UNIVERSITY OF CALIFORNIA, SAN DIEGO

High-Efficiency and High-Linearity SiGe BiCMOS Power Amplifiers for WCDMA Handset Applications

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Electrical Engineering (Electronic Circuits and Systems)

by

Junxiong Deng

Committee in charge:

Lawrence E. Larson, Chair
Peter M. Asbeck
Paul Yu
Chung-Kuan Cheng
KL Paul Sung
Prasad S. Gudem

2005
For my wife, parents, sister and brother-in-law

給我的妻子，父親，母親，姐姐和姐夫
# Table of Contents

SIGNATURE PAGE..........................................................................................................III

DEDICATION ..................................................................................................................IV

TABLE OF CONTENTS ..................................................................................................... V

LIST OF FIGURES ........................................................................................................ VIII

LIST OF TABLES ......................................................................................................... XIII

ACKNOWLEDGEMENTS ...............................................................................................XIV

VITA.............................................................................................................................XVI

ABSTRACT ................................................................................................................ XVIII

CHAPTER 1 INTRODUCTION ....................................................................................... 1

  1.1 Power Amplifiers for WCDMA Handsets ......................................................... 4
    1.1.1 WCDMA System ..................................................................................... 4
    1.1.2 WCDMA Handset Transceiver ............................................................. 8
    1.1.3 WCDMA Power Amplifier Specifications ............................................ 10
    1.1.4 Survey of Reported CDMA/WCDMA Handset Power Amplifiers ...... 11
  1.2 Power Amplifier Classification ........................................................................ 12
  1.3 Power Amplifier Technologies......................................................................... 19
  1.4 Research Focus................................................................................................. 23
  1.5 Dissertation Organization ................................................................................. 25

CHAPTER 2 EFFICIENCY IMPROVEMENT BY DYNAMIC BIASING TECHNIQUES .... 27

  2.1 Survey of PA Efficiency Boosting Techniques ............................................... 29
    2.1.1 Constant Current/Voltage Biasing ......................................................... 29
    2.1.2 Doherty Amplifier ................................................................................ 32
    2.1.3 Envelope Elimination and Restoration ................................................. 34
    2.1.4 Envelope Tracking .............................................................................. 36
    2.1.5 LINC Amplifier ................................................................................... 37
List of Figures

Figure 1.1 Mobile handset production in the world from 1998 to 2006 [3]...................2
Figure 1.2 Distribution diagram of handset market in 2004 [3].....................................3
Figure 1.3 Evolution diagram of mobile systems [6].....................................................5
Figure 1.4 Operating band diagram in WCDMA systems [8]........................................6
Figure 1.5 Principle of spread spectrum technique. (a) Frequency division duplex. (b) The wideband signal after the spreading operation. (c) Despreading for user 1........7
Figure 1.6 ACPR spectrum definition of WCDMA systems........................................8
Figure 1.7 Simplified schematic of a typical WCDMA handset transceiver. ...............9
Figure 1.8 Simplified schematic of a bipolar power amplifier operating in Class A/B/C ............................................................................................................................14
Figure 1.9 (a) Class D power amplifier circuit schematic. (b) Ideal voltage and current waveforms. ...................................................................................................................15
Figure 1.10 (a) Class E power amplifier circuit. (b) Ideal voltage and current waveforms. ............................................................................................................................16
Figure 1.11 (a) Class F power amplifier circuit. (b) Ideal voltage and current waveforms. ............................................................................................................................16
Figure 1.12 Output voltage and current waveforms at the collector of power amplifiers. (a) Class AB. (b) Class A. .................................................................................18
Figure 1.13 Schematic cross-sections of different devices. (a) GaAs NPN HBT. (b) Si NPN BJT. (c) SiGe NPN HBT. (d) NMOS. (e) LDMOS. ...........................................21
Figure 2.1 Representative CDMA probability distribution function (PDF) [46] and dc currents of a Class AB PA and a TxIC versus output power [47]...............................28
Figure 2.2 Schematic of constant current biasing (CCB)..............................................30
Figure 2.3 Large signal characteristic of CCB..............................................................30
Figure 2.4 Schematic of constant voltage biasing (CVB). .............................................31
Figure 2.5 Large signal characteristic of CVB.............................................................31
Figure 2.6 Block diagram of Doherty amplifier............................................................32
Figure 2.7 Power transfer characteristic of Doherty amplifier....................................33
Figure 2.8 Load line diagram of Doherty amplifier operation ........................................33
Figure 2.9 EER block diagram ..........................................................................................35
Figure 2.10 ET block diagram .........................................................................................36
Figure 2.11 LINC block diagram ....................................................................................38
Figure 2.12 Efficiency as a function of output amplitude [1]. A: Class A; B: Class B; ET: Envelope Tracking; DOH: Doherty amplifier; CHIR: Chireix outphasing amplifier (i.e. LINC); KAHN: Kahn amplifier (i.e. EER). ..........................................................38
Figure 2.13 BJT dc characteristics showing dynamic biasing techniques. ...................39
Figure 2.14 QPSK signal constellation diagram. .............................................................40
Figure 2.15 Simplified schematic of a typical SiGe HBT power amplifier. ....................42
Figure 2.16 Small signal model of HBT transistors. ......................................................42
Figure 2.17 Output stage transistors with dynamic current biasing. HBTs are biased “on” or “off” in response to output power requirements..............................................44
Figure 2.18 Simulated HBT cutoff frequency versus collector current with differing device sizes (single device: 25 µm²; m represents the number of devices in parallel). 45
Figure 2.19 Simplified schematic of a two-step switched dual dynamic bias in the output stage................................................................................................................... 46
Figure 2.20 Equivalent circuit including NFET switches .................................................50
Figure 2.21 Effect of NFET device size on power gain...................................................51
Figure 2.22 Effect of NFET device size on output 1dB compression point.................52
Figure 2.23 NFET large signal characteristics [61].......................................................53
Figure 2.24 NFET switch operation.................................................................................53
Figure 2.25 Schematic of power amplifier with bias network. .......................................55
Figure 2.26 Output matching network schematic..........................................................56
Figure 2.27 3-stage output matching network..................................................................57
Figure 2.28 Final schematic of the power amplifier with SDCB....................................59
Figure 2.29 Emitter ballasting resistor configuration......................................................62
Figure 2.30 Simplified schematic of the driver stage including the inter-stage matching (CVB: constant voltage biasing). .................................................................63
Figure 2.31 Inter-stage matching network......................................................................64
Figure 2.32 Inter-stage matching terminations. (a) Internal termination. (b) External termination............................................................................................................................................ 65
Figure 2.33 Final schematic of the two-stage PA with SDDB................................. 66
Figure 2.34 Cross-sections of oppces and opperes resistors.................................. 69
Figure 2.35 Cross-sections of MOSCAP33 and MIM capacitors. ....................... 70
Figure 2.36 Cross-section of AM octagonal inductors with the PC groundplane...... 71
Figure 2.37 Layout topology of single transistor....................................................... 72
Figure 2.38 Two different layout topologies of 100 transistors. (a) N=1 and M=100.
(b) N=5 and M=20........................................................................................................ 73
Figure 2.39 Gate tie down diagram............................................................................. 75
Figure 2.40 Bipolar-based diode ESD structure. (a) typical configuration. (b) stacked
configuration................................................................................................................. 76
Figure 2.41 Layout of single-stage PA with SDCB................................................... 77
Figure 2.42 Layout of two-stage PA with SDDB....................................................... 77
Figure 3.1 Frequency domain model of Volterra series............................................ 84
Figure 3.2 Flow diagram of PDCV analysis.............................................................. 86
Figure 3.3 Equivalent circuit of the SiGe HBT power amplifier.............................. 88
Figure 3.4 Nonlinear circuit of SiGe HBT for Volterra series calculation............... 88
Figure 3.5 Simulation data of $C_\pi$ and its third-order and fifth-order fitting curve..... 90
Figure 3.6 Nonlinearity coefficients $C_l$ and $K_3C_\pi$. $I_{CQ} = 110\text{mA}$............... 91
Figure 3.7 Simulation data of $i_C$ and its third-order curve..................................... 93
Figure 3.8 Nonlinearity coefficients $g_m$ and $K_3g_m$. $I_{CQ} = 110\text{mA}$........... 93
Figure 3.9 Block diagram of IMR3 calculation......................................................... 95
Figure 3.10 First-order equivalent circuit of SiGe HBT power amplifier................. 96
Figure 3.11 Equivalent circuit for the computation of the third-order intermodulation
product $2\omega_1 - \omega_2$................................................................................................. 97
Figure 3.12 Test bench used for experimental verification of linearity analysis........ 99
Figure 3.13 IMR3 comparison between SPECTRE simulation, Volterra calculation,
and measurement for the circuit of Figure 3.12...................................................... 100
Figure 3.14 Individual contributions to amplitude and phase of output IMR3........... 101
Figure 3.15 IMR3 with different values of $L_e$. $I_{CQ} = 110\text{mA}$.............................103
Figure 3.16 Comparison of simulated and calculated power gains with different values of $L_e$. ...........................................................................................................................103
Figure 3.17 Simulated power added efficiencies at 26 dBm output power with different values of $L_e$. .................................................................104
Figure 3.18 Simplified schematic of the dynamically-biased RF amplifier.........105
Figure 3.19 Comparison between calculation and simulation of IMD asymmetry with varying envelope injection phase. .................................................................109
Figure 3.20 Vector diagram showing IMD3 asymmetry. ..................................110
Figure 3.21 Vector diagram showing optimal IMD3 cancellation. .....................111
Figure 3.22 IMD3 reduction versus input power in the circuit of Figure 3.18. ....112
Figure 4.1 Simplified schematic of RF feedback amplifiers.................................117
Figure 4.2 Schematic of Cartesian modulation feedback.....................................118
Figure 4.3 Schematic of polar modulation feedback.............................................120
Figure 4.4 Schematic of feedforward system.......................................................122
Figure 4.5 Schematic of predistortion systems. (a) general form. (b) compensation principle..........................................................123
Figure 4.6 Schematic of cubic predistorters.........................................................124
Figure 4.7 Parallel diode predistortion schematic and characteristics. ..............125
Figure 4.8 Block diagram of digital predistortion..................................................127
Figure 4.9 Different digital predistortion algorithms [113].................................129
Figure 4.10 Look-Up-Table (LUT) size comparison between mapping and polar predistortion..........................................................130
Figure 4.11 Conceptual implementation diagram of complex-gain predistortion [121]. ........................................................................................................132
Figure 4.12 Predistortion magnitude rescaling.....................................................134
Figure 4.13 Test setup of the digital predistortion system.....................................135
Figure 4.14 Comparison of measured ACPRs with and without the digital predistortion..........................................................136
Figure 4.15 Input/output behavior without digital predistortion – measured data (dots), ideal performance (dashed line), curvefit (light line) ................................................. 137
Figure 4.16 Input/output behavior with digital predistortion – measured data (dots), ideal performance (dashed line), curvefit (light line) ................................................. 137
Figure 5.1 A typical microwave probe system ........................................................... 140
Figure 5.2 Cross section of the Micro Lead Frame package ....................................... 142
Figure 5.3 Photograph of the packaged die of the prototype power amplifier with SDDB. .................................................................................................................. 142
Figure 5.4 Photograph of the test PCB board with the package .................................. 143
Figure 5.5 Test bench Setup ...................................................................................... 144
Figure 5.6 Output power probability distribution $P_e$ and measured DC current for different biasing techniques. The switch point from 100 devices to 20 devices occurs at $P_{out} = 18$ dBm. $V_{CC}=3$V. ............................................................... 145
Figure 5.7 Measured power added efficiencies with constant voltage with fixed area (CV) and switched dynamic current biasing with varied area (SDCB) power amplifiers ........................................................................................................... 146
Figure 5.8 Measured power gains with constant voltage with constant area (CV), SDCB with varied HBT area and DCB with fixed HBT area power amplifiers...... 147
Figure 5.9 Measured ACPRs of SDCB SiGe HBT power amplifier. ......................... 148
Figure 5.10 Output power probability distribution function $P_e$ and measured DC current comparison for different biasing techniques (CV: constant voltage biasing; SDDDB: switched dual dynamic biasing). The switching point from high-power mode to low-power mode is $16$ dBm. .................................................................................. 149
Figure 5.11 Measured ACPRs of the SDDDB power amplifier with digital predistortion (DP) (before DP and after DP). ................................................................. 150
Figure 5.12 Measured gain and power added efficiency (PAE) of the SDDDB power amplifier with DP. ................................................................................................. 151
List of Tables

