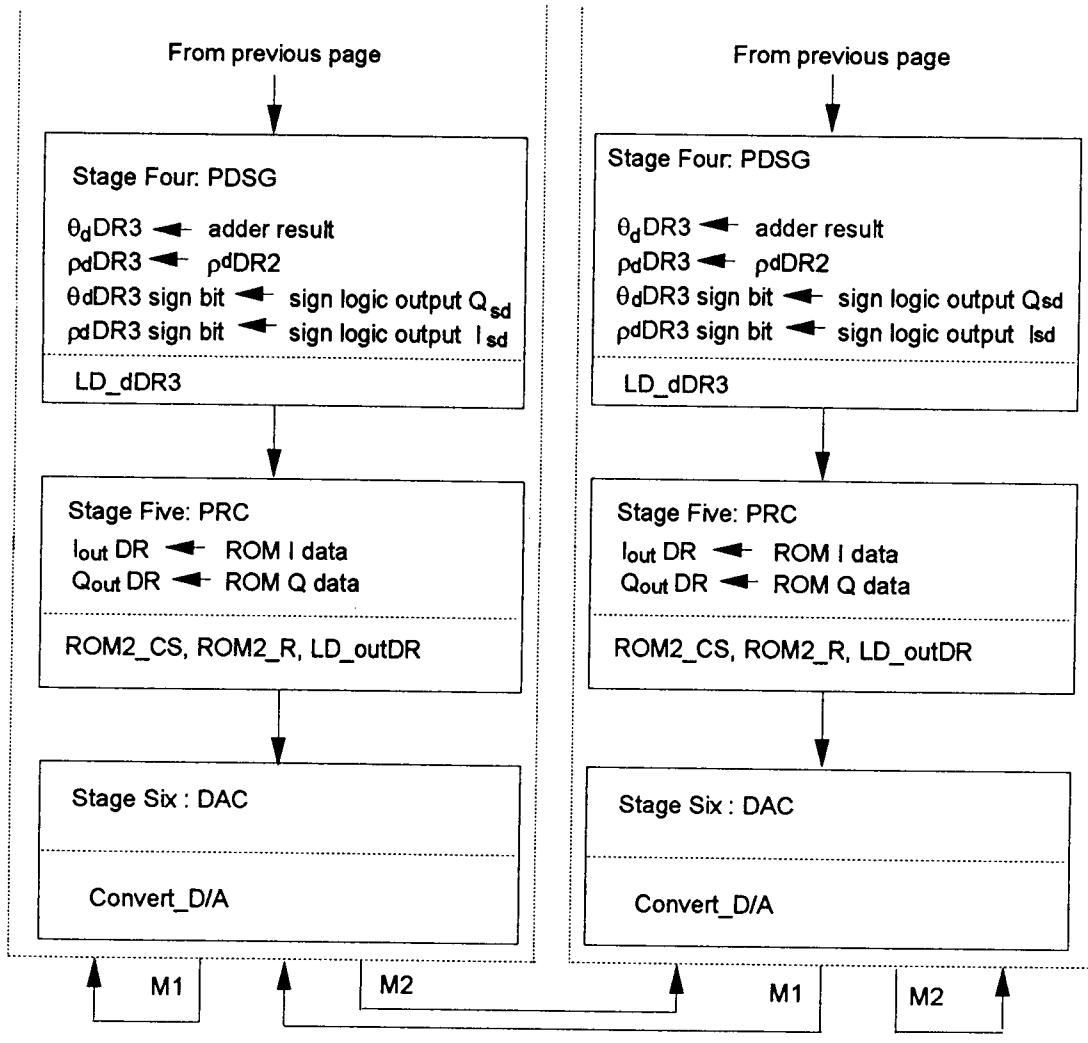

Figure 3.10 Micro-operation diagram (Continued)

3.3.5 The controller

The controller generates control signals for the predistorter unit. Since only a few operations are carried out by the unit, the controller is relatively simple. It consists of a 3-bit command register and a ROM. The control command from the host computer is stored in the command register. The output of this register provides addresses to a ROM which produces the control signals for the unit. There are a total of 20 control signals. A ROM of at

least 8 x 20 bits is enough to provide control signals. By using the ROM, the controller is much more flexible than a controller designed with hardwired logic. The content of the ROM can easily be modified to fit any design changes.

The block diagram is shown in Figure 3.11.

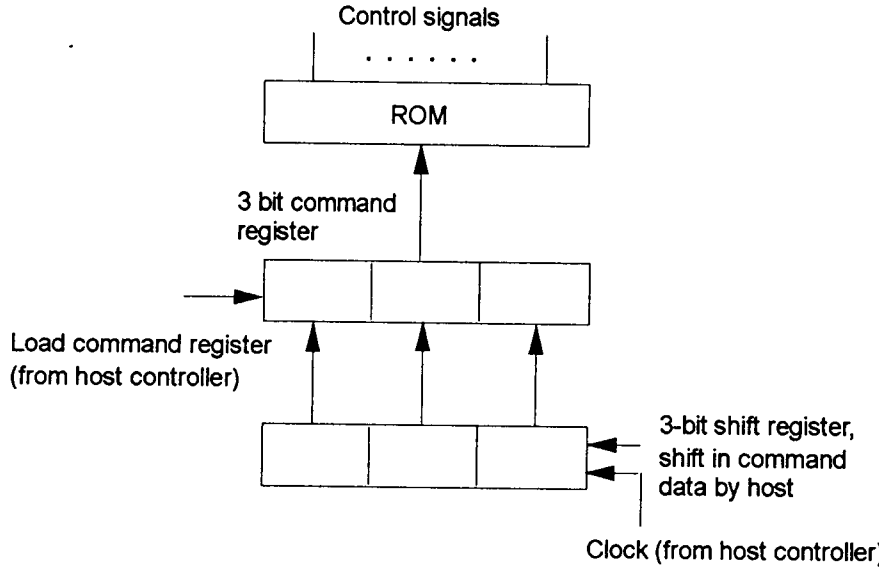

Figure 3.11 The controller

The communication between host and the unit is in serial format (this will be explained in next section). A command issued by the host computer is first shifted into a 3-bit shift register then loaded to the command register. The control signals associated with this command are then generated from the ROM. Refer to Figure 3.10, the Micro-operation diagram, for the use of the control signals.

The control signals are as follows:

	signal name	Function
1	Convert_A/D	Start A/D data conversion

2	LD_inDR	load input data registers, $I_{in}DR$ and $Q_{in}DR$
3	ROM1_CS	Rectangular to polar conversion ROM1 chip select
4	ROM1_R	ROM1 read
5	LD_DR1	Load data registers, $\rho DR1$ and $\theta DR1$
6	M1_CS	Look-up table memory Table-set1 chip select
7	M1_R/W	M1 read/write
8	M2_CS	Look-up table memory Table-set2 chip select
9	M2_R/W	M2 read/write
10	M1E	Enable Table-set1 as real time look-up table
11	M1UE	Enable Table-set1 as update table
12	M2E	Enable Table-set2 as real time look-up table
13	M2UE	Enable Table-set2 as update table
14	LD_dDR2	Load predistorted data registers, $\rho_d DR2$ and $\theta_d DR2$
15	LD_dDR3	Load predistorted data registers, $\rho_d DR3$ and $\theta_d DR3$
16	ROM2_CS	Polar to rectangular conversion ROM2 chip select
17	ROM2_R	ROM2 read
18	LD_outDR	Load output data registers, $I_{out}DR$ and $Q_{out}DR$
19	Convert_D/A	Start D/A data conversion.
20	Reset	Reset all data registers except the update register

During the data processing operation, the control signals for each stage are issued in every clock cycle because of the pipeline structure. Therefore, some of the control signals can be shared. For example, all the register loading signals, LD_inDR, LD_DR1, LD_dDR2, LD_dDR3, and LD_outDR can be reduced to a single load signal. Similarly, memory chip select signals, A/D and D/A convert signals can be reduced. The total number of the control signals are reduced to 11. The reduced control signals are as follows:

New signal	Replaced old signals
Convert	Convert_A/D, Convert_D/A
LD_REG	LD_inDR, LD_DR1, LD_dDR2, LD_dDR3, LD_outDR
MEM_CS	ROM1_CS, ROM2_CS, M1_CS, M2_CS
ROM_R	ROM1_R, ROM2_R

Actually, because of the pipeline structure, control signals are issued at each cycle. Only three different control signals, each associates with an operation mode, are required. For the purpose of clarity, the name of the control signals remain unchanged.

3.3.6 Interface to host computer

The unit is designed to interface with a general purpose microcontroller or a personal computer. The host uses an 8-bit parallel port to control the unit. Since speed is not a concern, data from the host computer to the unit are in a serial format. Data transmission is controlled by the host using a bit-bashing technique. The function of each interface bit is defined as follows:

bit number	function
1	command data bit
2	command shift register clock bit
3	load command register
4	reset address generate counter
5	address counter increment
6	memory table data bit
7	memory table shift register clock
8	load memory data to update data register

There are four registers and one address counter in the unit that can be controlled by the host. Figure 3.12 shows the interface to the host.

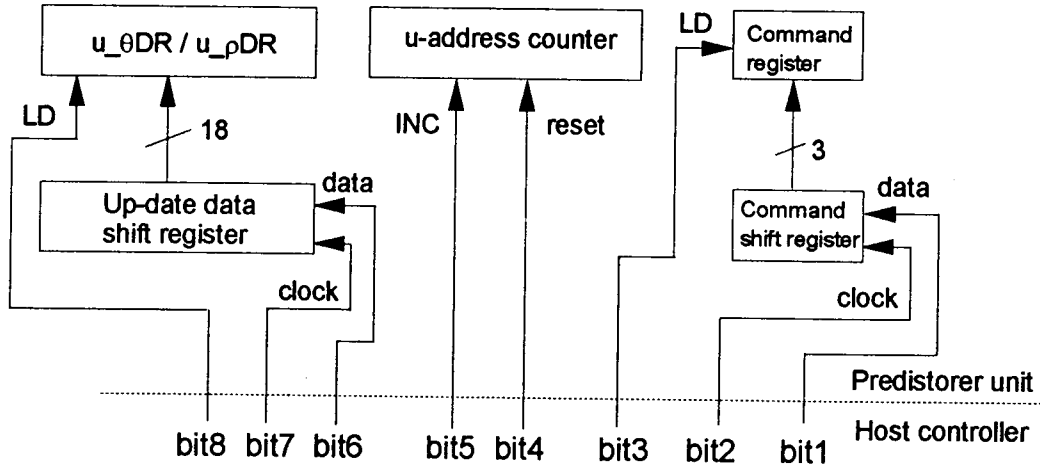

Figure 3.12 Interface to host computer

The 3-bit command shift register receives serial data from host. It loads the data to the command register in parallel. Three interface bits are required, bit1 carries data, bit2 is a shift clock and bit3 is a load signal. Similarly, bit6, bit7 and bit8 are for loading the update data registers. As indicated in Figure 3.7, a u-address counter is used to provide address for update the LUT. The update is performed in sequential order from lower address to higher address. Bit4 and bit5 are for resetting and incrementing the u-address counter respectively.

The data transfer to the idle memory table is controlled by the host controller while the unit is working at its own real-time speed with the active memory table.

When the system starts to run, the host computer first sends a start command to initialize the predistorter unit. The unit executes this command, resets all data registers and sets the memory Table-set1 ready for receiving data from the host. Next, the host asserts bit4 to reset the address counter. It then sends a bit-string of one data word to the update data shift register. This word is then transferred in parallel to the u_θDR and u_ρDR and loaded to Table-set1. This completes one data word update cycle. The host computer then advances the memory address by asserting the counter INC bit, bit5 and starts next data word update cycle until all data are loaded into Table-set1. When Table-set1 is fully loaded, the host sends a command, M1, to start data processing mode.

The host computer monitors the output of the amplifier, generates a modified table and loads the new table to Table-set2 while the unit is working with the LUT in Table-set1. When Table-set2 is updated, the host sends a command to switch from Table-set1 to Table-set2. The host computer continuously monitors the performance of the amplifier and makes adjustments to the LUT. The LUT is updated without interrupting the real-time operation. The rate at which the table needs to be updated depends on the optimization algorithm and components drift. This needs further study. Since the adjustment is mainly for drift of the amplifier's nonlinearity, the update speed should be on the order of minutes.

Chapter 4. Simulation Performance

4.1 Simulation model

The predistortion amplifier system can be modeled as two cascaded blocks, the predistorter block and the amplifier block.

The quadrature demodulation and modulation in the system do not participate in the amplifier nonlinearity cancellation process. These functions are used to down-convert the RF signal to the baseband and up-convert the predistorted signal back to RF. They are assumed to be ideal linear functional blocks when the nonlinearity cancellation of the amplifier is discussed.

4.1.1 Amplifier and predistorter model

A commonly used model for an amplifier with a weak nonlinearity is the power series (16). The instantaneous output, e_o, of the amplifier with input, e_i, is given as:

$$e_o = a_1 e_i + a_2 e_i^2 + a_3 e_i^3 + \ldots\ldots = \sum_{k=1}^{\infty} a_k e_i^k$$

Where the a_k's are device-dependent coefficients.

For most nonlinear cellular amplifiers the coefficients a_k decrease with k. Thus, the contribution of $a_k e_i^k$ to the energy of e_o from higher order terms (k>1) is small compared to the linear term (k=1). Coefficients of the higher order terms are usually rather small compared with those first few terms. Therefore, the first few terms are normally sufficient to describe the amplifier (19). Using more terms gives more accuracy but adds computation complexity. A power

series up to the fifth order gives a reasonable compromise between accuracy and complexity, and is the model used here. Higher order terms are assumed to be negligible.

Consider the amplifier input, $x_d = A\cos\omega t$, then y can be written as:

$$y = a_1 A\cos\omega t + a_2 A^2\cos^2\omega t + a_3 A^3\cos^3\omega t + a_4 A^4\cos^4\omega t + a_5 A^5\cos^5\omega t$$

$$= [(1/2)a_2A^2 + (3/8)a_4A^4] + [a_1A + (3/4)a_3A^3 + (5/8)a_5A^5]\cos\omega t$$

$$+ [(1/2)a_2A^2 + (1/2)a_4A^4]\cos 2\omega t + [(1/4)a_3A^3 + (5/16)a_5A^5]\cos 3\omega t$$

$$+ [(1/8)a_4A^4]\cos 4\omega t + [(1/16)a_5A^5]\cos 5\omega t$$

The output signal consists of a component at the fundamental frequency, ω, and spurious signals at the dc, the second, third, fourth and fifth harmonic frequencies, 2ω, 3ω, 4ω, and 5ω. The second and higher harmonic components in y are normally not of interest for cellular RF amplifier since they are outside the working frequency band and can be eliminated by a band pass filter.