Table 1.1 Transmit power classes in different WCDMA handsets [10]. ..................... 10
Table 1.2 3GPP WCDMA handset power amplifier specifications. ......................... 11
Table 1.3 Comparison of recently reported CDMA/WCDMA handset power
amplifiers. ...................................................................................................................... 12
Table 1.4 Comparison of peak possible efficiency for different operation classes of
power amplifiers. ........................................................................................................... 14
Table 1.5 Comparison of typical technologies for RF power amplifiers [40].............. 20
Table 2.1 Average power efficiency comparison of SDCB in an output stage......... 48
Table 2.2 Average power efficiency comparison of SDDB in an output stage......... 60
Table 2.3 Characteristics of metal levels in IBM SiGe BiCMOS 6HP technology.... 68
Table 2.4 Resistor comparison in IBM SiGe BiCMOS 6HP technology............... 69
Table 2.5 Parameter comparison of NPN transistors. ............................................. 72
Table 4.1 Comparison of different PA linearization techniques. ......................... 115
Table 5.1 Average power efficiency comparison of reported dynamic biasing
techniques. ..................................................................................................................... 146
Acknowledgements

I would like to express my sincere thankfulness to my research advisors: Professor Lawrence Larson and Dr. Prasad Gudem. Professor Lawrence Larson always guides me with his lightening advice and constructive suggestions, which are indispensable to this dissertation. His continuous support and encouragement make this work come true. Dr. Prasad Gudem, at all times, offers his unselfish help and incisive comments to my research. Without him, this long journey will not be so unforgettable and enjoyable.

I would also like to thank the members of my committee, Professors Peter Asbeck, Paul Yu, Chung-Kuan Cheng, and Paul Sung, for their invaluable suggestions and recommendation to this dissertation.

I wish to acknowledge Mr. Joe Pusl of Magis Networks for helpful comments on the SiGe power amplifier design; Dr. Xiaojun Yuan of IBM for his timely support during the chip implementation; Dr. Donald Lie of SPAWAR for many inspiring conversations. Many thanks would go to my colleagues at UCSD for their inspiring discussions and suggestions. I would especially like to acknowledge: Dr. Chengzhou Wang (now with Marvell), Dr. Liwei Sheng (now with Koycera), Dr. Vincent Leung (now with IBM), Dr. John Fairbanks (now with L3 Communications), Mr. Don Kimball, Mr. Mingyuan Li, Mr. Farzad Parvaresh, Mr. Rahul Kodkani, Mr. Joe Jamp, Mr. Yang Sun, Mr. Yiping Han, and Mr. Himanshu Khatri. I also want to gratefully
acknowledge the continuous support of the Center of Wireless Communications at UCSD, and its member companies and a University of California Discovery Grant.

The text of Chapter Two, Three, Four and Five, in part or in full, is a reprint of the material as it appears in our published papers or as it has been submitted for publication in *IEEE Transactions on Microwave Theory and Techniques, IEEE Radio Frequency Integrated Circuits Symposium, and IEEE Radio and Wireless Conference*. The dissertation author was the primary researcher and the first author listed in these publications. He directed and supervised the research which forms the basis for these chapters.
Vita

1997 B.S. EE, Southeast University, Nanjing, China
2003 M.S. EE, University of California, San Diego
2005 Ph.D. EE, University of California, San Diego
Dissertation: *High-Efficiency and High-Linearity SiGe BiCMOS Power Amplifiers for WCDMA Handset Applications*

Publications


**Fields of Study**

Major Field: Electrical Engineering

   Professor Lawrence E. Larson and Dr. Prasad S. Gudem
ABSTRACT OF THE DISSERTATION

High-Efficiency and High-Linearity SiGe BiCMOS Power Amplifiers
for WCDMA Handset Applications

by

Junxiong Deng

Doctor of Philosophy in Electrical Engineering (Electronic Circuits and Systems)
University of California, San Diego, 2005

Professor Lawrence E. Larson, Chair

In recent years, silicon germanium (SiGe) has become a competitive candidate for the development of cellular handset power amplifiers (PAs) of third-generation (3G) wireless communication systems, since SiGe exhibits good linearity, low-cost and compatibility with BiCMOS technology. Wideband Code Division Multiple
Access (WCDMA), one of 3G cellular system standards, promises wideband services and high data rate applications, such as video teleconferencing and web browsing. This poses stringent requirements on efficiency and linearity of RF handset power amplifiers (PAs).

This dissertation proposes switched dynamic biasing techniques to boost the average power efficiency of WCDMA power amplifiers. Furthermore, the gain-variation issue in the existing dynamic biasing techniques is overcome by keeping the collector current density constant, which further helps with the error vector magnitude (EVM) of WCDMA handsets. Two prototype WCDMA handset PAs have been implemented in a 0.25\(\mu\)m SiGe BiCMOS process to demonstrate the effectiveness of the proposed techniques.

The linearity of SiGe HBT power amplifiers has been analyzed with power-dependent coefficient Volterra (PDCV) analysis. A SiGe HBT power amplifier is analyzed to identify the dominant sources of nonlinearity and provide directions to design linear SiGe HBT high-power amplifiers. A dynamically-biased RF amplifier with envelope signal injection is also analyzed to clarify the cause of third-order intermodulation product (IMD\(_3\)) asymmetry. Furthermore, the linearity of the amplifier can be improved by optimizing the phase and amplitude of the injected envelope signal. Both analyses are confirmed with simulation and experimental results.
Digital predistortion is employed to improve the linearity of WCDMA power amplifiers, since it can be conveniently integrated, flexibly controlled, and economically implemented without sacrificing other performance, such as power gain and efficiency. The operation principle of digital predistortion is derived rigorously. The effectiveness of digital predistortion is verified through experimental results.
CHAPTER 1
Introduction

Wireless communications has evolved for over one century. As far back as 1896, Guglielmo Marconi demonstrated wireless telegraphy to the English telegraph office. His pioneering work in this field resulted in his award of the Nobel Prize for physics in 1909. However, the modern revolution in telecommunications started with the first digital cellular phone system in the United States in 1991. Since then, the AMPS (Advanced Mobile Phone System), CDMA (Code Division Multiple Access), and GSM (Global System for Mobile Communications) wireless cellular systems have all witnessed exponential growth. To accommodate the increasing requirements for wideband services and high data rate applications in wireless communication systems, third-generation (3G) cellular system standards, including CDMA2000 [1] and WCDMA (Wideband CDMA) [2], have been proposed and implemented into commercial versions. 3G wireless systems promise a new array of mobile communication services, such as video teleconferencing and web browsing. There is no doubt that this telecommunication revolution has affected every aspect of our lives profoundly.
The key to successful 3G systems is the implementation of low-cost and high-power-efficiency handsets. The cost factor determines the number of subscribers and consequently the deployment scope of 3G systems. On the other hand, a high efficiency handset supports a longer battery life and more complicated and power-consuming applications, which make 3G systems more attractive. For any type of wireless handset, a transmitter is always necessary, and for any type of transmitter, an RF power amplifier is an essential component. As shown in Figure 1.1 [3], the number of mobile handset power amplifiers is projected to keep increasing. The major portion of this market is generated by replaced handsets and rapid expansion of new subscribers coming from low-income regions such as China and India.

Figure 1.1 Mobile handset production in the world from 1998 to 2006 [3].
The rapid development of 3G systems results in more stringent requirements on efficiency and linearity of RF power amplifiers. As shown in Figure 1.2 [3], there are 50% non-entry-level handset power amplifiers that have to meet these rigorous requirements and their percentage is expected to expand in the future. In order to obtain high efficiency handsets, the power amplifier has to be power-efficient since it is the major power consumer and dominates the talking time and battery life in a handset. In order to increase data rate, most 3G systems involve high peak-to-average signals. This requires that all the components in the transmitter, especially the power amplifier at the front end, must be able to handle peak power levels without linearity degradation. Otherwise, there will be spurious signals and in-band interference, which makes the detection of the desired signal difficult. Therefore, the main challenge in the 3G power amplifier design is to achieve high efficiency and simultaneously high linearity. The focus of this dissertation is the design of high efficiency and high linearity SiGe power amplifiers (PAs) for 3G WCDMA handset applications.

![Figure 1.2 Distribution diagram of handset market in 2004 [3].](image)
This introductory chapter will cover the research background. To begin with, we will introduce WCDMA systems, handset transceivers, and specifications of power amplifiers. Reported CDMA/WCDMA power amplifiers will be also summarized. Subsequently, the classification of power amplifiers and different power amplifier technologies will be discussed. Next, the research motivation and focus will be presented and finally we will conclude this chapter by outlining the organization of this dissertation.