The fundamental component of y has an amplitude of $(a_1A + 3/4\ a_3A^3 + 5/8\ a_5A^{5)}$. The gain of the amplifier at the fundamental frequency is given as:

$$g = (a_1A + 3/4\ a_3A^3 + 5/8\ a_5A^{5)}/A$$

$$= a_1 + 3/4\ a_3A^2 + 5/8\ a_5A^4$$

g is usually expressed in dB. This is:

$$G = 20\log(g)$$

The linear gain of the amplifier is defined as:

$$G_1 = 20\log(a_1A/A) = 20\log(a_1)$$

If a_3 and a_5 are negative, the resulting amplitude of the fundamental component is smaller than the linear amplitude, a_1A, of the amplifier, especially at high input level. This property is called gain compression. Most RF amplifiers are compressive. Gain compression generates AM-AM distortion.

When the input signal amplitude is small, the terms, 3/4 a_3A^3 and 5/8 a_5A^5 are negligible. a_1A is a dominant term to the fundamental component and the amplifier behaves linearly. With increases in the input amplitude, the higher order terms play a more important role, and gain compression becomes more significant. When the input amplitude reaches a certain level, the output will not increase further. The amplifier is said to be saturated. The output power of an amplifier is usually defined as the output power at which the gain drops by 1 dB compared to the small input signal gain. This point is referred as the 1 dB gain compression point.

Notice, only the a_k's of the odd order terms in the power series generate the gain compression. Even order terms have no effect on the amplitude of the fundamental component. For this reason, the amplifier is modeled as:

$$y = a_1x_d + a_3x_d^3 + a_5x_d^5$$

Where y and x_d are the output and input, and a_1, a_3, and a_5 are the coefficients of the amplifier model.

Now consider the predistorter. Let the input to the predistorter be x_i, then the output, x_d, can be represented in power series as:

$$x_d = b_1x_i + b_2x_i^2 + b_3x_i^3 + b_4x_i^4 + b_5x_i^5 + ...$$

where b's are the coefficients which are chosen to cancel the nonlinearity of the amplifier. Since only the odd order terms generate gain compression, there is no need for compensation by the even predistortion terms. So all b_i's for which i = 2, 4, 6, ... can be set to zero. For the same consideration as in the amplifier, the predistorted power series is truncated to the fifth order as:

$$x_d = b_1x_i + b_3x_i^3 + b_5x_i^5$$

4.1.2 Predistortion system model

As shown in Figure 4.1, two truncated power series are used to represent the predistorter block and the amplifier block in the system. They are:

$$x_d = b_1 x_i + b_3 x_i^3 + b_5 x_i^5$$

$$y = a_1 x_d + a_3 x_d^3 + a_5 x_d^5$$

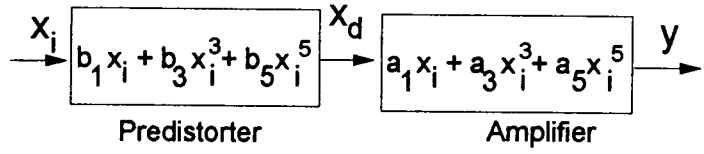

Figure 4.1 The predistortion system model

In practice the amplifier model is obtained from measured characteristic. A least square error fitting algorithm is used to generate a_1, a_3, and a_5. This analytical function is then used to determine the initial coefficients of the predistorter.

Assuming perfect up and down conversion, then the overall system input and output can be represented as:

$$y = a_1(b_1 x_i + b_3 x_i^3 + b_5 x_i^5) + a_3(b_1 x_i + b_3 x_i^3 + b_5 x_i^5)^3 + a_5(b_1 x_i + b_3 x_i^3 + b_5 x_i^5)^5$$

$$= C_1 x_i + C_3 x_i^3 + C_5 x_i^5 + C_7 x_i^7 + \text{higher order terms}$$

where the overall system power series coefficients are:

$$C_1 = a_1 b_1$$

$$C_3 = a_1 b_3 + a_3 b_1^3$$

$$C_5 = a_1 b_5 + a_5 b_1^5 + 3 a_3 b_1^2 b_3$$

...

By properly selecting the coefficients of the predistorter b_i's, the third and the fifth order terms in the overall system can be canceled out or reduced. Perfect cancellation of the nonlinearity requires the use of an infinite number of power series terms. This is not feasible. In most RF amplifiers, the higher order terms are small and can be omitted. Based on this assumption, only the third and fifth orders are considered here.

The cancellation conditions for the third and fifth order are given as:

$a_1 b_3 + a_3 b_1^3 = 0$

$a_1 b_5 + a_5 b_1^5 + 3a_3 b_1^2 b_3 = 0$

Choosing $C_1 = a_1$, then $b_1 = 1$, and b_3 and b_5 are given as:

$b_3 = -a_3 / a_1$

$b_5 = (-a_5 - 3a_3 b_3) / a_1$

These conditions are used as the initial start up amplitude to amplitude (AM-AM) correction in the predistorter. They are adjusted by an adaptive process to optimize the correction.

Since the amplifier also has amplitude to phase (AM-PM) distortion, a correction on AM-PM is also required in the predistorter. The measured AM-PM characteristic, $\theta = a'_1 x_d + a'_3 x_d^3 + a'_5 x_d^5$, of the amplifier is used to determine the initial AM-PM compensation in the predistorter.

The predistorter uses two look-up tables to produce the AM-AM and AM-PM correction signals, as described in Chapter 3.

The amplifier used in this study is a three-stage class A power amplifier, Celwave model PA9340 used in Cellular base stations. Its linear gain is 32 dB and designed output power is 6 Watt. The amplifier is operated from 869 MHz to 894 MHz with 50Ω resistive source and load impedance. The output power

at the 1 dB compression point of PA9340 is measured as 38.29 dBm at 880 MHz, the saturated power is 40.8 dBm.

The measured amplitude and phase characteristics of the PA9340 with the fifth order least squared fitting curves are shown in Figure 4.2 and Figure 4.3 respectively. The characteristics are measured with an HP model 8753A network analyzer. The fitting curves are generated by a minimum mean squared error fitting algorithm.

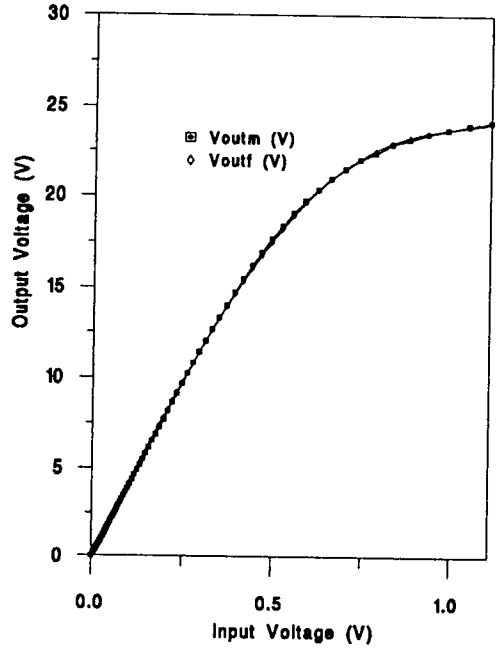

Figure 4.2 Amplitude characteristic of PA9340

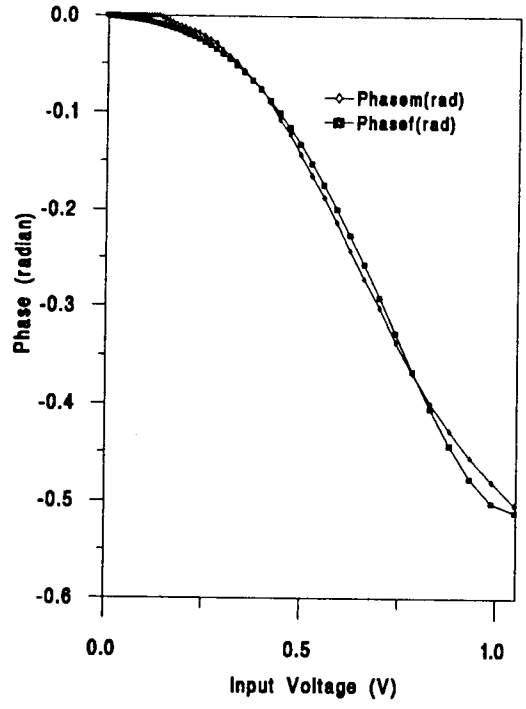

Figure 4.3 Phase characteristic of PA9340

In Figure 4.2, Voutm (V) is the measurement amplitude characteristic. Voutf (V) is the fifth order least squared fit curve. The coefficients of the AM-AM function are: a_1=40.61, a_3=-23.21, a_5= 6.38.

In Figure 4.3, Phasem(rad) is the measurement phase characteristic. Phasef(rad) is the fifth order least squared fit curve. The coefficients of the AM-PM function are: a'_1=-0.049, a'_3=-1.038, a'_5= 0.581.

4.2. The predistortion tables

4.2.1 AM to AM correction table

The following considerations are used to determine the input range on the amplifier for proper operation of the predistortion.

From Figure 4.2, it can be seen that the non-linearity increases drastically near the saturated output power region. It is difficult to try to linearize an amplifier up to that area. Cavers (1) pointed out that in practice the amplifier could be linearized up to 95 % - 98 % of its saturated power.

Let the maximum output power, P_{max}, at which the system tries to linearize be 95% of the saturated power, P_{sat}, of the PA9340. $P_{max.}$ in dBm is given by:

$$P_{max} = P_{sat} + 10 \log (0.95)$$
$$= P_{sat} - 0.22 \text{ dBm}$$
$$= 40.8 - 0.22 \text{ dBm}$$
$$= 40.58 \text{ dBm}$$

The P_{max} limits the input power, P_d, of the amplifier to P_{dmax}. Since P_d is from the output of the predistorter, P_{dmax} sets the output limit of the predistorter.

From the measurement data, shown in Figure 4.4, P_{dmax} is equal to 13.4 dBm with P_{max} at 40.58 dBm. In a 50 Ω system, the power is equivalent to the voltage amplitude of:

$$\rho_{dmax}(\text{rms}) = \sqrt{(10^{(13.4/10)} \times 0.001 \times 50)} = 1.05 \text{ V}$$

because: $P(\text{in dBm}) = 10\log(1000 * V^2/R)$

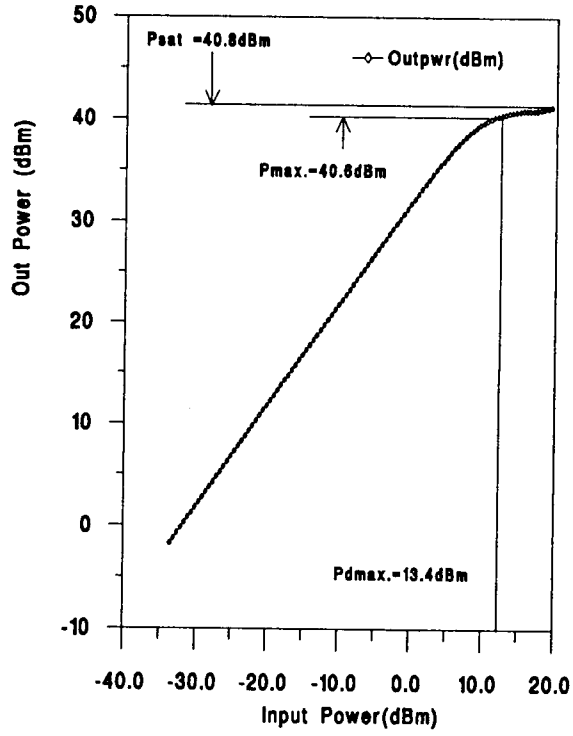

Figure 4.4 Maximum In/Out power of PA9340

The maximum input power, P_{inmax}, with respect to P_{max} of the system is set by the linear gain and the P_{max} of the amplifier as:

$P_{inmax} = P_{max}$ - linear gain (in dBm)

In practice, the linear gain is measured with a low level input, since at low signal levels the amplifier behaves linearly. With the linear gain of 32 dB, P_{inmax} is 8.6 dBm. This is equivalent to ρ_{inmax} = 0.60 V. Figure 4.5 illustrates the maximum input and output limits.

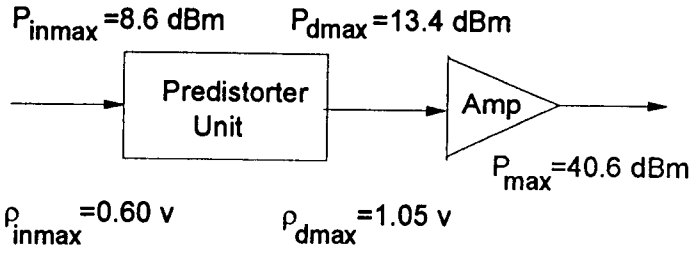

Figure 4.5 Illustration of maximum input/output limit

ρ_{inmax} and ρ_{dmax} are the maximum input voltage and output voltage of the amplitude correction table respectively. Let the input and output of the table be ρ_{in} and ρ_{out} respectively, then the input and output relation of the table is given as:

$$\rho_d = b_1\rho_{in} + b_3\rho_{in}^3 + b_5\rho_{in}^5$$

where ρ_{in} and ρ_{out} are bound by ρ_{inmax} and ρ_{dmax} respectively. The b_is are coefficients of the predistorter which can be adjusted by the adaptive process to optimize the nonlinearity cancellation. The initial value of b_is are: b_1=1, b_3=0.5715, b_5=0.823.

For example, if ρ_{in} = 0.20 V, ρ_d will be 0.22V. This 0.22 is stored in AM-AM table at the location corresponding to ρ_{in} = 0.20V, as shown in Figure 4.6.

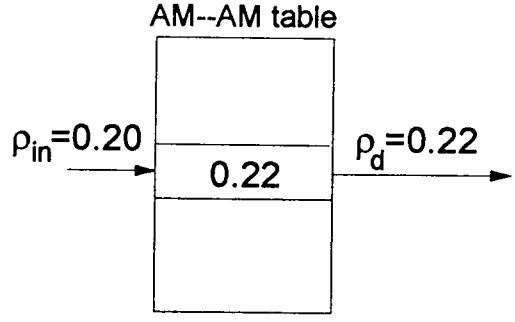

Figure 4.6 The relation of ρ_{in} and ρ_d

4.2.2 AM to PM correction table

As shown in Figure 4.3, the phase drops with an increase in the signal level. To keep the input phase and output phase equal, a phase shift, $\Delta\theta$, is added to the input phase of the signal for each input level. $\Delta\theta$ is stored in the AM-PM correction table and added to the original phase, θ_{in}, to generate the phase offset signal, θ_d, for the amplifier.