1.1 Power Amplifiers for WCDMA Handsets

With the advancement of wireless communication systems, a greater challenge is imposed on reducing the size, cost and power consumption of the handset chipset. As a result, more stringent requirements are set for power amplifiers. This section will provide the background for WCDMA handset power amplifiers.

1.1.1 WCDMA System

Wideband Code Division Multiple Access (WCDMA) is one of the technologies selected for the air interface of the 3GPP standard [4]-[5], also known as UMTS (Universal Mobile Telecommunications System). This interface supports data rates of 2 Mbps in selected areas (e.g. indoors) and 384Kbps everywhere on a common 5MHz frequency carrier. It provides a variety of wideband services such as high-speed web access, high quality image transmission and live video conferencing in a flexible and efficient way with global roaming capability.
As shown in Figure 1.3, WCDMA systems evolve from Global System for Mobile Communications (GSM) technology, which is the mostly widely deployed air interface standard in the world. In general, the WCDMA system adopts the frequency division duplex (FDD) scheme [7], where the system operates in the uplink (from mobiles to base stations) at the band of 1920-1980MHz and in the downlink (from base stations to mobiles) at the band of 2110-2170MHz. As shown in Figure 1.4, both uplink and downlink bands are divided into twelve frequency channels. Each channel has a bandwidth of 5MHz. Any two 190MHz-apart channels form a duplex pair for two-way simultaneous communication [8].
Figure 1.4 Operating band diagram in WCDMA systems [8].

WCDMA is a direct sequence spread spectrum system [9]. Its spreading and despreading principle is illustrated in Figure 1.5. The direct sequence spreading process multiplies an information stream with a high chip rate pseudo-noise (PN) code. WCDMA systems spread the bandwidth of an information stream to a much wider bandwidth and a lower power spectral density (PSD) than other multiple access systems, such as frequency division multiple access (FDMA) and time division multiple access (TDMA). Figure 1.5(a) shows the original wideband signal for several users with frequency division duplex, in which the signal of user 1 is the desired data. Figure 1.5(b) illustrates the wideband signal after the spreading operation in the transmitter, which is the same as that received at the receiver. The despread signal spectrum is shown in Figure 1.5(c), where only the signal of user 1 will be retrieved by the receiver and the rest of the signals appear as noise. If the interference between users in a cell does not exceed the system specification, new users can be added into the cell sharing the same frequency band. Consequently, there
is no absolute upper limit on the number of users that can be accommodated simultaneously in each cell. This feature distinguishes WCDMA systems from FDMA and TDMA systems.

Figure 1.5 Principle of spread spectrum technique. (a) Frequency division duplex. (b) The wideband signal after the spreading operation. (c) Despreading for user 1.

On the other hand, with an increase of the number of users in the same cell, the interference may increase to intolerable levels. To maintain high communication quality, WCDMA handsets need to comply with the linearity specifications defined by 3GPP standard in terms of adjacent channel power ratio (ACPR). As shown in Figure
ACPR is defined as the ratio of the leakage power measured in the adjacent channel (channel $n+1$) to the transmitted power in the transmission channel (channel $n$). The maximum allowable ACPRs over a 3.84MHz channel bandwidth at 5 MHz offset (channel $n+1$) and 10 MHz offset (channel $n+2$) are -33 dBC and -43 dBC respectively.

![Diagram of ACPR spectrum definition of WCDMA systems.](image)

**Figure 1.6 ACPR spectrum definition of WCDMA systems.**

### 1.1.2 WCDMA Handset Transceiver

A handset chipset of WCDMA systems consists of two major blocks: a transmitter and a receiver. The simplified schematic of a typical WCDMA handset transceiver is shown in Figure 1.7, where the top path after the duplexer is the receiver and the bottom path is the transmitter. Analog-to-Digital (A/D) and Digital-to-Analog (D/A) converters are not shown at the baseband side. Power control, automatic frequency control and automatic gain control are not displayed at the system level. They are all essential for the full implementation of WCDMA systems, but ignored
here for simplicity. The transmitter forms the uplink of the WCDMA handset system and the power amplifier is the front end block of the transmitter.

Figure 1.7 Simplified schematic of a typical WCDMA handset transceiver.

To accommodate different types of WCDMA wireless applications, different power classes are required as the maximum transmit power at the output of the transmitter. The classification of power classes is detailed in Table 1.1 according to the 3GPP specifications [10]. Table 1.1 indicates that both power Class 3 and 4 are applicable to three operation bands (band I: 1920-1980MHz; band II: 1850-1910MHz; band III: 1710-1780MHz). Furthermore, power Class 3 deals with higher power level than power Class 4 does. Thus, power Class 3 is the most applicable power class in practice and is therefore the design goal in this dissertation.
Table 1.1 Transmit power classes in different WCDMA handsets [10].

<table>
<thead>
<tr>
<th>Operating Band</th>
<th>Power Class 1</th>
<th>Power Class 2</th>
<th>Power Class 3</th>
<th>Power Class 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Power (dBm)</td>
<td>Tol (dB)</td>
<td>Power (dBm)</td>
<td>Tol (dB)</td>
</tr>
<tr>
<td>Band I</td>
<td>+33</td>
<td>+1/-3</td>
<td>+27</td>
<td>+1/-3</td>
</tr>
<tr>
<td>Band II</td>
<td>-</td>
<td>-</td>
<td>+24</td>
<td>+1/-3</td>
</tr>
<tr>
<td>Band III</td>
<td>-</td>
<td>-</td>
<td>+24</td>
<td>+1/-3</td>
</tr>
</tbody>
</table>

1.1.3 WCDMA Power Amplifier Specifications

Based on the discussions of WCDMA handset requirements in the previous sections, the specifications of WCDMA handset power amplifiers can be derived as follows. The maximum output power $P_{out}$ of PAs for different classes is found to be from +23dBm to +35dBm by adding 2dB to the maximum output powers of different power classes in Table 1.1, where 2dB is the power loss in the duplex filter between the power amplifier and the antenna. Since the reported maximum output power of the WCDMA TxIC driver stage (the RF block before the PA) is in the range between +6dBm and +10dBm [11], the average value of +8 dBm is taken as the input power $P_{in}$. Therefore, the power gain $G$ of the PA is obtained to be from 15dB to 27dB with $G = P_{out} - P_{in}$. At the maximum output power, the PA has to meet the linearity requirements specified in terms of ACPR, i.e. -33dBc and -43dBc at 5 and 10MHz offsets respectively. To support long talk time and advanced multimedia services, a
stringent requirement is imposed on efficiency. For SiGe WCDMA PAs, the maximum power added efficiency is typically around 30% [12]. The specifications of WCDMA handset PAs are summarized in Table 1.2.

Table 1.2 3GPP WCDMA handset power amplifier specifications.

<table>
<thead>
<tr>
<th>Specifications</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Frequency</td>
<td>1.92GHz – 1.98GHz</td>
</tr>
<tr>
<td>Maximum Output Power</td>
<td>23dBm – 35dBm</td>
</tr>
<tr>
<td>Power Gain</td>
<td>15dB – 27dB</td>
</tr>
<tr>
<td>PAE @ Maximum Output Power</td>
<td>&gt; 30%</td>
</tr>
<tr>
<td>ACPR (over a 3.84MHz channel bandwidth)</td>
<td>-33dBc @ 5MHz offset</td>
</tr>
<tr>
<td></td>
<td>-43dBc @ 10MHz offset</td>
</tr>
</tbody>
</table>

1.1.4 Survey of Reported CDMA/WCDMA Handset Power Amplifiers

Power amplifiers implemented in different semiconductor technologies have their respective advantages and disadvantages. For example, InGaP/GaAs HBT power amplifiers typically achieve better efficiency and linearity than CMOS or SiGe HBT power amplifiers. However, the cost of InGaP/GaAs HBT processes is higher than that of CMOS or SiGe HBT process.

Table 1.3 compares the performance of recently reported CDMA/WCDMA handset power amplifiers in different technologies. To the author’s knowledge, SiGe power amplifiers are becoming increasingly popular in the market for 3G handset
power amplifiers. The reasons for the popularity of SiGe technology are explained in Section 1.3.

Table 1.3 Comparison of recently reported CDMA/WCDMA handset power amplifiers.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Technology</th>
<th>Pout (dBm)</th>
<th>PAE@Pout</th>
<th>Gain (dB)</th>
<th>Bias Voltage</th>
<th>ACPR@Pout</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim '04 [13]</td>
<td>InGaP/GaAs HBT</td>
<td>27</td>
<td>33 %</td>
<td>25</td>
<td>3.4 V</td>
<td>-30 dBC @5 MHz</td>
<td>WCDMA</td>
</tr>
<tr>
<td>Wang '04 [14]</td>
<td>CMOS</td>
<td>24</td>
<td>29 %</td>
<td>23.9</td>
<td>3.3 V</td>
<td>-35 dBC @5 MHz</td>
<td>WCDMA</td>
</tr>
<tr>
<td>Jeon '04 [15]</td>
<td>InGaP/GaAs HBT</td>
<td>28</td>
<td>37 %</td>
<td>27</td>
<td>3.5 V</td>
<td>-47 dBC @1.25MHz</td>
<td>CDMA</td>
</tr>
<tr>
<td>Noh '04 [16]</td>
<td>InGaP/GaAs HBT</td>
<td>28.3</td>
<td>52.4 %</td>
<td>11.5</td>
<td>3.4 V</td>
<td>-33 dBC @5 MHz</td>
<td>WCDMA</td>
</tr>
<tr>
<td>Staudinger '02 [17]</td>
<td>PHEMT</td>
<td>30</td>
<td>38 %</td>
<td>27</td>
<td>3.5 V</td>
<td>-42 dBC @5 MHz</td>
<td>WCDMA</td>
</tr>
<tr>
<td>Hau '01 [18]</td>
<td>HJFET</td>
<td>26</td>
<td>57.4 %</td>
<td>n/a</td>
<td>3.5 V</td>
<td>-40 dBC @5 MHz</td>
<td>WCDMA</td>
</tr>
<tr>
<td>Luo '01 [19]</td>
<td>Si BiCMOS</td>
<td>28.2</td>
<td>30 %</td>
<td>21.5</td>
<td>3.6 V</td>
<td>-45 dBC @1.25MHz</td>
<td>CDMA</td>
</tr>
<tr>
<td>Vintola '01 [20]</td>
<td>AlGaAs/GaAs HBT</td>
<td>24</td>
<td>27 %</td>
<td>30</td>
<td>3.5 V</td>
<td>-36 dBC @5 MHz</td>
<td>WCDMA</td>
</tr>
<tr>
<td>Tseng '00 [21]</td>
<td>SiGe HBT</td>
<td>28</td>
<td>36 %</td>
<td>22</td>
<td>4 V</td>
<td>-44 dBC @1.25MHz</td>
<td>CDMA</td>
</tr>
<tr>
<td>Iwai '00 [22]</td>
<td>InGaP/GaAs HBT</td>
<td>27</td>
<td>42 %</td>
<td>30.5</td>
<td>3.5 V</td>
<td>-38 dBC @5 MHz</td>
<td>WCDMA</td>
</tr>
<tr>
<td>Kawamura '00 [23]</td>
<td>GaAs HBT</td>
<td>27.6</td>
<td>44 %</td>
<td>21</td>
<td>3.4 V</td>
<td>-37 dBC @5 MHz</td>
<td>WCDMA</td>
</tr>
<tr>
<td>Hanington '99 [24]</td>
<td>GaAs MESFET</td>
<td>24</td>
<td>20 %</td>
<td>12</td>
<td>Dynamic biasing</td>
<td>-26 dBC @1.25MHz</td>
<td>CDMA</td>
</tr>
</tbody>
</table>