Let the input and output of the table be ρ_{in} and $\Delta\theta$ respectively, then the input and output relation of the table is given as:

$$\Delta\theta = b'_1\rho_{in} + b'_3\rho_{in}^3 + b'_5\rho_{in}^5$$

The b'_is are determined initially by the AM-PM relation of the amplifier. They can be adjusted by the adaptive process later. The initial value of b'_is are: b'_1=0.049, b'_3=1.038, b_5=-0.581.

For example, at ρ_{in} = 0.30V, the phase shift of the amplifier is -0.036 radian. A $\Delta\theta$ = 0.143 radian is stored in the location corresponding to 0.30V input amplitude. Figure 4.7 shows the relationship of ρ_{in}, θ_{in}, $\Delta\theta$ and θ_d.

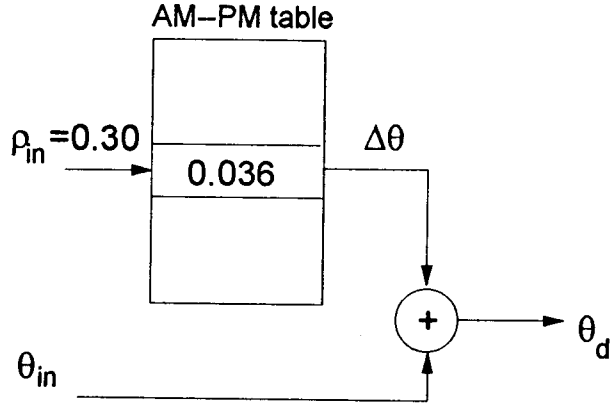

Figure 4.7 The relation of ρ_{in}, θ_{in}, and θ_d

4.3 Single signal test

One way to characterize distortion at high power levels uses a 1 dB gain compression point as discussed in 4.1.1. In practice, the 1 dB compression point is measured by increasing the input power level until the gain drops by 1 dB with respect to the small signal gain. The output power at that point is defined as the output power of the amplifier.

The output characteristic of the overall system with predistortion is obtained by simulation. The simulation program is written in MATLAB. The detailed information and simulation program is listed in Appendix C. In the simulation, the input signal is first passed through an I/Q demodulator and low pass filter to generate the I and Q channel signals. Let the input signal be X_i, the output of the I/Q demodulator be I' and Q', then:

$X_i = A\sin\omega_1 t$

$I' = X_i \cos \omega_0 t$

$= A \sin \omega_1 t \cos \omega_0 t = (A/2) [\sin (\omega_1+\omega_0)t + \sin (\omega_1 -\omega_0)t]$

$Q' = X_i \sin \omega_0 t$

$= A \sin \omega_1 t \sin \omega_0 t = (A/2) [\cos (\omega_1+\omega_0)t - \cos (\omega_1 -\omega_0)t]$

The low pass filtered I and Q are:

$I = (A/2)\sin(\omega_1 -\omega_0)t$

$Q = (A/2)\cos(\omega_1 -\omega_0)t$

The I and Q signals then converted into polar format, ρ and θ using:

$\rho = \sqrt{(I^2 + Q^2)}$

$\theta = \tan^{-1}(Q/I)$

The ρ and θ are used to generate predistorted signal ρ_d, $\Delta\theta$ and θ_d by:

$\rho_d = b_1\rho + b_3\rho^3 + b_5\rho^5$

$\Delta\theta = b'_1\rho + b'_3\rho^3 + b'_5\rho^5$

$\theta_d = \Delta\theta + \theta$

Where bi's are the coefficients that discussed early in this chapter. These values are adjusted by adaptive process later to optimize the performance. The ρ, θ and ρ_d, $\Delta\theta$, θ are 10-bit data.

The predistorted signals then converted back to I and Q format by:

$I_d = \rho_d \cos\theta_d$

$Q_d = \rho_d \sin\theta_d$

I_d and Q_d are then modulated back to X_d to drive the amplifier.

$X_d = I_d \cos\omega_o t + Q_d \sin\omega_o t$

This signal is fed into the amplifier to produce the overall system output. The real measured data of the amplifier are used in the simulation. Interpolation is used to calculate the data point between measurement points.

The amplitude of the input signal is varied from 0 to maximum allowable level to generate response of the system. Figures 4.8 and 4.9 show the amplitude and phase responses of the overall system respectively in comparison with the PA9340.

The simulation result indicates that the 1 dB compression point of the overall system with predistortion does not appear until the system is saturated. The output power at this point is 40.53 dBm. This is very close to the expected maximum output power, P_{max}=40.58dBm, as discussed in the previous section. In other words, the output power of the amplifier with predistortion is increased from 38.29 dBm to 40.53 dBm. Notice, it is difficult to linearize the amplifier up at the saturation region using only a fifth order predistorter model. The gain drops when close to saturation. A higher order model is required to reduce the gain drops. But at this region, the output power is limited by

saturation. Using a higher order model would not have much effect. The achieved improvement is the increase of the linear range. The amplifier can be operated almost up to saturation, so the 6 Watt amplifier is equivalent to a 10 Watt (40dBm = 10 Watt) amplifier.

The 1 dB compression point analysis is simple but does not give a direct measurement of the amplifier performance when the amplifier amplifies two or more signals simultaneously. In this case, a two-tone test is preferred.

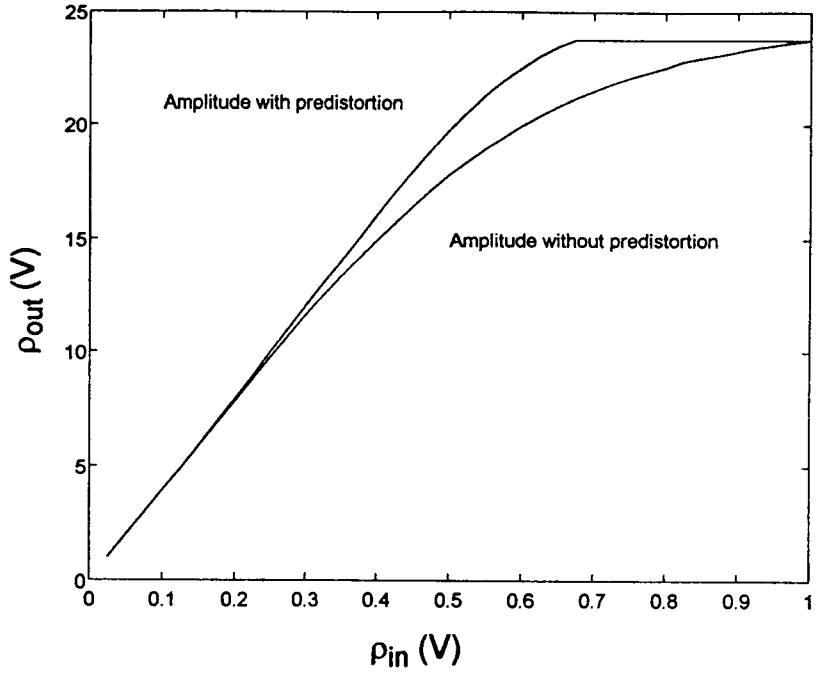

Figure 4.8 Amplitude characteristic of the PA9340 with/without predistortion

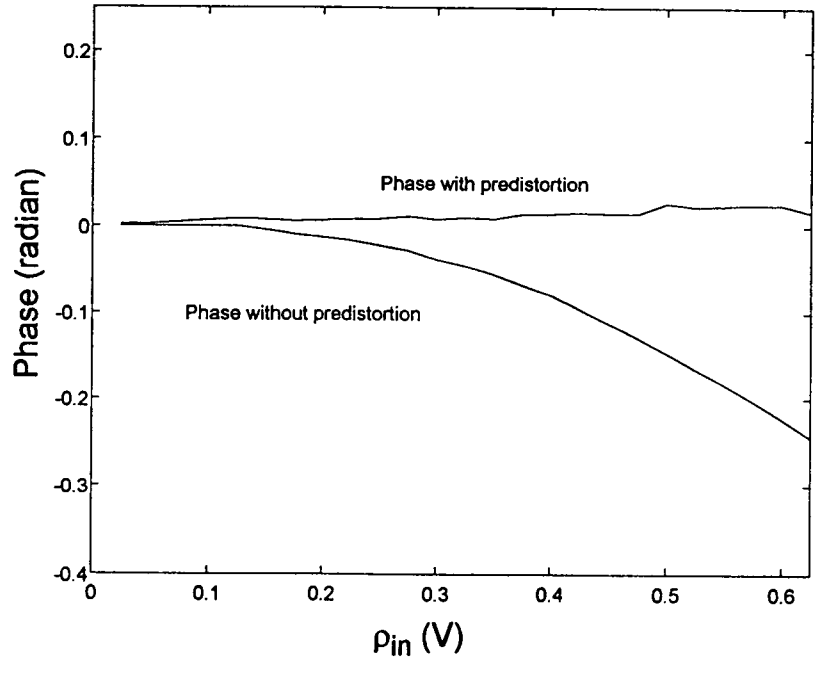

Figure 4.9 Phase characteristic of the PA9340 with/without predistortion

4.4 Two tone test

When multiple signals are passed through an amplifier, the nonlinearity of the amplifier causes intermodulation products (IMs) to be generated. One common method to measure the intermodulation product is by exciting the amplifier with two signals of equal amplitude and different frequencies, ω_1 and ω_2, and measuring the level of the IMs. These two signals are close in frequency and both within the amplifier's bandwidth.

Consider an input signal $x_i = A(\cos\omega_1 t + \cos\omega_2 t)$. Applying x_i to the system's fifth order power series yields:

$$\begin{aligned} y = {} & [C_1A + (9/4)C_3A^3 + (50/8)C_5A^5]\cos\omega_1 t \\ & + [C_1A + (9/4)C_3A^3 + (50/8)C_5A^5]\cos\omega_2 t \\ & + [(3/4)C_3A^3 + (25/8)C_5A^5]\cos(2\omega_1 - \omega_2)t \\ & + [(3/4)C_3A^3 + (25/8)C_5A^5]\cos(2\omega_2 - \omega_1)t \\ & + [(5/8)C_5A^5]\cos(3\omega_1 - 2\omega_2)t + [(5/8)C_5A^5]\cos(3\omega_2 - 2\omega_1)t \\ & + \text{DC term and higher frequency terms} \end{aligned}$$

where the C_is are coefficients of the predistorted amplifier system.

The output signal consists of components at dc, the fundamental frequencies, ω_1 and ω_2, the second to fifth harmonics, $2\omega_1$, $2\omega_2$, $3\omega_1$, $3\omega_2$, $4\omega_1$, $4\omega_2$, $5\omega_1$, $5\omega_2$ and the second to fifth-order intermodulation products at $\omega_1 \pm \omega_2$, $2\omega_1 \pm \omega_2$, $2\omega_2 \pm \omega_1$, $2(\omega_1 \pm \omega_2)$, $4\omega_1 \pm \omega_2$, $4\omega_2 \pm \omega_1$, $3\omega_1 \pm 2\omega_2$, $3\omega_2 \pm 2\omega_1$. The order of the IM is determined by the sum of the coefficients of ω_1 and ω_2. For example, the third order IMs are produced by the terms $2\omega_1 \pm \omega_2$ and $2\omega_2 \pm \omega_1$.

In cellular systems, the spurious signals with frequencies at $2\omega_1 - \omega_2$, $2\omega_2 - \omega_1$, and $3\omega_1 - 2\omega_2$ and $3\omega_2 - 2\omega_1$, referred to as the third and fifth order IMs, will fall

within the pass-band and can distort the in-band signal. The other spurious signals are far outside the pass band and can be filtered out.

For example, if an input signal consisting of two equal amplitudes and different frequencies, 840 MHz and 842 MHz, is applied to an amplifier with a frequency band of 824 MHz to 849 MHz, the output of the amplifier will consist of third order IMs at frequencies:

$2f_1 - f_2 = 2*840 - 842 = 838$ MHz

$2f_2 - f_1 = 2*842 - 840 = 844$ MHz

The fifth order IMs at frequencies:

$3f_1 - 2f_2 = 3*840 - 2*842 = 836$ MHz

$3f_2 - 2f_1 = 3*842 - 2*840 = 846$ MHz

These IMs fall within the pass band and may interfere with the in-band signal if their level is too high. All others are far away from the passband. The spectrum is shown in Figure 4.10. The figure is for illustration purposes and the amplitude of the signals are not to scale.

The third order intermodulation products are most significant compared to all other IMs for an amplifier with weak nonlinearity. The third order intercept point is commonly used for measuring third order intermodulation distortion. The definition of the intercept point is given in appendix B.

The output third order intercept point, IP_3, is defined as the output power at the third order intercept point. Increasing amplifier's linearity causes a higher value for the third order intercept point.

Let P_o and $P_{(2\omega_1-\omega_2)}$ be the linear output power at fundamental frequency, ω_1 or ω_2, and third order IM frequency, $2\omega_1-\omega_2$ or $2\omega_2-\omega_1$, respectively, then, as discussed in appendix B, the third order intercept point is given as:

$$IP_3 = (3P_o - P_{(2\omega_1-\omega_2)}) / 2$$

Since $P_o \approx P_{\omega 1}$ with low level signals, the third order intercept point is determined by measuring the output power at the fundamental frequency, ω_1 or ω_2, and at the third order frequency, $2\omega_1-\omega_2$ or $2\omega_2-\omega_1$, respectively, then calculating according to the equation above. In fact, this is the method by which the third order intercept point is measured in practice.

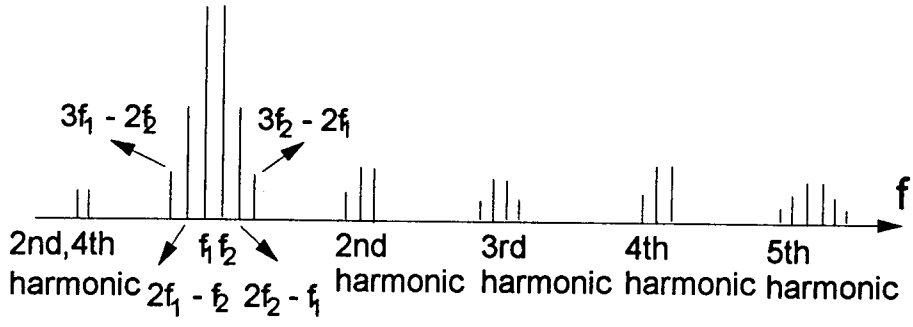

Figure 4.10 Two signal intermodulation spectrum

Another approach to measure the third order IMs is by setting the amplifier output power to a certain level and then measuring the third order IM level. The difference of the two is called third order intermodulation rejection (IMR). The more linear the amplifier the larger the IMR. The equation, $IP_3 = (3P_o - P_{(2\omega_1-\omega_2)})/2$, can be reorganized to:

$$IP_3 = (IMR/2) + P_o$$

With increases in input signal level , the amplifier starts to experience gain compression which causes the IMR to drop. Normally, the IMR is in the range of 45 to 50 dB when IP_3 is measured.

The IP_3 of the overall system with predistortion is obtained by simulation. The simulation method is identical to that used in the single signal test except that the input signal is $X_i = A(\sin \omega_1 t + \sin \omega_2 t)$. ω_1 and ω_2 are selected so that

the third and fifth order IMs are located inside the frequency band. The output of the amplifier is fed into a Fourier transform routine to obtain the output spectrum.

Since the initial contents in the predistortion table are generated by the amplifier's fifth order model, the initial cancellation is not perfect. A simple optimization is used to minimize the IMs. The optimization adjusts the coefficients of the predistorter one by one and monitors the IMs at the output. Since there are few coefficients in the predistorter model, it takes about 15 iterations to obtain 10 dB better performance. This shows that the adaptive technique allows the performance to be improved. An optimization algorithm should be the topic of a further study.

Figure 4.11 shows the two tone test spectrum of the amplifier without predistortion. The output power at the fundamental frequency, ω_1 or ω_2, is 24.88 dBm, the third order intermodulation product at frequency $2\omega_1-\omega_2$ or $2\omega_2-\omega_1$ is -20.78 dBm. therefore, the IP_3 is:

$IP_3 = ((24.88+20.78)/2) + 24.88 = 47.71$ dBm

The actual measurement of IP3 with same input level is 47.8 dBm which is very close to the simulation result.

Figure 4.12 shows the two tone test spectrum of the amplifier with only AM-AM correction. The output power at the fundamental frequency is 24.96 dBm. The third order IM at frequency $2\omega_1-\omega_2$ or $2\omega_2-\omega_1$ is -21.88 dBm. Only 1 dB of additional suppression on IMR is shown.

Figure 4.13 shows the two tone test spectrum of the amplifier with only AM-PM correction. The output power of the fundamental frequency is 24.88 dBm. The third order IM at frequency $2\omega_1-\omega_2$ or $2\omega_2-\omega_1$ is -26.45 dBm. 6 dB of improvement in IMR is achieved.

Figure 4.14 shows the two tone test spectrum of the amplifier with both AM-AM and AM-PM corrections. The output power of the fundamental frequency remains at 24.96 dBm, The third order IM at frequency $2\omega_1-\omega_2$ or $2\omega_2-\omega_1$ is dropped to below -36.91 dBm. 17 dB of improvement in IMR is achieved.

In order to see what improvement the predistortion system can make at higher level, the input signal level is increased to -1 dBm. The results of the simulation are shown in Figure 4.15 and Figure 4.16.

Figure 4.15 shows the spectrum of the amplifier without predistortion. the third order IM level is -0.39 dBm.

Figure 4.16 shows the spectrum of the amplifier with both AM-AM and AM-PM corrections. The third order IM is lowered to -20.41 dBm which is 20 dB better than the amplifier without predistortion.

4.5 Summary

The simulation result shown in section 4.3 and 4.4 indicates that the nonlinear distortion of an amplifier can be reduced by using predistortion correction proposed in Chapter 3.

The linear region of the amplifier is increased because of the distortion cancellation by a predistorter. Therefore, the 1 dB gain compression point, or the output power, of the amplifier system is increased. In the case of PA9340, the output power is increased from 6 Watt to 10 Watt.

The two tone test simulation result demonstrates the IM cancellation of the predistortion. Due to the amplitude and phase distortion nature of the amplifier, only AM-AM or AM-PM correction by themselves does not show good

IM cancellation result. By applying both amplitude and phase correction, the third order IMs in the PA9340 is reduced by 20 dB.

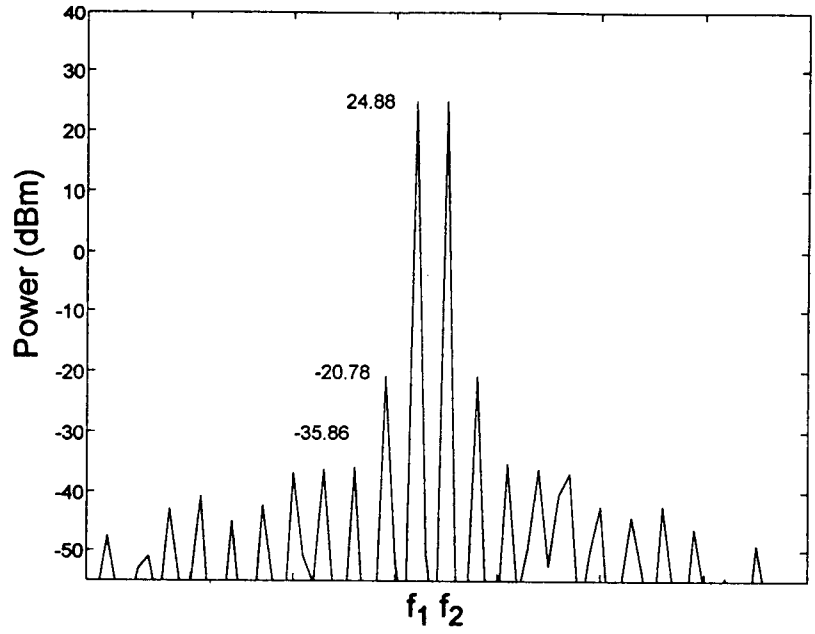

Figure 4.11 Two tone test spectrum of the PA9340 without predistortion (-7 dBm input)

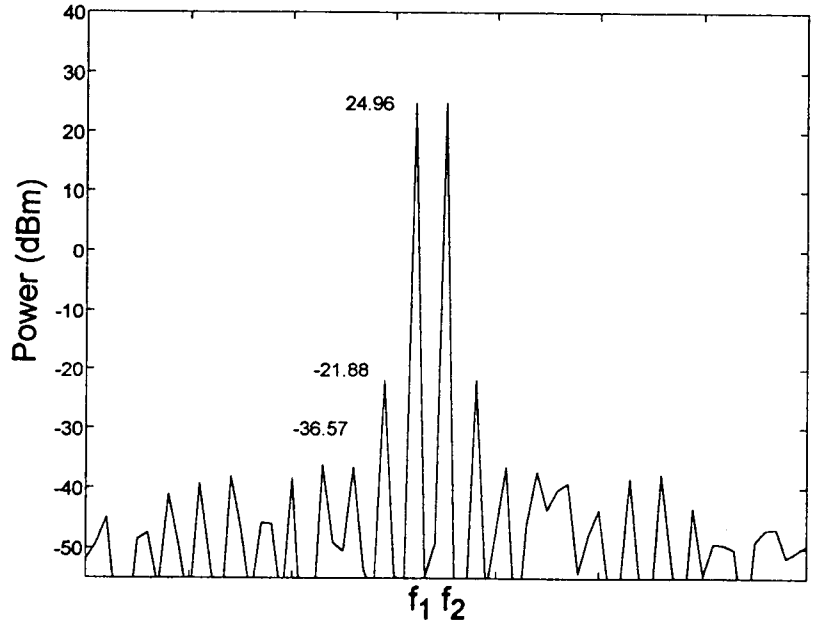

Figure 4.12 Two tone test spectrum of the PA9340 with AM-AM predistortion (-7 dBm input)

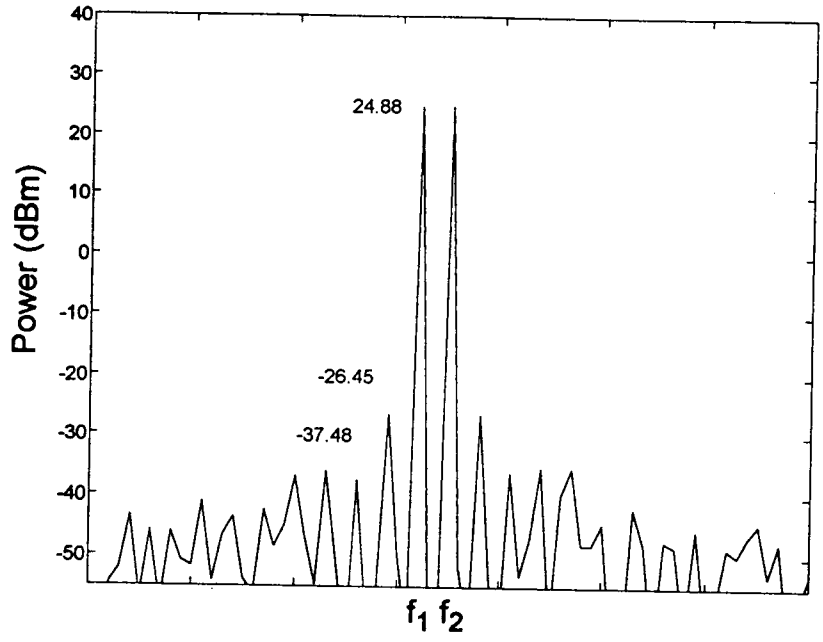

Figure 4.13 Two tone test spectrum of the PA9340
with AM-PM predistortion (-7 dBm input)

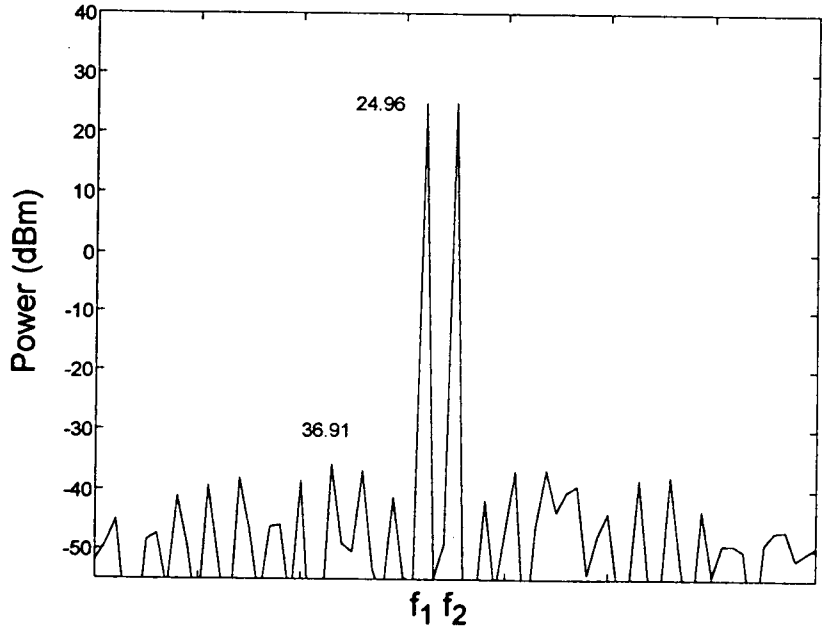

Figure 4.14 Two tone test spectrum of the PA9340
with full predistortion (-7 dBm input)

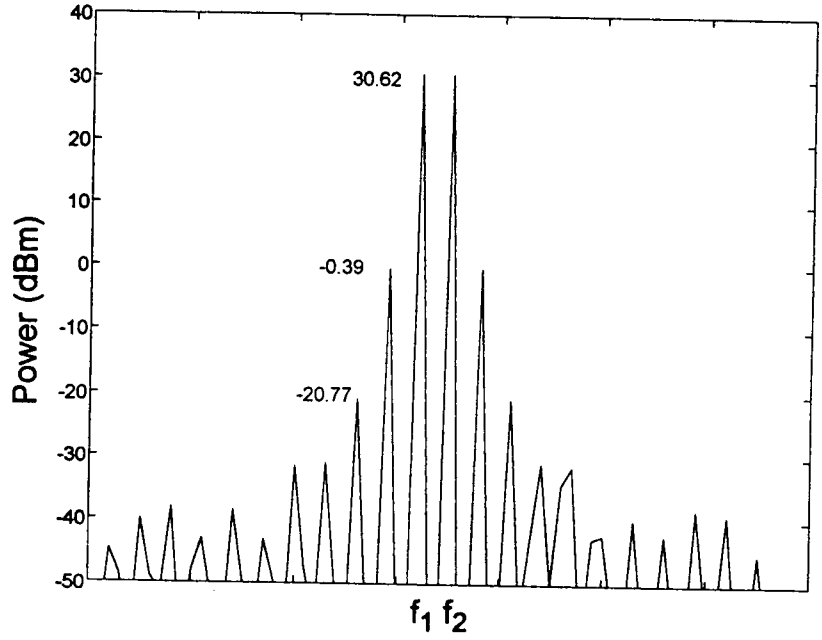

Figure 4.15 Two tone test spectrum of the PA9340 without predistortion (-1 dBm input)

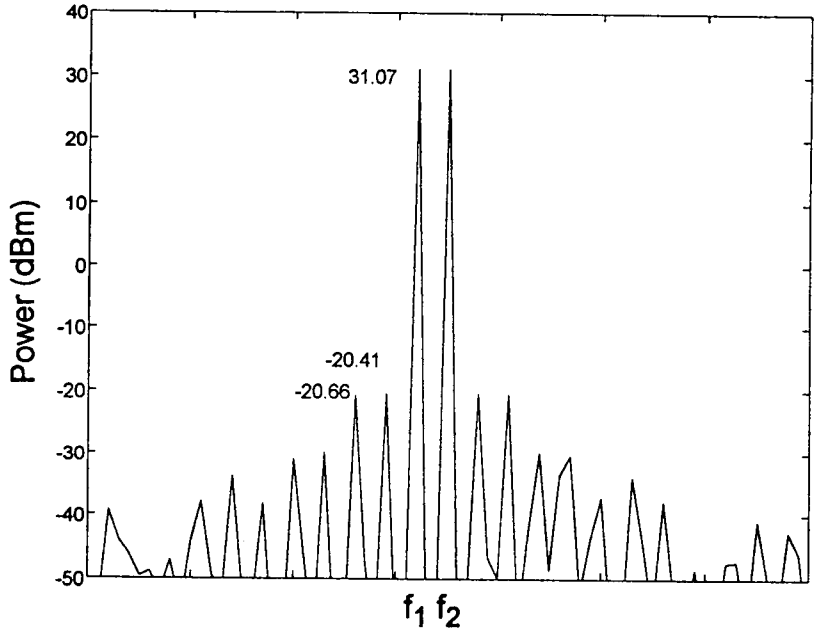

Figure 4.16 Two tone test spectrum of the PA9340 with full predistortion (-1 dBm input)

Chapter 5. System Bandwidth Versus Cost

5.1 Bandwidth limiting factor

It is useful to clarify the meaning of bandwidth. Two types of bandwidths, linearizing and operational bandwidths, are associate with linearization of an amplifier.