1.2 Power Amplifier Classification

Power amplifiers can be classified into different categories. The most common method is based on the maximum achievable efficiency of the power amplifier. It is necessary to review one of the most important definitions – efficiency in the PA design. There are two typical definitions for the efficiency in RF PAs [25]: output
efficiency and power added efficiency. Output efficiency is defined as the ratio of the RF output power $P_{out}$ to the dc power $P_{dc}$:

$$\eta = \frac{P_{out}}{P_{dc}}$$  \hspace{1cm} (1.1)

However, the above definition does not take into account the input power $P_{in}$ and power gain $G$, whose effects are significant in RF PAs. This results in the definition of Power Added Efficiency (PAE):

$$PAE = \frac{P_{out} - P_{in}}{P_{dc}}$$  \hspace{1cm} (1.2)

$$= \frac{P_{out} - \frac{P_{out}}{G}}{P_{dc}}$$  \hspace{1cm} (1.3)

$$= \eta \cdot \left(1 - \frac{1}{G}\right)$$  \hspace{1cm} (1.4)

According to their maximum possible efficiencies as shown in Table 1.4, RF power amplifiers can be classified as Class A, AB, B, C, D, E, and F [25]-[26]. Class A, AB, B, and C employ the same circuit topology as shown in Figure 1.8. The difference among them stems from the base bias of the transistor: Class A is always biased in the active region (where the conduction angle of the transistor is 360°); Class B is biased at the threshold of conduction (where the conduction angle is 180°); Class AB represents the case where the conduction angle is between 180° and 360°; Class C is biased below the threshold (where a typical value of the conduction angle is 150°).
Table 1.4 Comparison of peak possible efficiency for different operation classes of power amplifiers.

<table>
<thead>
<tr>
<th>Classification</th>
<th>A</th>
<th>AB</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Efficiency (%)</td>
<td>50</td>
<td>50~78.5</td>
<td>78.5</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 1.8 Simplified schematic of a bipolar power amplifier operating in Class A/B/C.

Class D PAs in Figure 1.9(a) use two transistors as switches to generate a square drain-voltage waveform shown in Figure 1.9(b). An output filter is used to pass only the fundamental-frequency component. When \( V_o \) is high, there is no current flowing through the bottom switching transistor. Therefore, an ideal Class D PA can achieve a maximum output efficiency of 100%. Nevertheless, practical Class D PAs exhibit poor efficiency because parasitic capacitance at the output results in substantial loss [27]. Recently, some Class D PAs have been reported with an operating frequency at 1GHz [28]. However, the application of Class D topology is limited beyond that frequency because of the dramatic increase of parasitic capacitance loss.
Class E is another topology for switching-mode PAs [29], which adopts a single transistor as a switch, as shown in Figure 1.10(a). Since the drain current and voltage of the switching transistor never coexist at the same time, the maximum achievable output efficiency can be as high as 100%. In an optimum Class E power amplifier, both the drain voltage $V_D$ and the slope of $V_D$ drop to zero when the switch turns on, as shown in Figure 1.10(b). The corresponding values of $L$ and $C$ in the figure are calculated from the load resistance $R$ [30]:

$$L = \frac{1.15 \cdot R}{\omega_o} \quad (1.5)$$

$$C = \frac{0.1836}{\omega_o \cdot R} \quad (1.6)$$

where $\omega_o$ is the fundamental frequency. At the RF frequency, variations in the load resistance and switch shunt susceptance result in the degradation of power efficiency.
Some work [31] has been published for Class E power amplifiers at the frequency below 1GHz.

Figure 1.10 (a) Class E power amplifier circuit. (b) Ideal voltage and current waveforms.

Figure 1.11 (a) Class F power amplifier circuit. (b) Ideal voltage and current waveforms.

Class F differs from Class A/B/C in the harmonic tuning network at the output [32], as shown in Figure 1.11. Ideally, $Z_C$ is infinite at odd harmonic frequencies (i.e. $(2n-1)f_o$, where $n$ is a positive integer and $f_o$ is the RF signal frequency) and zero at even harmonic frequencies $(2nf_o)$. Consequently, the collector voltage is shaped to
a square wave with the addition of odd harmonics and the collector current is shaped to a half sine wave with the addition of even harmonics. This waveform shaping technique dramatically reduces the power loss in the transistor and thus the maximum achievable output efficiency of Class F power amplifiers is 100% [32]. Nevertheless this is not realistic in practice. Experiments [33]-[34] have already confirmed that at RF frequencies the efficiency drops dramatically because of lossy on-chip components, particularly inductors [35].

Until now, we have introduced the general background of different operation classes of RF power amplifiers. In order to find out the one suitable for WCDMA handset PAs, it is necessary to comprehend the relationships of efficiency and linearity versus different operation classes. As an illustration, a Class AB bipolar power amplifier will be compared to its Class A and B counterparts. Its simplified schematic is shown in Figure 1.8. The collector dc bias current $I_c$ in Class AB is much lower than that in Class A (e.g. 100mA vs. 400mA). Since the power efficiency of the power amplifier is inversely proportional to the dc power consumption, the power efficiency in Class AB is better than in Class A.

There is another way to account for the reason Class AB may achieve better power efficiency than Class A [36]. In Class AB power amplifiers, the output voltage and current waveforms at the collector of the transistor are shaped such that for a certain period $T_{ov}$ no overlap occurs between these two waveforms, as shown in Figure 1.12(a). On the other hand, there is always some overlap between output
voltage and current waveforms in Class A as shown in Figure 1.12(b). Thus the power dissipated in the transistor in Class A is higher than that in Class AB. Furthermore,

\[ \eta = \frac{P_{out}}{P_{dc}} = \frac{P_{out}}{P_{out} + P_{diss} + P_{para}} \]  

(1.7)

where \( P_{out} \) is the output signal power, \( P_{dc} \) the total dc power consumption, \( P_{diss} \) the power dissipated in the transistor, and \( P_{para} \) the power lost in matching networks due to non-ideal matching components. Since the power dissipated in the transistor in Class AB is smaller, its power efficiency is better than that of Class A. A similar principle accounts for the reason the efficiency of Class B is even better.

![Figure 1.12 Output voltage and current waveforms at the collector of power amplifiers. (a) Class AB. (b) Class A.](image)

On the other hand, the linearity of the power amplifier degrades in the opposite trend. In Class B operation, the transistor is set to the threshold of conduction, which means the transistor is turned off for half of the period \( T = \frac{1}{f} \) (\( f \) is the fundamental frequency). Nonlinear effects in the power amplifier become stronger through the
cutoff behavior of the transistor [25], compared to those in Class A and AB. The linearity of the power amplifier degrades accordingly.

In mobile wireless systems with constant envelope signals\(^1\), like GSM with Gaussian Minimum Shift Keying (GMSK), power amplifiers are mainly designed for high efficiency and there is no specific linearity requirement. Therefore, nonlinear but high efficiency PA designs, such as Class B [37] and Class E [31], are typically employed in these systems. In wireless systems with non-constant envelope signals, such as CDMA and Wideband CDMA, stringent linearity requirements are posed in the design of power amplifiers. Class AB operation is typically selected [38]-[39], because it has the best tradeoff between efficiency and linearity.

### 1.3 Power Amplifier Technologies

In the market, there are a variety of technologies for RF power amplifiers [40]. RF power amplifiers are normally characterized in device physics by cutoff frequency\(^2\), breakdown voltage\(^3\), thermal conductivity\(^4\), integration level\(^5\), size and cost\(^6\).

---

\(^1\) Constant envelope signal means the maximum and minimum amplitude of the signal over one period is constant. For instance, the sine wave is an ideal constant envelope signal.

\(^2\) Cutoff frequency is the frequency at which the magnitude of the current gain is equal to unity. It determines power gain of the transistor.

\(^3\) Breakdown voltage of the transistor defines the maximum voltage swing at the output of the transistor.

\(^4\) Thermal conductivity indicates the heat flow rate of the material. Low thermal conductivity leads to thermal stability issue.

\(^5\) Integration with other circuit blocks (transmitter IC and digital circuit) on a single chip is another figure of merit to further reduce the total cost for the whole chipset.

\(^6\) Size and cost are the major considerations for successful commercial products.
Furthermore, they are assessed with efficiency or PAE, power gain, and linearity in terms of inter-modulation distortion (IMD)\(^7\) or ACPR. Table 1.5 summarizes several popular technologies for RF power amplifiers. Their devices are characterized with their schematic cross-sections, as shown in Figure 1.13.

**Table 1.5 Comparison of typical technologies for RF power amplifiers [40].**

<table>
<thead>
<tr>
<th>Technology</th>
<th>GaAs HBT</th>
<th>Si BJT</th>
<th>SiGe HBT</th>
<th>CMOS</th>
<th>LDMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>(f_T)</td>
<td>46GHz</td>
<td>27GHz (HF(^1)) 22GHz (HV(^2))</td>
<td>44GHz (HF) 25GHz (HV)</td>
<td>&lt; 20GHz</td>
<td>&lt; 20GHz</td>
</tr>
<tr>
<td>(BV_{ceo})</td>
<td>14.3V</td>
<td>3.3V (HF)</td>
<td>3.0V (HF)</td>
<td>5V</td>
<td>15V</td>
</tr>
<tr>
<td>Thermal Conductivity</td>
<td>0.49 W/cm-C</td>
<td>1.5 W/cm-C</td>
<td>1.5 W/cm-C</td>
<td>1.5 W/cm-C</td>
<td>1.5 W/cm-C</td>
</tr>
<tr>
<td>Normalized Cost 1mm(^2)</td>
<td>1</td>
<td>0.3</td>
<td>0.3</td>
<td>&lt; 0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Wafer Size</td>
<td>6'' wafer</td>
<td>8'' wafer</td>
<td>8'' wafer</td>
<td>12'' wafer</td>
<td>8'' wafer</td>
</tr>
<tr>
<td>PAE</td>
<td>&gt; 40%</td>
<td>&gt; 30%</td>
<td>&gt; 30%</td>
<td>&gt; 30%</td>
<td>&gt; 35%</td>
</tr>
<tr>
<td>Linearity</td>
<td>great</td>
<td>good</td>
<td>good</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Power Gain</td>
<td>great</td>
<td>good</td>
<td>great</td>
<td>poor</td>
<td>poor</td>
</tr>
<tr>
<td>Integration Level</td>
<td>poor</td>
<td>good</td>
<td>good</td>
<td>great</td>
<td>good</td>
</tr>
</tbody>
</table>

\(^1\)High frequency npn transistor.