The linearizing bandwidth refers to the bandwidth over which the system is capable of performing linearization. This bandwidth is determined by the bandwidth of the components and the architecture of the linearization circuit. For example, in Cartesian feedback, the linearizing bandwidth is determined by the bandwidth of the feedback loop. In the feed-forward and digital predistortion amplifiers, the linearizing bandwidth is defined by the circuit components.

The operational bandwidth refers to the bandwidth in which the amplifier is design to operate. It is the bandwidth within which the linearizing bandwidth can appear. The operational bandwidth for an RF amplifier for cellular communication application is typically several tens of megahertz. The linearizing bandwidth may be equal to or smaller than the operational bandwidth, depending on the linearization techniques used. For example, a typical RF power amplifier has an operational bandwidth of 25 MHz. When the digital predistortion technique with a bandwidth of 30 kHz is applied to that amplifier, the amplifier would have 30 kHz linearizing bandwidth. However, this 30 kHz channel may be tuned to appear anywhere in the 25 MHz spectrum allocation.

In this chapter, the bandwidth is referred to as linearizing bandwidth.

As mentioned in Chapter 2, all the digital predistortion systems reported in the literature are single channel narrow-band systems. The bandwidth of such systems is on the order of tens of kHz. The microprocessors with 10MHz to 30MHz clock rates are fast enough to handle the data processing requirement of the system. The other system components such as A/D, D/A and memory also have no problem to meet the speed requirement of the system.

When the bandwidth increases to the order of several MHz, the processing speed of the components becomes a significant factor.

The linearizing circuit of the predistortion system consists of a modulator/demodulator and a predistorter unit as shown in Figure 3.1, the system block diagram. The feedback path is only for adapting the slow changes of the amplifier nonlinearities. This path impacts the optimization but not the bandwidth. The modulator and demodulator are analog components which can be designed to have a bandwidth in the same order as the amplifier's bandwidth. The predistorter unit consists of A/D, D/A, and digital data processing elements such as adders, registers and memory tables. When an analog signal is converted into a digital signal by A/D, the Nyquist sampling theorem is applied. This, in turn, sets the speed requirement of the components in the predistorter unit.

To achieve a 25 MHz signal bandwidth, the data sample rate needs to be at least 50 MHz. The processing time in the predistorter unit for each data sample has an upper limit of 20 nanoseconds. Referred to the block diagram of the predistorter unit in Figure 3.5, A/D, D/A and the memory table are major components in the predistorter unit and their speed will impact the linearizing bandwidth. The total cost of the linearization circuit should not excess 30 to

40% of the amplifier cost. For example, the reasonable cost on linearizing a $1500 amplifier is about $600. The following section will discuss them in terms of speed, performance and cost.

5.1.1 Analog to digital converter

For a given application, resolution, speed and cost are three important factors which influence the choice of an A/D converter. To some extent, the converter's architecture determines the value for these three parameters. There are many different types of A/D converters available in the market, such as flash, successive approximation, subranging, integrating and sigma-delta converters.

Flash converters offer speeds as fast as 500 Mega samples per second (MSPS), but the resolution is limited to a maximum of 8 to 10 bits. This limitation is a result of its architecture. Figure 5.1 shows the basic architecture of a flash converter. A n-bit flash converter has 2^n-1 comparators. A resistive voltage divider sets the reference voltage for each comparator. The reference voltage at each comparator is one least significant bit (LSB) higher than the comparator immediately below it. When an analog signal is applied to the converter, the comparators having a reference voltage below the level of the input signal give a logic "1" output while the ones having a reference voltage above the input level give a logic "0" output. The outputs then are fed into a encode logic and form a n-bit output word. The conversion is parallel, and all bits are processed at the same time. The number of comparators grows exponentially with the number of resolution bits. An 8-bit flash converter has 255 comparators while a 10-bit converter will have 1023 comparators. It is not

practical to have a flash converter beyond 10 bit of resolution. The flash converter is the fastest and most expensive type of converter. An 8-bit, 500 MSPS flash converter costs around $300.

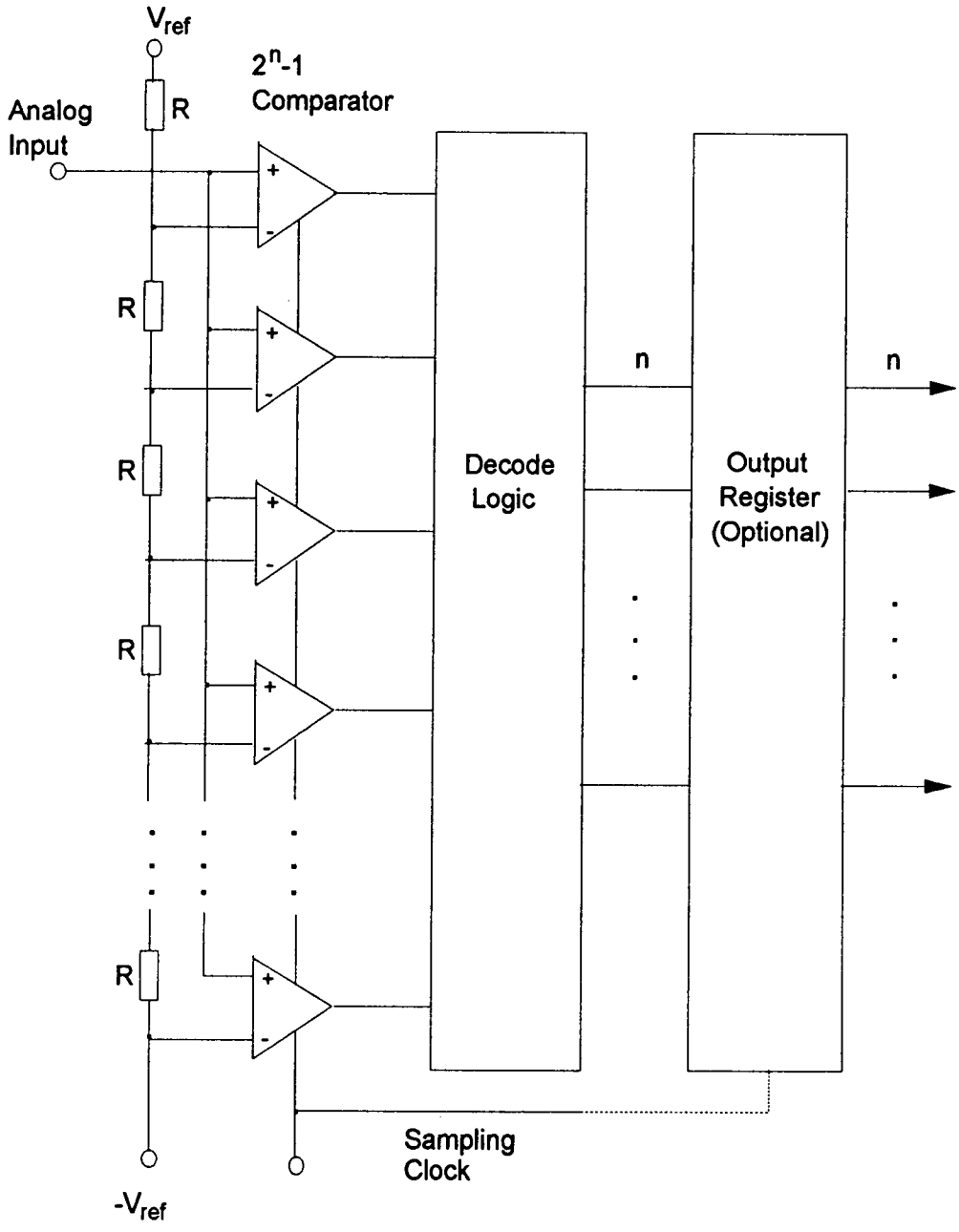

Figure 5.1 Block diagram of a flash converter

A successive approximation converter, as shown in Figure 5.2, uses a comparator, a successive approximation register, a reference D/A converter and control/timing logic to perform conversions. The conversion is sequential in bit-by-bit order. The approximations start with the most significant bit (MSB) and continues through the least significant bit until the output of the D/A is within 0.5 LSB of the input. The decision for each bit requires one clock period. Thus a n-bit conversion requires n-clock period. At the end of the conversion, the contents of the register form a n-bit binary word corresponding to the magnitude of the input signal. Depending on whether the input signal is stable during the conversion period, the converter sometimes requires a sample and hold (S/H) or track and hold (T/H) circuit. The speed and resolution limit is approximately 1 MSPS and 12 bits (25). Higher resolution is possible at slower speeds.

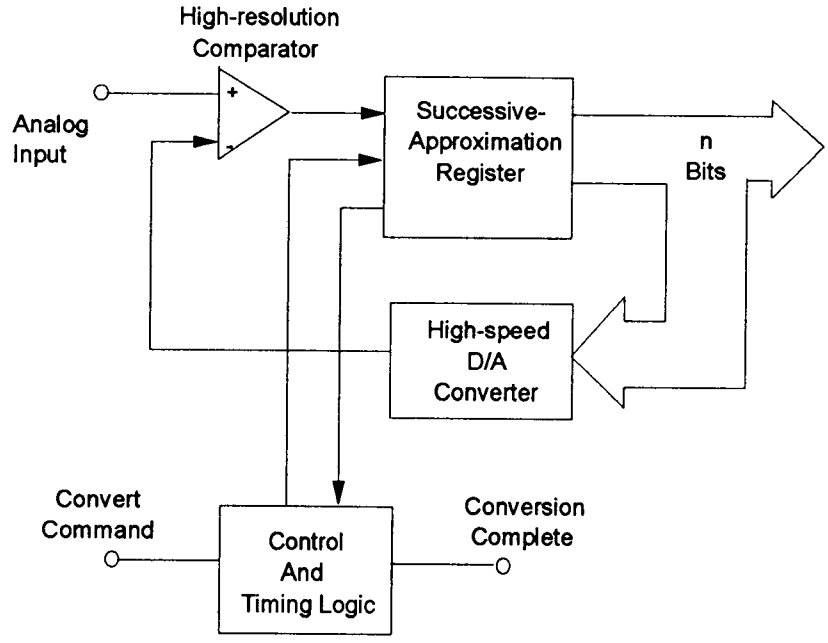

Figure 5.2 Block diagram of a successive-approximation converter

Subranging converters combine the direct and sequential features of flash and successive approximation converters to obtain both high speed and high resolution. This type of converter offers a good compromise in both speed and resolution. The speed can be up to 60 MSPS with a resolution up to 16 bits. The price range is from $30 to $200.

Figure 5.3 shows a two-step subranging architecture with two 4-bit flash converters to produce an 8 bit output. The input signal passes through the first 4 bit flash converter and produces 4 higher order output bits. The summing amplifier then subtracts the D/A output from the input signal and applies the resulting signal to the second flash converter to produce the lower 4 bits. If the signal which goes into the second converter does not fill the range of the converter, the converter can exhibit nonlinearities and missing codes. Most subranging converters include digital error-correction logic to minimize nonlinearities and the possibility of missing codes.

Subranging converters can have different configurations. Some only use a single flash converter like the successive approximation converter, but process the data more than one bit at a time. The more recent ones employ a pipeline structure to speed up the conversion. All subranging converters require a S/H circuit to provide a stable input signal during the conversion process.

Integrating converters and Sigma-delta converters are primarily for low speed applications. A typical conversion speed for integrating converters is 10 sample per second. This is useful for precisely measuring slowly varying signals. Sigma-delta converters have speed in the range of 10 Hz to 100 KHz. These two types are not applicable to the predistorter application because of inadequate speed.

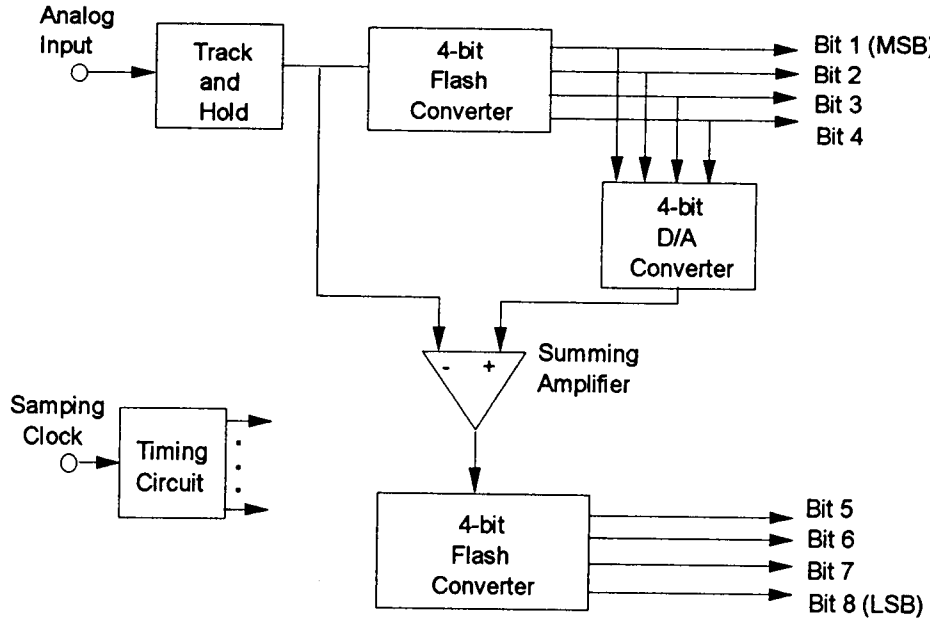

Figure 5.3 Block diagram of a two stage subranging converter

As discussed in Chapter 2, an A/D converter with at least 10-bit resolution and 40 MSPS is required for a 20 MHz bandwidth predistorter. A Flash A/D converter has no problem in meeting the speed requirement, but it is difficult to find a 10-bit resolution flash converter on the market at a reasonable price (less than $100). Subranging converters provide remarkable speed-resolution characteristics, and their performance is close to the requirement of the predistorter unit. Table 5.1 lists some commercially available A/D converters with resolution, speed and cost. Other parameters such as power supply voltage, package, signal to noise ratio are not listed. They can be found in corresponding manufacturer's data books. In general, the cost increases with resolution and speed.

Table 5.1 lists 10-bit and higher resolution A/D converters with MSPS higher than 1 MSPS. A few 8-bit flash A/D converters are given to show the cost and speed compared to subranging devices.

Table 5.1 A/D converters

Part Number	Manufacturer	Resolution Bit	Conversion MSPS	Logic Interface	Price $ /100	Comments
HI1276	HARRIS	8	500	ECL	329.65	Flash
HI1166	HARRIS	8	250	ECL	169.94	Flash
HI1396	HARRIS	8	125	ECL	59.49	Flash
HI1386	HARRIS	8	75	ECL	47.49	Flash
AD9002	Analog Devices	8	125	ECL	90	Flash
AD9012	Analog Devices	8	75	TTL	70	Flash
AD875	Analog Devices	10	15	TTL	15	Pipeline
AD876	Analog Devices	10	20	TTL	9.95/1000	Subranging
AD9040	Analog Devices	10	40	CMOS	50	Subranging
AD9020	Analog Devices	10	60	TTL	165	
AD9060	Analog Devices	10	75	ECL	185	1/2 flash
AD9050	Analog Devices	10	40	TTL	25.5	Subranging
HI5702	HARRIS	10	40	TTL	45.99	
HI5710	HARRIS	10	20	CMOS TTL	16.85	
TDA8760	Philips	10	50	TTL	39.28	Subranging
CXD2310R	Sony	10	20	CMOS TTL	20	
SPT7820	Signal process-- ing tech.	10	20		49	

Table 5.1 A/D converters (Continued)

SPT7855	Signal process-- ing tech.	10	25	CMOS	32	
SPT7860	Signal process-- ing tech.	10	40	CMOS	55	
SPT7920	Signal process-- ing tech.	12	10		122	
SPT7921	Signal process-- ing tech.	12	20		147	
SPT7922	Signal process-- ing tech.	12	30		198	
AD872	Analog Devices	12	15	TTL	140	
AD9022	Analog Devices	12	20	TTL	165	Subranging
AD9023	Analog Devices	12	20	ECL	165	Subranging
AD9026	Analog Devices	12	31	TTL	180	Subranging
AD9027	Analog Devices	12	31	ECL	180	Subranging
AD871	Analog Devices	12	5	TTL	110	Pipeline
HI5800	HARRIS	12	3	CMOS TTL	76.67	Subranging
AD9032	Analog Devices	12	25	ECL	950	Subranging
AD1671	Analog Devices	12	1.25	CMOS	55	Subranging
ADS-130	DATEL	12	10	TTL	775 and up	Subranging
ADS-131	DATEL	12	5	TTL	549 and up	Subranging
ADS-132	DATEL	12	2	TTL	346	Subranging
ADS-193	DATEL	12	1	TTL	289	Subranging
ADS605	Burr-Brown	12	10	TTL	125/1000	Subranging
ADS119	DATEL	12	10		245	
CLC949	Comlinear Corp	12	20		98/1000	
ADS-944	DATEL	14	5		494 and up	
ADS-945	DATEL	14	10		866 and up	

Table 5.1 A/D converters (Continued)

AD878	Analog Devices	14	2.2		75	
MN5430	Micro Networks	14	10		800	

5.1.2 Digital to analog converter

The D/A converters have more selection in terms of desirable speed, resolution and price compared to A/D converters. It is easy to find a D/A converter with 10 to 12 bit resolution, high enough speed and low cost. Table 5.2 lists some D/A converters.

Table 5.2 D/A converters

Part Number	Manufacturer	Resolution Bit	Update MSPS	Logic Interface	Price $ /100	Comments
AD9720	Analog Device	10	400	ECL	42.5	
AD9721	Analog Device	10	100	TTL/CMOS	34	
HI5721	HARRIS	10	100	COMS/TTL	38	
AD9712B	Analog Device	12	100	ECL	29.75	
AD9713B	Analog Device	12	80	TTL/CMOS	29.75	
MAX555	MAXIM	12	250	ECL	68	
AS768	Analog Device	16	32	CMOS	19.95	

5.1.3 Memories

There are two types of memories used in the predistorter unit, RAM and ROM. RAM is used for θ and ρ correction tables. The size is small (512 words each for 10-bit system). Static RAM (SRAM) is a suitable type of RAM for those correction tables. The access time for high speed SRAM is on the order of a few nanoseconds (about 10 nanoseconds) which is enough to meet the need of the predistorter unit. ROM is used for two R/P and P/R conversion tables. The size is relatively large (256k words for 10-bit system). Either EPROM or PROM will meet the system requirement as long as the access time is on the order of 20 nanoseconds or less. Unfortunately, the access time of fast EPROM or PROM with 256k to 1M-bit densities currently available in the commercial market is around 35 nanoseconds to 45 nanoseconds. One way to solve the problem is by using SRAM. There are several companies that make higher speed SRAM with access time on the order of 10 ns to 30 ns. Some hardware has to be added to provide initial data loading to the SRAM.

Table 5.3 lists some memories with size and access times.

Table 5.3 Memories

Part #	Manufacturer	Size (Bit)	Type	Access Time (ns)	Interface	Price ($/100)
CYM1441	Cypress	256k x 8	SRAM	25	TTL	
27HC010-357	ISSI	128k x 8	EPROM	15/35		12.98 (35 ns)
MCM32257	Motorola	256k x 32	SRAM	20/25/35	TTL	
MS621002A	Mosel Vitelic	256k x 4	SRAM	20/25	TTL	

Table 5.3 Memories (Continued)

AM27H	AMD	128k x 8	CMOS EPROM	45	TTL	
CY7C101	Cypress	256k x 4	SRAM	12/15/20 /25/35	TTL	38/35/28/25
CY7C261	Cypress	8k x 8	PROM	20/25/35/45	CMOS	22/19/15/14
CY7C266	Cypress	8k x 8	PROM	20/25/35/45	CMOS	25/21/15/14
CY7C271	Cypress	32k x 8	PROM	30/35/45/55	CMOS	31/23/19/13
CY7C276	Cypress	16k x 16	PROM	25	CMOS	24/20/18
CYM1831	Cypress	16k x 32	SRAM	15/20/25/30 /35/45	TTL/CMOS	84/81/71 /77/78
CYM1836	Cypress	128k x 32	SRAM	20/25/30 /35/45	TTL/CMOS	152/137 /133/131
CYM1840	Cypress	256k x 32	SRAM	25/30/35	TTL/MOS	333/320/320
IS61C256AH	ISSI	256k	SRAM	15		7.5
TC55V328	Toshiba	32k x 8	SRAM	15/20/25		5.5/4.7/4.5
STK11C88	Simtek	256k x 8	SRAM	25/35/45		24.98

5.1.4 Conclusion

A/D converters are the key components in limiting the linearizing bandwidth of the predistortion system. From table 5.1, the speed limit for a 10 to 12 bit A/D converter that is not too expensive is around 50 MSPS. Therefore the maximum analog signal bandwidth is limited to around 25 MHz. Higher speed A/D converters are required for wider bandwidth systems.

The D/A converter, memory and other components are generally able to meet the 20 MHz to 25 MHz requirement of the system.

5.2 Bandwidth versus cost curve

In general, the wider the linearizing bandwidth the higher the system cost. Consider the A/D converter, D/A converter and memory as major components in the predistorter unit. If Nyquist sampling theorem is used to determine the bandwidth, then a bandwidth versus cost curve for a 10-bit predistortion system can be drawn as in Figure 5.4.

Other system components, such as the up-down converters, the microcontroller and the components in the feed back path have less influence on the bandwidth. The estimated cost for those components is on the order of $150. This cost is not shown in Figure 5.4.

Higher bandwidth requires faster components. When the bandwidth increases to 15 MHz, the cost of the memory becomes a major expense in the system. For the two 256k words R/P conversion tables, the cost goes up from $20 for 200ns access time memory to $350 for 16ns access time. The bandwidth is limited by the speed of the A/D converters but the total cost is more strongly influenced by the cost of the memory. Fortunately, the cost of memory devices is constantly decreasing while the performance of A/D converters is improving with the development of new technology. Components with better performance and lower cost are expected in the future.

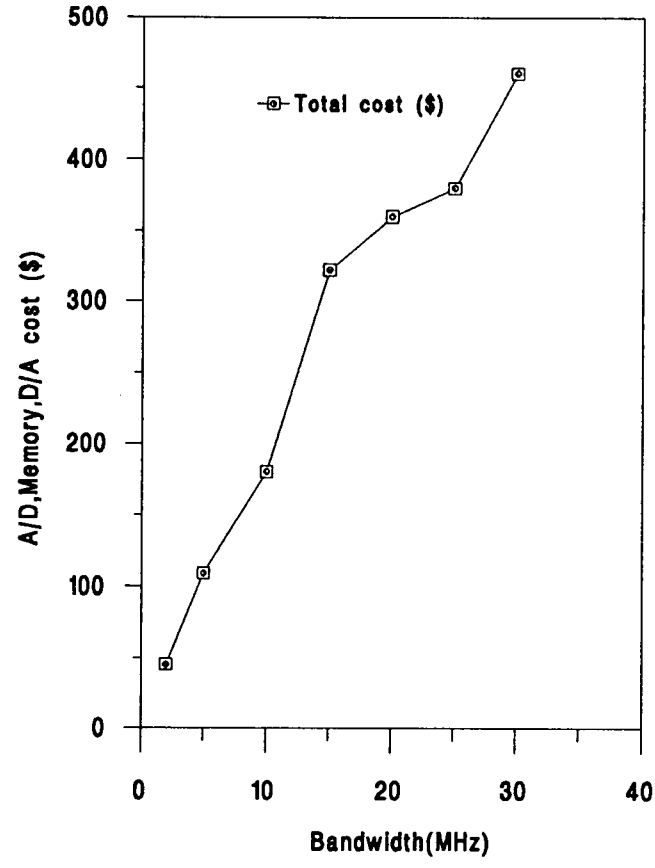

Figure 5.4 Bandwidth versus cost

Chapter 6. Summary and areas of further study

The design features, simulations and component searches presented in this thesis give a good overall analysis and proof of concept necessary to implement the predistorter unit.

The pipelined architecture of the predistorter unit proposed allows fast data processing speed, and is capable of providing linearization for a wideband amplifier. The predistorted correction signal to compensate the nonlinearity of the amplifier is provided with two look-up tables. The contents of the table can be modified according to the characteristic of an amplifier. This allows the unit to be used as a front end building block for different amplifiers.

The simple interface makes the unit easy to be controlled by a general purpose microcontroller or personal computer. The two-set memory look-up table structure allows the up-date to the table be performed without interrupting real time data processing.

The simulation results show significant improvement in the amplifier's performance using the predistortion technique. The linear region of the amplifier is increased and the level of the output IMs is reduced. The improvements are due to the distortion cancellation by the predistorter.

A physical predistorter can be implemented with commercially available components as a first step. The design of the predistorter on ASICs would be another project.

The following areas need further study.

To obtain optimal correction performance, it is found by the simulation that the contents of the correction tables have to be adjusted. The data transfer between the predistorter unit and the host controller should be parallel to

speed up the updates. As mention in Chapter 4, during one update iteration, the simulation monitors the level of the IMs, varies the coefficients (b_i's) of the predistorter one at a time, if the change gives better performance, the program keeps the change and start next iteration. It takes about 15 iterations to achieve 10 dB better IM suppression. This is just to show that the adaptive ability provide in the design allows the performance to be improved. An optimization algorithm should be further developed to ensure the best performance of the system.

As indicated in Chapter 5, the cost of the two rectangular to polar/polar to rectangular conversion tables are a large percentage of the total cost of the system. Cost may be further reduced by considering the symmetry of the I and Q (refer to Figure 3.6) in the first quadrant. It is possible to use the relation of I and Q in the first half of the quadrant to calculate I and Q in the other half, so that the size of these two conversion tables can be reduced by one half. Some hardware is required to implement this. By doing this the cost of the system can be cut significantly.

Bibliography

1. Cavers, James K. "A linearizing predistorter with fast adaptation" Proc of the IEEE Vehic Tech Conf, p. 41- 47 1990.

2. Faulkner, Michael, and Johansson, Mats, "Adaptive Linearsation Using Predistortion-Experimental results" IEEE Transactions on Vehicular Technology V43 no.2 May 1994 p.323-332

3. Johansson, Mats and Thomas Mattsson, "Transmitter Linearization Using Cartesian Feedback for Linear TDMA Modulation" Proc of the IEEE Vehic Tech Conf, p 439- 444 1991

4. Pedro, Jose C. and Jorge Perez, " A Novel Concept of Intermodulation Distortion Linearization " Microwave and Optical Technology Letters Vol 6 No 7 June 5 1993

5. Stapleton, S.P., G. S. Kandola, and J.K.Cavers, "Simulation and Analysis of an Adaptive Perdistorter Utilizing a Complex Spectral Convolution" IEEE Transactions on Vehicular Technology V41 no.4 Nov. 1992 p. 387-394

6. Stapleton, Shawn P., and Flaviu C. Costescu, "An Adaptive Predistorter for a Power Amplifier Based on Adjacent Channel Emissions" IEEE Transactions on Vehicular Technology V41 no.1 Feb.1992 p. 49-56

7. Lerfino, Nick and Harris Farinon, " Increasing Linearity in Amplifiers with IF Predistortion", RF Design March 1994.

8. Nagata, Yoshinori. "Linear Amplification Technique for Digital Mobile Communications", Proc of the IEEE Vehic Tech Conf, p 159-164, 1989.

9. Nojima, Toshio and Tohru Konno, " Cuber Predistortion Linearizer for Relay Equipment in 800 MHz Band Land Mobile Telephone System" IEEE Transactions on Vehicular Technology V34 no.4 Nov 1985 p.169-177

10. Kenington, P.B., and A. Bateman. " RF Linear Amplifier Design", RFexpo West March 1994 San Jose, California

11. Coimbra, Mauro de Lima, and Rui FRagassi de Souza, "Some Improvements on Feedforward Amplifiers", RFexpo West March 1994 San Jose, California

12. Briffa, Mark A. and Michael Faulkner, "Dynamically Biased Cartesian Feedback Linearization", Proc of the IEEE Vehic Tech Conf, p 672-675 1993

13. Larkin, R., and M. Testa, " Cellular Telephone System for Rural Areas" , Electrical Communication, ALCATEL, 1st Quarter 1994

14. Hsieh, Chi-chia and Shu-Park Chan. "A Feedforward S-Band MIC Amplifier System", IEEE Journal of Solid-state Circuits, Vol. SC-11, No. 2, April, 1976.