\(^2\)High breakdown npn transistor.

\(^7\) IMD is characterized by the appearance in the output of frequencies equal to the sums and differences of integral multiples of the two or more component frequencies present in the input waveform. For instance, if two tones (\(\omega_1\) and \(\omega_2\)) pass through a nonlinear PA, then at the output of the PA, frequency components \(2\omega_1-\omega_2\) and \(2\omega_1+\omega_2\) are generated by the third-order inter-modulation.
The GaAs HBT is the preferred technology [41] in the current commercial handset PA market. Its efficiency and linearity are so great that fast engineering development becomes feasible to meet the design specifications on the first design.
iteration. High cost and low integration level are the two drawbacks of the GaAs HBT.

From the standpoints of fabrication cost and integration level, CMOS is the best among the four technologies. However, there are a number of issues limiting its application in the PA market. For the same functionality, CMOS PAs are larger than their GaAs or SiGe counterparts. Unlike GaAs or SiGe HBT technology, which incorporates vertical transistors, the footprint of the MOSFETs in CMOS increases rapidly when scaled up to accommodate high output powers. This tends to cancel out the advantage of having a low-cost manufacturing process. Also, the low breakdown voltage in CMOS limits the maximum voltage swing and its low transconductance-to-current ratio results in relatively poor power gain.

LDMOS has a low fabrication cost and good integration level. However, the bias voltage in LDMOS PAs is quite high, which makes this technology typically suitable for base-station PAs, instead of handset applications. Similar to CMOS, the application of LDMOS in PA market is also limited due to its poor power gain.

In recent years, the SiGe HBT has become a competitive candidate [42] for the development of cellular handset power amplifiers, since the SiGe HBT exhibits good linearity, low-cost and compatibility with BiCMOS technology [43], even though the SiGe HBT has a lower breakdown voltage and efficiency than its GaAs counterpart and is also affected by the thermal runaway issue. Compared with the Si BJT, the SiGe HBT has a higher cutoff frequency $f_T$ for the same bias current and a higher
current gain $\beta$. For applications (such as low-noise amplifiers) where the noise performance is critical, the SiGe HBT can tradeoff its high $\beta$ for lower base resistance $r_b$ by doping the base region heavily. This approach not only lowers the thermal noise of the device, but increases the maximum oscillation frequency $f_{\text{max}}$ according to [44]

$$f_{\text{max}} = \sqrt{\frac{f_T}{8\pi C_{\text{fc}} r_b}}$$

(1.8)

These advantages make the SiGe HBT very attractive in the market of handset RF PAs.

1.4 Research Focus

To increase talk time and battery life, high-efficiency power amplifiers are necessary in handset designs. However, with the advancement of WCDMA systems, the traditional Class AB power amplifiers may not satisfy the stringent requirements for both efficiency and linearity. To solve these issues, this dissertation presents new integrated dynamic biasing techniques with digital pre-distortion (DP) in a SiGe BiCMOS process for WCDMA handset applications. A few of the key results of this dissertation are summarized here:

- New dynamic biasing techniques have been proposed to improve the average power efficiency of RF PAs. Furthermore, the proposed techniques can keep the power gain of the power amplifier constant. Non-constant power gains in previous
dynamic biasing techniques would create problems in the power control loop operation of the WCDMA handset. Prototype PAs with proposed dynamic biasing techniques have been implemented in a 0.25µm SiGe BiCMOS process. Experimental results indicated that average power efficiency was improved by more than 160% and power gain was kept within the 2dB variation while the specifications of WCDMA Class 3 handsets are satisfied.

- A varied-coefficient Volterra series technique has been applied to analyze the nonlinear behavior of SiGe HBT power amplifiers. The dominant nonlinearity sources in SiGe HBT PAs are highlighted. Our analysis includes the effect of emitter bond-wire inductance, which is crucial to accurately predict the intermodulation distortion of power amplifiers. The analysis was verified by simulation and experimental results.

- A three-tone Volterra model using the “Nonlinear Current Method” has been derived and applied to the analysis of an adaptive power amplifier with envelope signal injection [45]. The model was well confirmed by simulation and experimental results.

- Baseband adaptive memoryless digital pre-distortion was used to improve the linearity of WCDMA power amplifiers and further boost its efficiency. Experimental results demonstrated that the overall ACPR improvement with the digital pre-distortion is more than 8dB and the peak power added efficiency is improved by 60%.
1.5 Dissertation Organization

This dissertation is organized as follows.

Chapter 2 presents the efficiency boosting techniques. We will begin with an in-depth review of the existing efficiency improving methods. Subsequently two new dynamic biasing techniques will be proposed to combat the limitations of the existing methods. Besides improving average power efficiency, the new techniques can keep the power gain of the power amplifier constant. Design methodology and layout implementation will be covered in detail.

Chapter 3 analyzes the linearity of SiGe HBT power amplifiers by using a varied-coefficient Volterra Series. Different linearity analysis methods are first compared to illustrate the issues in high-power amplifier linearity analysis. We will then introduce the principle of the power dependent coefficient Volterra analysis. A SiGe HBT high-power amplifier will be analyzed to clarify its main nonlinearity contributors. An adaptive RF amplifier with envelope signal injection will be also analyzed to illustrate that the linearity of the amplifier can be improved by optimally injecting the envelope signal. Both analyses are confirmed with simulation and experimental results.

Chapter 4 describes the implementation of memoryless digital pre-distortion in the dynamically biased power amplifiers described in Chapter 2. Besides a brief introduction of different pre-distortion techniques, the implementation of the digital
pre-distortion system will be illustrated. Experimental results will be provided to demonstrate the effectiveness of the digital pre-distortion technique.

Chapter 5 presents the measured results of the dynamically biased power amplifiers. Key design parameters, such as dc current consumption, power gain, power added efficiency, and ACPR will be presented. The prototype power amplifiers will be compared to other reported power amplifiers with dynamic biasing techniques.

Chapter 6 concludes the dissertation with a summary of this work. Future directions are discussed in the field of high-efficiency and high-linearity power amplifiers for advanced handset applications.
CHAPTER 2
Efficiency Improvement by Dynamic Biasing Techniques

Advanced wireless communication systems, such as 3G WCDMA systems, support data-oriented applications, which require higher dc power consumption. In the near future, cellular battery technology will not improve dramatically to accommodate this stringent requirement. Therefore, low-power and high-efficiency circuits are highly desirable in handset applications.

Power amplifiers are the bottleneck of RF power consumption in handsets. As shown in Figure 2.1, a typical Class AB power amplifier consumes more power than a transmitter IC (TxIC) and therefore the power amplifier is the biggest RF dc power consumer in handsets. Since the dc power consumption remains constant at the low-power region, the efficiency of the power amplifier is degraded as the RF output power decreases. Furthermore, more than 90 percent of the output power occurs between −15 dBm and +15 dBm, where the efficiency is low. As a result, average power efficiency (over the full range of output powers), instead of peak power efficiency, is the key factor determining the battery life and talk time for portable wireless applications [48].
Average power efficiency is a measure of the ratio of the total energy transmitted to the total energy drawn from the battery, i.e.

$$\langle \eta \rangle = \frac{\langle P_{\text{out}} \rangle}{\langle P_{\text{dc}} \rangle} = \frac{\int P_{\text{out}} p(P_{\text{out}}) dP_{\text{out}}}{\int \frac{P_{\text{out}} p(P_{\text{out}}) dP_{\text{out}}}{\eta(P_{\text{out}})}}$$  \hspace{1cm} (2.1)$$

where \(P_{\text{out}}\) is the output power, \(p(P_{\text{out}})\) is probability of a certain output power \(P_{\text{out}}\), and \(\eta(P_{\text{out}})\) is the PAE at \(P_{\text{out}}\). Using the PDF of Figure 2.1, the average power efficiency of a typical Class-AB amplifier is very low – typically below 2%. Therefore, a high average-efficiency power amplifier is crucial to prolong the battery life and support advanced applications in WCDMA handsets.

To begin with, different efficiency improving methods will be reviewed to evaluate their strengths and weaknesses. Then two new dynamic biasing techniques
will be proposed to overcome the limitations of the existing methods. Design considerations and layout details will be presented, which are followed by concluding remarks.

2.1 Survey of PA Efficiency Boosting Techniques

A variety of efficiency enhancement techniques [49]-[60] have been proposed and implemented for power amplifiers. Their relative advantages and disadvantages are discussed below.

2.1.1 Constant Current/Voltage Biasing

Constant current biasing (CCB) and constant voltage biasing (CVB) are two typical biasing methods in PA design. As indicated by their names, CCB and CVB provide constant dc bias current and voltage respectively to the base of the power amplifier. A typical schematic of CCB is shown in Figure 2.2. At node $A$, a constant voltage is obtained when the external current source $I_{dc}$ remains constant. By connecting node $A$ to the base with a resistor $R_{bias}$, a constant bias current $I_{bias}$ ($V_A - V_{BE}$) is provided to the base of the transistor.

When RF input power increases, the collector dc current does not change due to the fixed base dc current. This results in a drop of the quiescent base-emitter voltage as shown in Figure 2.3. When the quiescent base-emitter voltage drops, the collector current would clip at high input powers and generate unwanted harmonics. Consequently, the linearity of the amplifier will be degraded.
A simplified schematic of CVB is shown in Figure 2.4. At the base of the transistor, a constant bias voltage is provided through a large inductor. However, the large on-chip inductor takes much area and is not desirable in practical designs. On the other hand, the implementation of an off-chip inductor will degrade the circuit integration and bring extra cost. As a result, it is often replaced with an op-amp circuit, which will be discussed in Section 2.3.4.
Figure 2.4 Schematic of constant voltage biasing (CVB).

The large signal characteristic of CVB is shown in Figure 2.5. When the RF input power increases, the base dc current of CVB increases, so does the collector dc current. Consequently, the collector current will clip less than that of CCB at high input powers, which helps with the linearity of the amplifier. This explains the reason CVB is more popular in RF PA design. However, as shown in Figure 2.1, a Class AB PA with CVB still has low efficiencies at low-power levels.
2.1.2 Doherty Amplifier

The Doherty configuration is an efficiency enhancement technique, which was originally proposed in 1936 [49]. The block diagram of a Doherty amplifier is shown in Figure 2.6. It consists of a main amplifier, a peak amplifier, a phase adjuster (quarter-wave transmission line), and an amplitude attenuator. With phase and amplitude adjustments, output powers from both devices can be combined effectively. At low input powers, only the main amplifier is active; at high input powers, both the main and peak amplifiers contribute to the output power.

![Block diagram of Doherty amplifier.](image)

**Figure 2.6 Block diagram of Doherty amplifier.**

The power transfer characteristic is shown in Figure 2.7, where the upper 6dB power range is the region for power combination. At the low-power level, the load impedance is $R_L$; at the high-power level, the load impedance dynamically decreases to accommodate more current swing (contributed from both devices) while maintaining maximum voltage swing, as shown in Figure 2.8. The effective load impedance reduces to $R_L \left(1 + \frac{I_{\text{peak}}}{I_{\text{main}}} \right)$, where $I_{\text{main}}$ and $I_{\text{peak}}$ are the currents of
the main and peak devices respectively and they are in anti-phase. This dynamic-load feature is called an active load-pull technique.

Figure 2.7 Power transfer characteristic of Doherty amplifier.

Figure 2.8 Load line diagram of Doherty amplifier operation.

The efficiency of the Doherty configuration also benefits from the dynamic-load feature. Assume that the main amplifier operates in Class B and the Doherty amplifier
is capable of producing a maximum efficiency of 78.5% over the upper 6 dB power range. The efficiency of a two-way Doherty amplifier has been derived by Raab [50], as indicated in (2.2). The detailed derivation is omitted here for simplicity.

\[ \eta = \frac{\pi}{2} \frac{\left( \frac{v_{\text{in}}}{V_{\text{max}}} \right)^2}{3 \left( \frac{v_{\text{in}}}{V_{\text{max}}} \right)^2 - 1} \]  

(2.2)

where \( v_{\text{in}} \) is the input voltage and \( V_{\text{max}} \) is the maximum input voltage. Doherty power amplifiers deliver much better efficiency than Class B power amplifiers, as shown in Figure 2.12. The linearity of the Doherty configuration may be improved by adopting feed-forward or pre-distortion techniques since it is highly nonlinear. It also has other drawbacks [51] such as gain degradation, intermodulation distortion (IMD), and narrow bandwidth. Gain degradation and IMD are primarily caused by the peaking amplifier. Narrow bandwidth stems from the use of a quarter-wave transmission line as the phase adjuster. Furthermore, considering its complexity, the Doherty configuration is typically adopted in base-station power amplifiers, not for handset applications.

### 2.1.3 Envelope Elimination and Restoration

Envelope elimination and restoration (EER) was first proposed in 1952 by Kahn [52]. The general principle is to provide a highly efficient and linear power amplification by combining a highly efficient switching-mode amplifier with a linear low-frequency amplifier (modulator). As shown in Figure 2.9, the limiter before the
nonlinear switching-mode PA (in Class D or Class E) eliminates the envelope and generates a constant-amplitude phase-modulated carrier signal. To enhance its efficiency, the nonlinear switching-mode amplifier operates in saturation at all times. The amplitude information will be restored through the amplitude modulation by the highly linear low-frequency modulator.

**Figure 2.9 EER block diagram.**

Since the switching-mode amplifier always operates in saturation, EER amplifiers may achieve very high efficiency as shown in Figure 2.12. However, since the envelope detecting signal path consumes quite an amount of power, the overall efficiency will be degraded considerably. At the same time, the bandwidth of EER systems is limited by the modulator and thus an even greater design challenge is posed for wideband applications. Furthermore, a good envelope control circuit is rather complex. These above issues make the EER configuration not well suited for handset power amplifiers.
2.1.4 Envelope Tracking

The envelope-tracking (ET) architecture is similar to the EER system. Both of them have an envelope-detecting path and an RF path. However, the RF drive of the RF power amplifier in ET preserves both amplitude and phase information, whereas the RF drive in EER only retains the phase information. The combination of an envelope detector and a DC-DC converter in ET systems, as shown in Figure 2.10, dynamically varies the supply voltage of the RF power amplifier. Since the dc power consumption is reduced at low powers, the efficiency is improved as shown in Figure 2.12.

![Figure 2.10 ET block diagram.](image)

Some ET systems [53] have been reported, but there are still a number of practical issues during its monolithic implementation. First, an off-chip inductor and capacitor are typically used in the DC-DC converter [54], which degrades the integration performance. The monolithic DC-DC converter with on-chip passive components has been implemented [55], but low-Q components considerably reduce
the overall efficiency of the ET system. Second, when the bias voltage and current are adjusted dynamically, the power gain of the RF PA varies dramatically [24]. The gain variation degrades the system performance in terms of error vector magnitude (EVM), which will be discussed in Section 2.1.6.

### 2.1.5 LINC Amplifier

The concept of LINC (LInear amplification using Nonlinear Components, also known as Chireix Outphasing Amplifier) was proposed by Chireix in 1935 [56]. The general principle is shown in Figure 2.11. First, an envelope modulated waveform $(S_{in}(t) = A(t) \cdot \cos(\omega t + \phi(t)))$ is decomposed into two out-phased constant envelope signals $(S_1(t) = \cos(\omega t + \phi(t) + \cos^{-1}(A(t)))$ and $S_2(t) = \cos(\omega t + \phi(t) - \cos^{-1}(A(t)))$. Subsequently, these two signals are individually applied to highly-nonlinear power amplifiers. The resulting output signals are then recombined through a passive combiner into the final output signal $S_{out}(t)$, where

$$S_{out}(t) = G[S_1(t) + S_2(t)]$$
$$= 2GA(t)\cos(\omega t + \phi(t)) \quad (2.3)$$

Since each amplifier operates in a highly nonlinear and efficient mode, the LINC system achieves a high efficiency. Power efficiencies of different efficiency enhancement techniques are compared in Figure 2.12. Except for the Kahn (i.e. EER) technique, which always has the highest efficiency, and Class A, which always has the lowest efficiency, every other technique has its own favorite region of high efficiency.
For instance, ET is better than LINC and Doherty at the low-power region and LINC is better than ET and Doherty at the high-power region.

Figure 2.11 LINC block diagram.

Figure 2.12 Efficiency as a function of output amplitude [1]. A: Class A; B: Class B; ET: Envelope Tracking; DOH: Doherty amplifier; CHIR: Chireix outphasing amplifier (i.e. LINC); KAHN: Kahn amplifier (i.e. EER).

The primary issue in LINC systems is the gain and phase imbalances between the two amplification paths. The typical tolerance for most applications is
approximately 0.1∼0.5dB in gain matching or 0.4∼2 degree in phase matching. This is nearly impossible to achieve in most practical situations. Although some advanced calibration algorithms [57]-[58] have been proposed to minimize the gain and phase errors, more work in this field must be done before LINC can be readily integrated into handset applications.

### 2.1.6 Dynamic Biasing

Dynamic biasing techniques are derived from the definition of the power efficiency in (1.1), where the power efficiency is inversely proportional to the dc power consumption. As a result, it is straightforward to increase the power efficiency by reducing the dc power consumption. This leads to three different techniques as shown in Figure 2.13: dynamic voltage biasing (DVB), dynamic current biasing (DCB), and dual dynamic biasing (DCB plus DVB). For instance, envelope tracking [53] is a type of DVB.

![Figure 2.13 BJT dc characteristics showing dynamic biasing techniques.](image-url)
Considering the performance of integration and control flexibility, most commercial power amplifiers, if they vary the bias at all, are using DCB [59]-[60]. Although the power efficiency is improved, the power gain of these power amplifiers drops drastically when switched from the high-power region into the low-power region. This may create problems in the operation of power control loop for WCDMA handsets, as well as degrade their error vector magnitude (EVM).

![QPSK signal constellation diagram](image)

Figure 2.14 QPSK signal constellation diagram.

EVM is a useful system-level figure of merit for the accuracy of digitally-modulated signals. As illustrated in Figure 2.14, after a QPSK signal passes through a power amplifier with a certain nonlinearity, the resulting output signal \( IQ(out) \) has some gain and phase mismatches (\( \Delta G \) and \( \Delta \theta \)) compared to the reference signal \( IQr(out) \). Accordingly, EVM is derived as:
\[ EVM = 100 \times \frac{|IQ(out) - IQR(out)|}{|IQR(out)|} \] (2.4)

The EVM specification of 3GPP WCDMA transmitters is 17.5% and in practice the EVM specification on the handset PAs is typically less than 5%. This poses a strict limitation on keeping gain and phase constant. Although this has not been highlighted in previous dynamic biasing techniques [59]-[60], this issue will be analyzed and solved in this dissertation.

### 2.2 Principles of Switched Dynamic Biasing Techniques

To overcome the issues discussed in the above efficiency enhancement techniques, the main objective of this dissertation is to provide a fully integrated solution with far better average power efficiency and constant power gain for WCDMA handset power amplifiers. Improved switched dynamic biasing techniques are proposed and implemented in this dissertation. Two types of switched dynamic biasing techniques, switched dynamic current biasing (SDCB) and switched dual dynamic biasing (SDDB), will be presented in detail.

#### 2.2.1 Switched Dynamic Current Biasing

To solve the gain-variation issue in previous dynamic current biasing techniques, it is necessary to find out the reason for this problem. A simplified schematic of a typical SiGe HBT power amplifier is shown in Figure 2.15, where IMN and OMN represent the input and output matching networks respectively. The small-signal model [61] of an HBT transistor is shown in Figure 2.16, where $r_b$ is the base
resistance, \( r_e \) and \( C_\pi \) are the base-emitter resistance and capacitance, \( C_{bc} \) is the base-collector capacitance, \( C_o \) is the transistor output capacitance, and \( g_m = \omega_f \cdot C_\pi \), in which \( \omega_f \) is the radian frequency at unity short-circuit current gain.

Figure 2.15 Simplified schematic of a typical SiGe HBT power amplifier.

Figure 2.16 Small signal model of HBT transistors.