15. Meyer, R.G., R. Eschenbach, and W.M. Edgerley, JR. "A Wide-Band Feedforward Amplifier", IEEE Journal of Solid-state Circuits, Vol. SC-9, No. 6, December, 1974.

16. Robert S. Larkin " Multiple-Signal Intermodulation and Stability Considerations in the Use of Linear Repeaters" Proc of the IEEE Vehic Tech Conf, pp 891-896 May 1991

17. Lubell, P.D., W.B. Denniston, and R.F. Hertz. "Linearizing Amplifiers for Multi-signal Use", Microwaves. April, 1974.

18. Johnson, A.K., and R. Myer. "Linear Amplifier Combiner", AT&T Bell Laboratories, Whippany, New Jersey 07981.

19. Ha, Tri T. "Solid-State Microwave Amplifier Design", John Wiley & Sons, New York. 1981

20. Freed, J. G. "Equations Provide Accurate Third-order IMD Analysis". Microwaves & RF August 1992.

21. Hewlett Packard "1994 Wireless Communications Symposium".

22. Bowick, Chris. "RF Circuit Design". Howard W. Sams & Company. 1982

23. Lathi, B. P. "Modern Digital and Analog Communication Systems". Holt, Rinehart and Winston, Inc.

24. Sedra, A.S. and Smith, K. C. "Microelectronic Circuits" Harcourt Brace Jovanovich College Publishers. 1991

25. Pryce, Dave "Speed-resolution Tradeoff Key to Choosing ADCs" EDN August 1992

26. Jeruchim, C.M., Balaban P. and Shanmugan S. K. "Simulation of Communication System" 1992 Plenum Press, New York

APPENDICES

Appendix A. Classification of Amplifier

The "class" of an amplifier is defined according to the collector current waveform that results when an input signal is applied (24).

A class-A amplifier is biased at a current I_c, greater than the amplitude of the signal current I_{cs}. Thus, the transistor in a class-A amplifier conducts for the entire cycle of the input signal. That is the conduction angle is 360°. The waveform for the case of sinusoid input signal is shown in figure A (a).

A class-B amplifier is biased at zero dc current. Thus a transistor in a class-B amplifier conducts for only half of the cycle of the input wave, resulting in a conduction angle of 180°. The waveform is shown in figure A (b).

A class-AB amplifier is an intermediate class between class-A and B. It involves biasing the transistor at a nonzero dc current much smaller than the current peak of the sine-wave signal. As a result, the transistor conducts for an interval slightly greater than half of a cycle, as illustrated in figure A (c). The conduction angle is greater than 180° but much less than 360°.

A class-C amplifier is biased such that under steady-state condition no collector current flows. The transistor idles at cutoff. The conduction angle is significantly less than 180°, as shown in figure A (d).

The power conversion efficiency of an amplifier is defined as

η = load power/supply power

For class-A amplifiers, assuming that the dc supply voltage is V, biasing current is I and output voltage is a sinusoid with the peak value of V_p. The average load power will be:

$P_L = (1/2)(V_p^2)/R_L$

where the R_L is a load resistor. The average supply power is given by:

$P_s = V\,I.$

Therefore,

$\eta = P_L / P_s$

$= [(1/2)(V_p^2)/R_L] / V\,I$

Since $V_p \leq (1/2)V$, and $V_p \leq IR_L$, so η is given by:

$\eta = [(1/2)(V_p^2)/ R_L] / V\,I$

$\leq (1/2)(V_p)(1/2)V / V\,I\,R_L$

$= (1/4)(V_p) / I\,R_L$

$\leq 1/4$

The maximum efficiency attainable is 25%. In practice the output voltage is limited to lower values in order to avoid transistor saturation and associated nonlinear distortion. Thus the efficiency achieved is usually lower than 25%.

Class-B amplifiers are more efficient than class-A amplifiers. It can be shown (24) that the maximum achievable efficiency for class-B amplifier is 78.5 %.

The efficiency of class-AB amplifiers is less than class-B amplifiers because of the small biasing power.

Class-C amplifiers can approach 85% efficiency (22). It is the most efficient type of amplifiers

Among all the amplifier types, the class-A amplifier is the most linear one. Linearity is a measurement of how close the output signal of the amplifier resembles the input signal. A linear amplifier is one in which the output signal is proportional to the input signal. A typical class-B amplifier is much less linear, since only half of the input waveform is amplified. The common configuration of a class-B amplifier is the push-pull arrangement which consists of two transistors. Each transistor conducts during half a cycle, one conducting

during the positive half cycle and the other during the negative half cycle. Linearity of class-C amplifier is the poorest of the classes of amplifiers.

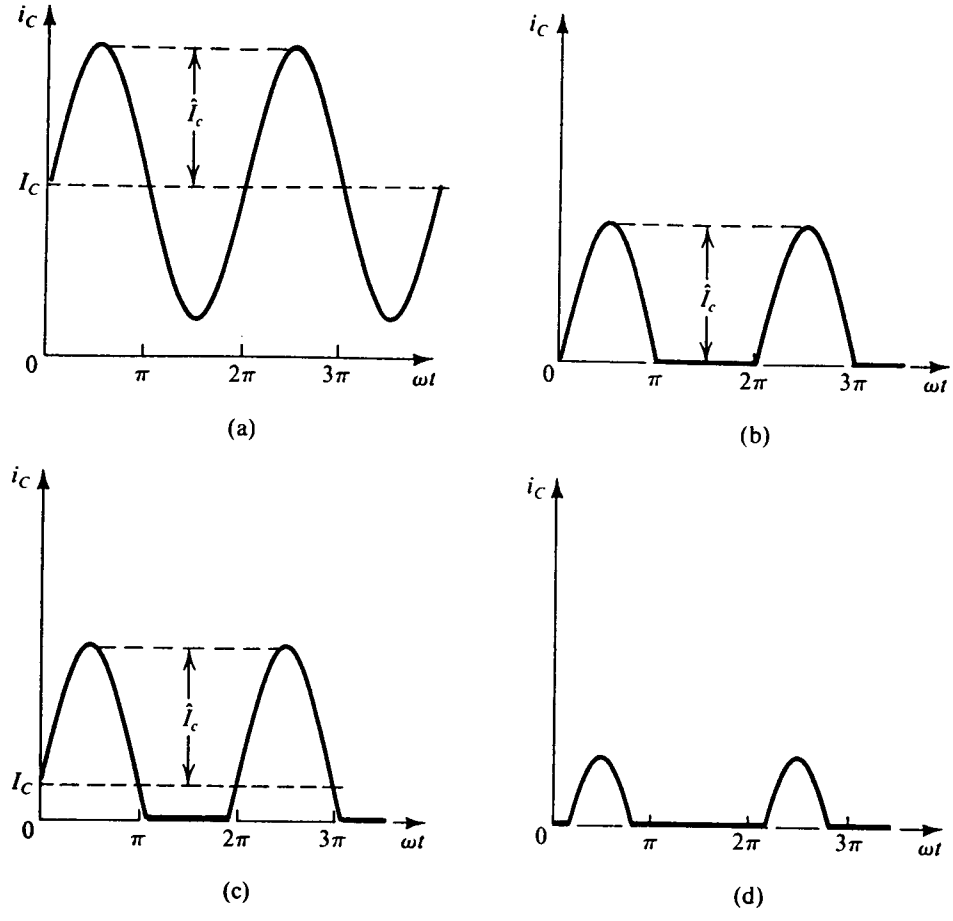

Figure A Collector current waveforms

a. Class A b. Class B

c. Class AB d. Class C

Appendix B. The third order intercept point

Third order intercept point is a commonly used measurement for third order intermodulation distortion.(19)

Consider an input signal $e_i = A(\cos\omega_1 t + \cos\omega_2 t)$. Applying e_i to the fifth order power series yields:

$$
\begin{aligned}
e_o = {} & [C_1 A + (9/4)C_3 A^3 + (50/8)C_5 A^5]\cos\omega_1 t \\
& + [C_1 A + (9/4)C_3 A^3 + (50/8)C_5 A^5]\cos\omega_2 t \\
& + [(3/4)C_3 A^3 + (25/8)C_5 A^5]\cos(2\omega_1 - \omega_2)t \\
& + [(3/4)C_3 A^3 + (25/8)C_5 A^5]\cos(2\omega_2 - \omega_1)t \\
& + [(5/8)C_5 A^5]\cos(3\omega_1 - 2\omega_2)t + [(5/8)C_5 A^5]\cos(3\omega_2 - 2\omega_1)t \\
& + \text{DC term and higher frequency terms}
\end{aligned}
$$

where Cs are the coefficients of the amplifier system.

In cellular systems, the spurious signals at frequencies $2\omega_1$-ω_2, $2\omega_2$-ω_1, and $3\omega_1$-$2\omega_2$ and $3\omega_2$-$2\omega_1$, referred to as the third and fifth order intermodulation products (IMs), will fall within the passband and can interfere with other in-band signals. All other spurious signals are far outside the pass-band and can be filtered out.

The third order intermodulation products are most significant compared to all the other IMs for an amplifier with weak nonlinearity. The third order IMs can be measured by using third order intercept point.

At low signal levels, the output power, P_o, is proportional to the amplitude of the input signal (because $C_1 A >> (9/4)C_3 A^3 + (50/8)C_5 A^5$), and the output of $P_{(2\omega_1-\omega_2)}$ is proportional to the cube of the amplitude of the input signal (because $(3/4)C_3 A^3 >> (25/8)C_5 A^5$). If these two curves are plotted in a log/log scale, the P_o and $P_{(2\omega_1-\omega_2)}$ have slopes of approximately 1 and 3,

respectively. If these two cures are extended, they will intersect at a point called the third order intercept point. Figure B illustrates the third order intercept point.

Intercept point is a useful measure of the intermodulation distortion. The output third order intercept point, IP_3, is defined as the output power at the third order intercept point.

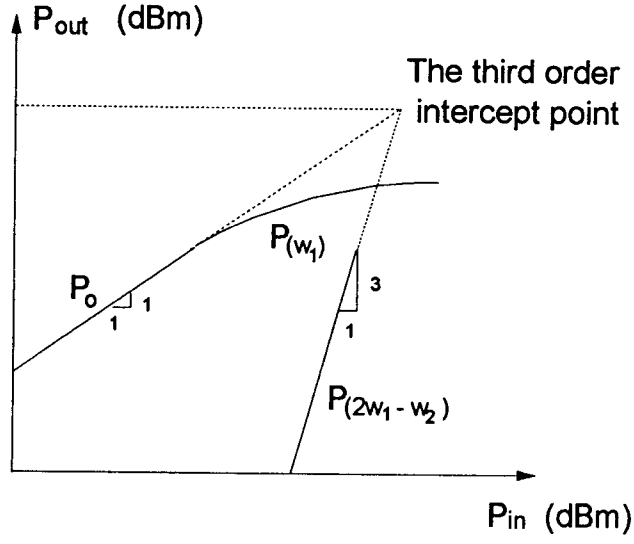

Figure B The third order intercept point

If the amplifier is linear, P_o in dBm in 50Ω system is given by:

$P_o = 10 \log \{(C_1A/\sqrt{2})^2(10^3/50)\}$

with the nonlinearity, the actual output power at fundamental frequency, ω_1 or ω_2, is:

$P_{\omega_1} = 10 \log \{[(C_1A + (9/4)C_3A^3 + (50/8)C_5A^5)/\sqrt{2}\,]^2(10^3/50)\}$

and the output power at frequency, $2\omega_1-\omega_2$ or $2\omega_2-\omega_1$, is:

$P_{(2\omega_1-\omega_2)} = 10 \log \{ [((3/4)C_3A^3 + (25/8)C_5A^5)/\sqrt{2}]^2 (10^3/50) \}$

At lower power levels, the terms contributed from the fifth order coefficient C_5 are rather small and can be ignored, therefore:

$$P_{\omega_1} = 10 \log \{[(C_1A + (9/4)C_3A^3)/\sqrt{2}]^2 (10^3/50)\}$$

$$P_{(2\omega_1-\omega_2)} = 10 \log \{ [((3/4)C_3A^3)/\sqrt{2}]^2 (10^3/50) \}$$

By definition, at intercept point the output power P_o is equal to $P_{(2\omega_1-\omega_2)}$, that is:

$$IP_3 = P_{\omega_1} = P_{(2\omega_1-\omega_2)}$$

The input signal level at intercept point is attained by solving this equation, the result is:

$$A^2 = (4/3)C_1/|C_3|$$

and therefore, the output power at intercept point, IP_3 is given by:

$$IP_3 = 10 \log \{ (2/3) (C_1^3/ |C_3|) (10^3/50) \}$$

The above equation shows that the third order intercept point is independent of the input power, so it is a useful measure of the system's nonlinearity. The more linear amplifiers will have higher third order intercept point.

By comparing P_o,$P_{(2\omega_1-\omega_2)}$ and A^2, a useful relationship is attained:

$$P_{(2\omega_1-\omega_2)} = 3P_o - 2IP_3$$

This equation gives a simple way to measure the third order intercept point as:

$$IP_3 = (3P_o - P_{(2\omega_1-\omega_2)}) / 2$$

Since $P_o \approx P_{\omega 1}$ with low level signals, the third order intercept point is given by measuring the output power at the fundament frequency, ω_1 or ω_2, and at the third order frequency, $2\omega_1-\omega_2$ or $2\omega_2-\omega_1$, respectively, then calculating according to the equation above. In fact, this is the way the third order intercept point is measured in practice.

For example, if the measured $P_{\omega 1}$ = 10 dBm, and $P_{(2\omega_1-\omega_2)}$ = -40 dBm, then

IP_3 = (30-(-40))/2 = 35 dBm

Appendix C. Simulation program

The simulations are designed to show that the predistortion technique proposed in this thesis can reduce the nonlinear distortion of an amplifier. The simulation program is written in MATLAB. The normalized frequency band, 75 to 100, is used to represent the high frequency band. Signals in this frequency range are down converted to the 0 to 25 frequency band for processing, then up converted back to the high frequency band. A sample rate of 1000/second is used to sample the signal. The local oscillation frequency is selected at the lower edge of the band, ω_0=75.