By putting the HBT model into the amplifier circuit, the complete expression for the transducer gain of the amplifier (including the effects of source impedance \( Z_S \), load impedance \( Z_L \) and bond-wire inductance \( L_e \)) is derived as:
\[ G_T = \frac{4Z_S Z_L}{\omega^2 \left( L_e + \frac{Z_S}{\omega r_c} + Z_S C_{bc} Z_L \right)^2 + \omega^4 L_e^2 C_{bc}^2 Z_S^2} \]  

(2.5)

where \( r_b \) and \( C_o \) are included in \( Z_S \) and \( Z_L \) respectively, and \( r_\pi \) is omitted because its impedance is much bigger than that of \( C_\pi \) at the desired RF frequency. To get more insight into the gain-variation issue, it is helpful to simplify (2.5) as:

\[ G_T = \frac{4}{Z_S} \left( \frac{\omega r_c}{\omega} \right)^2 \frac{Z_L}{(1 + \omega r_c Z_L C_{bc})^2} \]

(2.6)

Clearly, the power gain consists of three terms, in which the first and last terms are related to the input and output matching networks, and the second one is the \( \omega r_c \) term. Since \( \omega r_c \) is proportional to \( J_C \) before the peak \( \omega_T \), the decrease of \( J_C \) would reduce \( \omega_T \). From (2.6), the reduction of \( \omega_T \) consequently decreases the power gain. In previous dynamic current biasing techniques, when the bias current drops, the total active device size remains the same. The reduced \( J_C \) results in the drop of the power gain. Furthermore, (2.6) points out the way to keep the gain constant: if we keep the collector current density \( J_C \) constant, then the power amplifier is operated at a constant \( \omega_T \) (i.e. \( f_T \) since \( \omega_T = 2\pi f_T \)), which keeps the power gain constant.

To lower the collector current and still keep its density constant, we utilize low-loss NFET switches [62] at the bases of the transistors to dynamically bias the SiGe HBT transistors, as depicted in Figure 2.17. In the high-power mode, all transistors
are switched on; in the low-power mode, the right $4m$ transistors are turned off and the left $m$ transistors are still on.

Figure 2.17 Output stage transistors with dynamic current biasing. HBTs are biased “on” or “off” in response to output power requirements.

The number of “on” transistors is adjusted in response to changes in the desired output power; the collector current density is increased slightly at the low-power region so as to keep the power gain constant. As shown in Figure 2.18, the power amplifier shifts to a higher transition frequency ($f_{T1}$) in the low-power mode, nevertheless it operates at a lower transition frequency ($f_{T2}$) in the high-power mode. Operating the transistor at this higher $f_{T1}$ enables us to keep the gain relatively constant by overcoming the effects of the extra parasitics in low-power mode. Consequently, when the power amplifier is adjusted from the high-power mode to the low-power mode, the dc power consumption is reduced, and the power gain remains constant.
2.2.2 Switched Dual Dynamic Biasing

As shown in Figure 2.13, the dual dynamic biasing technique can achieve better efficiency than its dynamic current biasing counterpart. Therefore, to further improve average power efficiency, we propose a switched dual dynamic biasing (SDDB) technique that reduces both dc bias current and voltage (DCB and DVB).

The simplified schematic of the output stage with SDDB is shown in Figure 2.19. There are two groups of SiGe HBT transistors: a high-power group and a low-power group. In the high-power group, the transistors are biased at $V_{CC}$, whereas in the low-power group the transistors are connected in series and accordingly biased at $V_{CC}/2$. The switching operation between these two power groups is controlled by low-loss NFET switches on the bases of the transistors. For instance, in the low-power mode, the high-power group is switched off and the low-power group is switched on.
In addition to the lower dc bias voltage, the low-power group is also biased with a reduced dc bias current. As a result, the dc power consumption will be dramatically reduced.

Figure 2.19 Simplified schematic of a two-step switched dual dynamic bias in the output stage.

Similar to SDCB, the power gain of SDDB should remain roughly constant in different power modes. The following design methodologies are employed to achieve this goal.
• A constant collector current density is essential for achieving constant gain. The reason for this is discussed in Section 2.2.1.

• The connection between two power groups adds parasitics to each individual group, which degrades the overall gain. On the other hand, this connection decreases the difference between two power gains in two power modes respectively, with two power groups disconnected. For example, the simulated gain difference between the two groups is 3dB with no connection between them, but only 1dB when they are connected.

• The routing line inductance $L_{\text{line}}$ from the output of the low-power group to that of the high-power group is optimized to boost the power gain in the low-power mode close to that in the high-power mode, since the parasitic line inductance may help cancel out certain capacitance in the low-power mode.

• The complete PA is a two-stage design. The inter-stage matching between the driver stage and the output stage is selected to keep power gains in two power modes to be close to each other and the overall gain to be more constant.

2.3 Design with Switched Dynamic Current Biasing

In this section, the schematic design of a single output stage PA will be presented as a demonstration of switched dynamic current biasing. Details include the selections of power step, power device size, switch size, bias network, and matching networks.
2.3.1 Power Step Selection

Ideally, the more power steps the power amplifier has, the better its average power efficiency will be. In practice, however, more power stages also generate more parasitics, which will degrade the efficiency. As a result, the efficiency improvement with more power steps will not be as dramatic as expected. In Table 2.1, “Ideal SDCB” represents the case with infinite steps for SDCB, where the bias current will be adjusted continuously, not discretely. The efficiency difference between 1-step SDCB and “Ideal SDCB” is less than 1.5%. Furthermore, the circuit complexity of “Ideal SDCB” is much higher than that of 1-step SDCB. Considering the tradeoff between efficiency improvement and circuit complexity, a 1-step topology was finally chosen for the design of SDCB.

Table 2.1 Average power efficiency comparison of SDCB in an output stage.

<table>
<thead>
<tr>
<th>Bias Type</th>
<th>Average Power Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class AB with CVB</td>
<td>2.44%</td>
</tr>
<tr>
<td>Class AB with 1-step SDCB</td>
<td>5.88%</td>
</tr>
<tr>
<td>Class AB with Ideal SDCB</td>
<td>7.36%</td>
</tr>
</tbody>
</table>

2.3.2 Power Device Selection

A proper transistor type is first selected for the power device. In the IBM SiGe BiCMOS 6HP process, there are two types of NPN transistors: high- $f_T$ (npnrf) and high-breakdown (npnrb). Since high $f_T$ results in high power gain, npnrf is
employed for small-signal applications such as LNA designs. For power amplifiers, \( npnbb \) is always used to deliver more output power because the high breakdown voltage of the transistor accommodates a larger voltage swing at the output.

Next, the transistor size is optimized to achieve the optimal tradeoff between the maximum output power and the power gain of the PA. On the one hand, larger transistor size can handle more output power. On the other hand, for the same bias current, larger transistor size corresponds to smaller bias current density and more parasitic capacitance, both of which normally result in a reduced power gain.

Assume the maximum output power is 27dBm (0.5W), the collector-emitter breakdown voltage (\( BV_{ceo} \), i.e. \( V_{ce,max} \)) of \( npnbb \) is 5.2V and the minimum collector-emitter voltage (\( V_{ce,min} \)) is 0.2V. The peak output ac current \( I_p \) is calculated from the equation

\[
P_{out} = \frac{V_p \cdot I_p}{2} = \frac{(V_{ce,max} - V_{ce,min}) \cdot I_p}{4}
\]

Thus, \( I_p = 400\text{mA} \). Since the maximum output current \( I_{max} = 2I_p \), \( I_{max} = 800\text{mA} \).

The \( f_T \) operating point of each transistor is displayed in Figure 2.18. For each transistor with the emitter size of 48\( \mu \text{m} \times 0.44\mu \text{m} \), the maximum current is 8mA. Correspondingly, the number of transistors is \( 800/8 = 100 \).
2.3.3 Optimization of NFET Switch Size

After selecting the size of the power device, it is also critical to optimize the NFET switch size, because the addition of NFET switches to the bases of transistors degrades the performance of the PA (in particular, power gain and linearity). These effects will be analyzed in detail in this section. Through the analysis, the NFET switch size can be optimized to minimize these deteriorating effects.

The effect of NFET switches on the power gain is analyzed in an equivalent circuit including NFET switches at the input of transistors as shown in Figure 2.20, where $R_{MOS}$ and $C_{MOS}$ are the equivalent parasitic resistance and capacitance of NFET switches and $R_{in}$ and $C_{in}$ are the equivalent input resistance and capacitance of transistors. When the device size of NFET switches is small, the resistance loss dominates; when the device size of NFET switches is large, the capacitance loss becomes dominant.

![Figure 2.20 Equivalent circuit including NFET switches.](image)
To find out the optimum device size, SPECTRE simulations of the power gain versus different device sizes of NFET switches were carried out. The simulation result is shown in Figure 2.21, where the optimal size of each NFET switch is found to be 3 fingers×15µm×0.26µm. As discussed in the previous section, there are 100 HBT devices and each transistor emitter is 48µm×0.44µm.

![Figure 2.21 Effect of NFET device size on power gain.](image)

At the same time, a theoretical analysis was also performed to confirm the above simulation results. Since there are two sources of power loss due to NFET switches – resistance loss ($\Delta Gain_{R_{MOS}}$) and capacitance loss ($\Delta Gain_{C_{MOS}}$), the gain loss can be expressed as:

$$\Delta Gain = \Delta Gain_{R_{MOS}} + \Delta Gain_{C_{MOS}}$$

$$\approx 10 \log \frac{R_{in}^2}{(R_{MOS} + R_{in})^2}$$

(2.8)

where the capacitance loss is insignificant compared to resistance loss. The validity of (2.8) was verified with SPECTRE simulations. Using extracted parameters
\( R_{\text{MOS}} = 0.3 \text{ohm}, \quad C_{\text{MOS}} = 3.26 \text{pF}, \quad R_{\text{in}} = 1 \text{ohm}, \quad \text{and} \quad C_{\text{in}} = 204 \text{pF} \), the power loss due to NFET switches is 2.5dB, which is very close to the simulation result.

The effect of the NFET switch size on the linearity (output 1dB compression point\(^1\)) of the power amplifier was simulated. As shown in Figure 2.22, the optimum NFET switch size of 3 finger\( \times 15 \mu \text{m} \times 0.26 \mu \text{m} \) found in Figure 2.21 is also the best choice for the linearity. The corresponding gain loss at 1.95GHz was 2.5dB and 1dB compression point was degraded by 1~2dB.