In the single signal test, a input signal with frequency ω_1= 86, is first passed through an I/Q demodulator and low pass filtered to generate the I and Q channel signals. Let the input signal be X_i, the output of the I/Q demodulator be I' and Q', then:

$X_i = A\sin\omega_1 t$

$I' = X_i \cos \omega_0 t$

$= A \sin \omega_1 t \cos \omega_0 t = (A/2) [\sin (\omega_1+\omega_0)t + \sin (\omega_1 -\omega_0)t]$

$Q' = X_i \sin \omega_0 t$

$= A \sin \omega_1 t \sin \omega_0 t = (A/2) [\cos (\omega_1+\omega_0)t - \cos (\omega_1 -\omega_0)t]$

The low pass filtered I and Q are:

$I = (A/2)\sin(\omega_1 -\omega_0)t$

$Q = (A/2)\cos(\omega_1 -\omega_0)t$

The I and Q signals then converted into polar format, ρ and θ using:

$\rho = \sqrt{(I^2 + Q^2)}$

$\theta = \tan^{-1}(Q/I)$

The ρ and θ are used to generate predistorted signal ρ_d, $\Delta\theta$ and θ_d by:

$\rho_d = b_1\rho + b_3\rho^3 + b_5\rho^5$

$\Delta\theta = b'_1\rho + b'_3\rho^3 + b'_5\rho^5$

$\theta_d = \Delta\theta + \theta$

The bi's are the coefficients that discussed in Chapter 4. The ρ, θ and ρ_d, $\Delta\theta$, θ are truncated to 10-bit data.

The predistorted signals then converted back to I and Q format by:

$I_d = \rho_d \cos\theta_d$

$Q_d = \rho_d \sin\theta_d$

I_d and Q_d are then modulated back to X_d to drive the amplifier.

$X_d = I_d \cos\omega_o t + Q_d \sin\omega_o t$

This signal is fed into the amplifier, PA9340, to produce the overall system output. The real measured data of the PA9340 are used in the simulation. Interpolation is used to calculate the data point between measurement points.

The amplitude of the input signal is varied from 0 to the maximum allowable level to generate response of the system. The details are in the program "Onetest.m" .

In the two tone test, the simulation method is identical to that used in the single signal test except that the input signal is $X_i = A(\sin\omega_1 t + \sin\omega_2 t)$, where ω_1 and ω_2 are selected so that the third and fifth order IMs are located inside the frequency band. With ω_1= 86, and ω_1= 89, the third order IMs are located in ω= 83, and ω=92, and the fifth order IMs are located in ω= 80, and ω_1= 95. The output of the amplifier is fed into a Fourier transform routine to obtain the spectrum. The details are in the program "Twotest.m" .

The MATLAB simulation programs are list as follows.

Onetest.m

```
% Single signal test
s=0:0.001:1;
stepsize=0.025;
w0=75;
w1=86;
pwroffs=13.98;
fid=fopen('amplu.txt','r');
amplu=fscanf(fid,'%f %f',[2 inf]);      %amplitude=f(vinrms)
fclose(fid);
fid=fopen('rad.txt','r');
amprad=fscanf(fid,'%f %f',[2 inf]);    %phase=f(vinrms)
fclose(fid);
for j=1:40
b=stepsize*j;
a=b*sqrt(2);
f=a*sin(2*pi*w1*s);
fi=f.*cos(2*pi*w0*s);
fq=f.*sin(2*pi*w0*s);
fil=fi-(a/2)*(sin(2*pi*s*(w1+w0)));
fql=fq+(a/2)*(cos(2*pi*s*(w1+w0)));
if b>0.66
 fu=1*sqrt(2)*sin(2*pi*w1*s);
 fiu=fu.*cos(2*pi*w0*s);
 fqu=fu.*sin(2*pi*w0*s);
 filu=fiu-(1*sqrt(2)/2)*(sin(2*pi*s*(w1+w0)));
 fqlu=fqu+(1*sqrt(2)/2)*(cos(2*pi*s*(w1+w0)));
 pu=2*(1/sqrt(2))*sqrt(filu.*filu+fqlu.*fqlu);
end
p=2*(1/sqrt(2))*sqrt(fil.*fil+fql.*fql);
phase=atan(fql./fil);
for i=1:1001
  if (((fql(i)<0)&(fil(i)<0))|((fql(i)>0)&(fil(i)<0)))
    phase(i)=phase(i)+pi;
  end
end
p=round(p/(0.6/1024))*0.6/1024;    %10 bit, max=0.6
phase=round(phase/(4.712/1024))*4.712/1024; %phase max=4.712
pd=p+0.2715*p.^3+1.773*p.^5;
phdd=0.050*p+1.138*p.^3-.101*p.^5;
if b>0.66
  pd=pu;
end
pd=round(pd/(1.05/1024))*1.05/1024;      %10 bit, max=1.05
phdd=round(phdd/(0.5037/1024))*0.5037/1024; %phase max=0.5037
```

```
phd=phase+phdd;
% run thru an amplifier with input signal
 for k=1:107
  if p(1)<= amplu(1,k),break,end
 end
 if k==107
  amp(1)=amplu(2,107);
  ampph(1)=amprad(2,107);
 end
 if k==1
  sloplu=amplu(2,k)/amplu(1,k);
  amp(1)=p(1)*sloplu;
  slopph=amprad(2,k)/amplu(1,k);
  ampph(1)=p(1)*slopph;
 else
  sloplu=(amplu(2,k)-amplu(2,k-1))/(amplu(1,k)-amplu(1,k-1));
  amp(1)=amplu(2,k-1)+((p(1)-amplu(1,k-1))*sloplu);
  slopph=(amprad(2,k)-amprad(2,k-1))/(amplu(1,k)-amplu(1,k-1));
  ampph(1)=amprad(2,k-1)+((p(1)-amplu(1,k-1))*slopph);
 end
for i=2:1001
  amp(i)=amp(1);
  ampph(i)=ampph(1);
end
oallph=phase+ampph;          %amplifier overall phase
% run thru an amplifier with pd signal
 for k=1:107
  if pd(1)<= amplu(1,k),break,end
 end
 if k==107
  ampd(1)=amplu(2,107);
  ampdph(1)=amprad(2,107);
 end
 if k==1
  sloplupd=amplu(2,k)/amplu(1,k);
  ampd(1)=pd(1)*sloplupd;
  slopphpd=amprad(2,k)/amplu(1,k);
  ampdph(1)=pd(1)*slopphpd;
 else
  sloplupd=(amplu(2,k)-amplu(2,k-1))/(amplu(1,k)-amplu(1,k-1));
  ampd(1)=amplu(2,k-1)+((pd(1)-amplu(1,k-1))*sloplupd);
  slopphpd=(amprad(2,k)-amprad(2,k-1))/(amplu(1,k)-amplu(1,k-1));
  ampdph(1)=amprad(2,k-1)+((p(1)-amplu(1,k-1))*slopphpd);
 end
```

```
for i=2:1001
  ampd(i)=ampd(1);
  ampdph(i)=ampdph(1);
end
oallphd=phd+ampdph;  %amplifier overall phase with pred
fil=sqrt(2)*amp.*cos(oallph);
fql=sqrt(2)*amp.*sin(oallph);
fild=sqrt(2)*ampd.*cos(oallphd);
fqld=sqrt(2)*ampd.*sin(oallphd);
y=f;
outamp=fil.*cos(2*pi*w0*s)+fql.*sin(2*pi*w0*s);
outampd=fild.*cos(2*pi*w0*s)+fqld.*sin(2*pi*w0*s);
fy=fft(y,1000);
pfy=fy.*conj(fy)/1000;
pfydb=10*log10(pfy(1:500))-pwroffs;
pfyv(j)=sqrt(10^(pfydb(87)/10)*0.001*50);
phfy(j)=atan(real(fy(87))/imag(fy(87)));
famp=fft(outamp,1000);
pfamp=famp.*conj(famp)/1000;
pfampdb=(10*log10(pfamp(1:500))-pwroffs);
pfampv(j)=sqrt(10^(pfampdb(87)/10)*0.001*50);
phamp(j)=atan(real(famp(87))/imag(famp(87)));
fampd=fft(outampd,1000);
pfampd=fampd.*conj(fampd)/1000;
pfampddb=(10*log10(pfampd(1:500))-pwroffs);
pfampdv(j)=sqrt(10^(pfampddb(87)/10)*0.001*50);
phampd(j)=atan(real(fampd(87))/imag(fampd(87)));
pwrinf(j)=pfydb(87);
pwrout(j)=pfampdb(87);
pwroutpd(j)=pfampddb(87);
end
plot(pfyv,pfampv,'-',pfyv,pfampdv,'-');
plot(pfyv,phamp,'-',pfyv,phampd,'-');
axis([0 0.625 -0.4 0.25]);
```

Twotest.m

```
% Twotest
s=0:0.001:1;
stepsize=0.025;
w0=75;
w1=86;
w2=89;
pwroffs=13.98;
fid=fopen('amplu.txt','r');
```

```
amplu=fscanf(fid,'%f %f',[2 inf]);      %amplitude=f(vinrms)
fclose(fid);
fid=fopen('rad.txt','r');
amprad=fscanf(fid,'%f %f',[2 inf]);    %phase=f(vinrms)
fclose(fid);
for j=4:4
b=stepsize*j;
a=b*sqrt(2);
f1=a*sin(2*pi*w1*s);
f2=a*sin(2*pi*w2*s);
f=f1+f2;
fi=f.*cos(2*pi*w0*s);
fq=f.*sin(2*pi*w0*s);
fil=fi-(a/2)*(sin(2*pi*s*(w1+w0))+sin(2*pi*s*(w2+w0)));
fql=fq+(a/2)*(cos(2*pi*s*(w1+w0))+cos(2*pi*s*(w2+w0)));
p=2*(1/sqrt(2))*sqrt(fil.*fil+fql.*fql);
phase=atan(fql./fil);
for i=1:1001
  if (((fql(i)<0)&(fil(i)<0))|((fql(i)>0)&(fil(i)<0)))
    phase(i)=phase(i)+pi;
  end
end
p=round(p/(0.6/1024))*0.6/1024;     %10 bit, max=0.6
phase=round(phase/(4.712/1024))*4.712/1024; %phase max=4.712
pd=p+0.251*p.^3+1.705*p.^5;
phdd=0.050*p+1.108*p.^3-.09*p.^5;
pd=round(pd/(1.05/1024))*1.05/1024;       %10 bit, max=1.05
phdd=round(phdd/(0.5037/1024))*0.5037/1024; %phase max=0.5037
phd=phase+phdd;
fildba=sqrt(2)*pd.*cos(phd);
fqldba=sqrt(2)*pd.*sin(phd);
outpdba=fildba.*cos(2*pi*w0*s)+fqldba.*sin(2*pi*w0*s);
% run thru an amplifier with input signal
for n=1:1001
  for k=1:107
   if p(n)<= amplu(1,k),break,end
  end
  if k==107
   amp(n)=amplu(2,107);
   ampph(n)=amprad(2,107);  %should not need this
  end
  if k==1
   sloplu=amplu(2,k)/amplu(1,k);
   amp(n)=p(n)*sloplu;
```

```
   slopph=amprad(2,k)/amplu(1,k);
   ampph(n)=p(n)*slopph;
  else
   sloplu=(amplu(2,k)-amplu(2,k-1))/(amplu(1,k)-amplu(1,k-1));
   amp(n)=amplu(2,k-1)+((p(n)-amplu(1,k-1))*sloplu);
   slopph=(amprad(2,k)-amprad(2,k-1))/(amplu(1,k)-amplu(1,k-1));
   ampph(n)=amprad(2,k-1)+((p(n)-amplu(1,k-1))*slopph);
  end
 end
 oallph=phase+ampph;          %amplifier overall phase
 % run thru an amplifier with pd signal
 for n=1:1001
  for k=1:107
   if pd(n)<= amplu(1,k),break,end
  end
  if k==107
   ampd(n)=amplu(2,107);
   ampdph(n)=amprad(2,107);  %should not need this
  end
  if k==1
   sloplupd=amplu(2,k)/amplu(1,k);
   ampd(n)=pd(n)*sloplupd;
   slopphpd=amprad(2,k)/amplu(1,k);
   ampdph(n)=pd(n)*slopphpd;
  else
   sloplupd=(amplu(2,k)-amplu(2,k-1))/(amplu(1,k)-amplu(1,k-1));
   ampd(n)=amplu(2,k-1)+((pd(n)-amplu(1,k-1))*sloplupd);
   slopphpd=(amprad(2,k)-amprad(2,k-1))/(amplu(1,k)-amplu(1,k-1));
   ampdph(n)=amprad(2,k-1)+((pd(n)-amplu(1,k-1))*slopphpd);
  end
 end
 oallphd=phd+ampdph;  %amplifier overall phase with pred
 fil1=sqrt(2)*amp.*cos(oallph);
 fql1=sqrt(2)*amp.*sin(oallph);
 fild=sqrt(2)*ampd.*cos(oallphd);
 fqld=sqrt(2)*ampd.*sin(oallphd);
 y=f;
 outamp=fil1.*cos(2*pi*w0*s)+fql1.*sin(2*pi*w0*s);
 outampd=fild.*cos(2*pi*w0*s)+fqld.*sin(2*pi*w0*s);
 fy=fft(y,1000);
 pfy=fy.*conj(fy)/1000;
 pfydb=10*log10(pfy(1:500))-pwroffs;
 pfyv(j)=sqrt(10^(pfydb(87)/10)*0.001*50);
 phfy(j)=atan(real(fy(87))/imag(fy(87)));
```

```
famp=fft(outamp,1000);
pfamp=famp.*conj(famp)/1000;
pfampdb=(10*log10(pfamp(1:500))-pwroffs);
pfampv(j)=sqrt(10^(pfampdb(87)/10)*0.001*50);
phamp(j)=atan(real(famp(87))/imag(famp(87)));
for i=1:500
  if (pfampdb(i)<(pfampdb(87)-90))
   pfampdb(i)=(pfampdb(87)-90);
  end
end
fampd=fft(outampd,1000);
pfampd=fampd.*conj(fampd)/1000;
pfampddb=(10*log10(pfampd(1:500))-pwroffs);
pfampdv(j)=sqrt(10^(pfampddb(87)/10)*0.001*50);
phampd(j)=atan(real(fampd(87))/imag(fampd(87)));
for i=1:500
  if (pfampddb(i)<(pfampddb(87)-90))
   pfampddb(i)=(pfampddb(87)-90);
  end
end
pwrinf(j)=pfydb(87);
pwrout(j)=pfampdb(87);
pwroutpd(j)=pfampddb(87);
amp1(j)=amp(87);
ampph1(j)=ampph(87);
oallph1(j)=oallph(87);
ampd1(j)=ampd(87);
ampdph1(j)=ampdph(87);
oallphd1(j)=oallphd(87);
end
plot(pfyv,pfampv,'+',pfyv,pfampdv,'*');
plot(pfyv,phamp,'+',pfyv,phampd,'*');
plot(pfampdb(1:500));
axis([60 120 -55 40]); %rescale the plot
```