![Figure 2.22 Effect of NFET device size on output 1dB compression point.](image)

A number of design details in the implementation of NFET switches must be dealt with cautiously in order for the switches to function properly. First, NFETs must work in the triode region to make the on-resistance of NFET switches small, as

\(^1\) Output 1dB compression point in a PA represents the output power that drops by 1dB referred to the small-signal power gain of the PA.
indicated in Figure 2.23. In the triode region, \( V_{DS} < V_{GS} - V_t \). To make this condition valid, a bias voltage of 2.75V is placed on the gate of the NFET switch during its on-state. Second, a large-value resistor is inserted on the gate of every switch to reduce the leakage loss through the gate in the off-state. Third, a parallel NFET is added between the base and ground to prevent the base of the transistor from floating during the off-state. The operation details of NFET switches are shown in Figure 2.24.

![Figure 2.23 NFET large signal characteristics [61].](image)

**Figure 2.23 NFET large signal characteristics [61].**

![Figure 2.24 NFET switch operation.](image)

**Figure 2.24 NFET switch operation.**
2.3.4 Bias Network

In order to minimize its effects on the power gain and linearity of the power amplifier, the bias network is optimized for the following aspects:

- **Constant voltage biasing.** As discussed in Section 2.1.1, constant voltage biasing is desired from the standpoint of efficiency and linearity. The large inductor is replaced with an op-amp circuit (buffer) to save the chip area and cost. By adopting the topology as shown in Figure 2.25, the output impedance of the buffer is low at low frequencies, but high at the RF frequency.

- **Minimize the RF power loss due to the current leakage.** The output impedance of the buffer is so high at the RF frequency that the RF power leakage into the bias network is negligible.

- **Terminate dc or low-frequency components.** It has been proven [63] that the termination of dc or sub-harmonic components ($\Delta \omega = \omega_2 - \omega_1$, where $\omega_1$ and $\omega_2$ are two input frequencies) at the base of transistors is beneficial to the linearity of the PA. For instance, if there is no sub-harmonic termination, the sub-harmonic component ($\Delta \omega$) will mix with one fundamental frequency ($\omega_1$) to generate a third-order intermodulation product ($2\omega_1 - \omega_2$), and consequently degrade the linearity of the PA.

In summary, the bias network in Figure 2.25 provides constant voltage biasing to the base of the power amplifier and also terminates the sub-harmonic ($\Delta \omega$).
frequency at the input. In order to effectively terminate the sub-harmonic component, the buffer needs to satisfy certain bandwidth requirement. Since the channel bandwidth for WCDMA handset power amplifiers is 3.84MHz, the minimum bandwidth of the bias network should be greater than 3.84MHz. Using the feedback principle, we have

\[
Z_{\text{bias}}(\Delta \omega) \approx \frac{1}{g_{m,M3}(1 + g_{m,Q2}r_{o,Q2}/2)}
\]  

(2.9)

where \(g_{m,Q2}r_{o,Q2}/2\) is the approximate loop (A-Q2-B-M3-A) gain at \(\Delta \omega\). Since \(g_{m,Q2}r_{o,Q2}/2 \gg 1\), \(Z_{\text{bias}}(\Delta \omega) \ll 1/g_{m,M3}\), i.e. very close to zero. Based on simulations, \(Z_{\text{bias}}(\Delta \omega) \approx 0\) for \(\Delta \omega \leq 5\text{MHz}\).

Figure 2.25 Schematic of power amplifier with bias network.

2.3.5 Matching Networks

In PA design, input and output matching networks are designed according to different matching principles. To achieve maximum power gain, the input matching
network (IMN) adopts impedance matching, in which the source impedance is conjugately matched to the input impedance of transistors; whereas the output matching network (OMN) utilizes power matching to deliver maximum output power and therefore the load impedance is matched to the load-line impedance instead of the output impedance of transistors. From (2.7), we have

$$R_{\text{opt}} = \frac{V_p}{I_p} = \frac{V_{ce,\text{max}} - V_{ce,\text{min}}}{2 \cdot I_p} = 6\Omega$$  \hspace{1cm} (2.10)$$

In practice, a lower load-line impedance of 4Ω was selected to generate more output power. After compensating some power loss in the practical OMN, the desired maximum output power can be obtained at the load impedance. Our measurement results validated this assumption.

![Output matching network schematic.](image)

Since the load-line impedance is small compared to 50Ω, the accuracy of the output matching is very sensitive to parasitics. As a result, it is necessary to include all parasitics when designing the OMN. As shown in Figure 2.26, \(C_{\text{bondpad}}\) and \(C_{\text{lead}}\) represent parasitic capacitances due to bond-pad and lead frame respectively, and \(L_{\text{bondwire}}\)
and $L_{trace}$ are parasitic inductances due to bond-wire and soldering trace respectively, and $L_\infty$ is the choke inductor connected to the collector bias voltage. $Z_{opt} (R_{opt} + jX_{opt})$ is the optimum load impedance, whose real part $R_{opt}$ is the load-line resistance and imaginary part $X_{opt}$ is used to cancel out the output capacitance of transistors. Based on simulations, $X_{opt} (<0.2\,\Omega)$ is ignorable compared to $R_{opt}$ (4\,\Omega) and therefore $Z_{opt}$ can be replaced with $R_{opt}$. The typical parameters of parasitics are: $C_{bondpad} = 3\times75\,\text{fF}$, $L_{bondwire} = 0.9\,\text{nH}$, $C_{lead} = 0.3\,\text{pF}$, and $L_{trace} = 0.2\,\text{nH}$. The values of $C_{mo}$ and $L_{mo}$ can be found out through the following calculation procedure.

**Figure 2.27 3-stage output matching network.**

OMN in Figure 2.26 is analyzed as a 3-stage matching network in Figure 2.27. Starting from the left stage, it is straightforward to obtain that

\[
Z_{opt} = \frac{(Z_1 + j\omega L_{bondwire}) \cdot \frac{1}{j\omega C_{bondpad}}}{Z_1 + j\omega L_{bondwire} + \frac{1}{j\omega C_{bondpad}}} \quad (2.11)
\]

which results in
\[ Z_1 = \frac{j\omega L_{bondwire} \cdot (1 - Z_{opt} \cdot j\omega C_{bondpad}) - Z_{opt}}{Z_{opt} \cdot j\omega C_{bondpad} - 1} \]  

(2.12)

Considering the similarity between the first and second matching stages, the same procedure can be applied to obtain \( Z_2 \). Therefore,

\[ Z_2 = \frac{j\omega L_{trace} \cdot (1 - Z_1 \cdot j\omega C_{lead}) - Z_1}{Z_1 \cdot j\omega C_{lead} - 1} \]  

(2.13)

For the final matching stage, we have

\[ \text{real}\{Z_2\} = \frac{\omega^2 L_{mo}^2 R_L}{R_L^2 + \omega^2 L_{mo}^2} \]  

(2.14)

\[ \text{imag}\{Z_2\} = \frac{\omega L_{mo} R_L^2}{R_L^2 + \omega^2 L_{mo}^2} - \frac{1}{\omega C_{mo}} \]  

(2.15)

Combining (2.14) and (2.15), it is simple to obtain the initial values for \( L_{mo} \) and \( C_{mo} \), which always need to be tuned during lab testing. The final values are \( L_{mo} = 1.35 \text{nH} \) and \( C_{mo} = 2.2 \text{pF} \).

The input matching network (IMN) is implemented with a simple down-conversion topology as shown in Figure 2.28. Since the input impedance of transistors including the effect of bond-wires \( (Z_{in}) \) is 2.3-j2 \( \Omega \), it is straightforward to obtain that \( C_{mi} = 8 \text{pF} \) and \( L_{mi} = 0.9 \text{nH} \). Until now, we have derived all matching parameters. For clarification, the final schematic of the power amplifier with SDCB is shown in Figure 2.28, which includes four basic blocks: the bias network consisting of
a $\beta$ helper and a buffer, power transistors with NFET switches, input matching network, and output matching network.

![Diagram](image)

**Figure 2.28** Final schematic of the power amplifier with SDCB.

2.4 Design with Switched Dual Dynamic Biasing

In this section, the design considerations of the PA with switched dual dynamic biasing (SDDB) will be covered in detail. The PA is a two-stage topology and adopts SDDB in the output stage. Some design considerations that are the same as those in SDCB, including the selections of power device and NFET switches, the bias network, OMN and IMN, are omitted here for simplicity.

2.4.1 Power Step Selection

As discussed in Section 2.3.1, there is a tradeoff between efficiency and circuit complexity in the selection of power step. According to Table 2.2, the efficiency
difference between 1-step and 2-step SDDB is 1.2%. However, the circuit complexity of 2-step SDDB increases dramatically. Therefore, 1-step SDDB was finally selected.

Table 2.2 Average power efficiency comparison of SDDB in an output stage.

<table>
<thead>
<tr>
<th>Bias Type</th>
<th>Average Power Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class AB with CVB</td>
<td>2.44%</td>
</tr>
<tr>
<td>Class AB with 1-step SDDB</td>
<td>7.75%</td>
</tr>
<tr>
<td>Class AB with 2-step SDDB</td>
<td>8.94%</td>
</tr>
</tbody>
</table>

2.4.2 Output Stage Design

Since the critical specifications (such as efficiency and linearity) are mostly determined by the output stage, it is the key part in PA design. In practice, the logical way to design a multiple-stage power amplifier is to begin with the output stage and then trace back to the first stage. As shown in Figure 2.19, the output stage with SDDB consists of high-power and low-power groups. Both groups adopt the single-ended common-emitter topology and are biased in Class AB. The quiescent bias currents are 110mA and 22mA for high-power and low-power groups respectively. Since there are 100 and 20 devices in high-power and low-power groups respectively, their current densities are the same so as to keep the power gain constant. The operation principle has been illustrated in Section 2.2.2.
One key issue during the design of the output stage is stability. There are several steps to check the stability of power amplifiers. First of all, the following stability criteria are applied to check the small-signal stability of amplifiers [64]:

\[
K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|}
\]  

(2.16)

\[
B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2
\]  

(2.17)

where \( \Delta = S_{11}S_{22} - S_{12}S_{21} \). If \( K > 1 \) and \( B_1 > 0 \) simultaneously, then the amplifier is unconditionally stable. In our designs, a series R-C feedback network is added between the collector and the base to achieve more phase margin through the pole-splitting [61] and make the circuit stable. Besides the small-signal stability, the large-signal stability has to be checked in high-power amplifiers with large-signal transient simulations (e.g. two-tone test at maximum output power). Both simulation and experimental results have verified that the designs in this dissertation are large-signal stable.

In high-power HBT amplifiers, there also exists the thermal stability issue, which is caused by the positive feedback between transistor current and temperature [65]-[66]. With a positive temperature coefficient of \( V_{BE} \), the collector current in HBT transistors increases in response to a rise in temperature if \( V_{BE} \) is held constant. Assuming that one transistor happens to become a little hotter than other transistors in parallel, the collector current in that transistor will increase. The more current it carries, the hotter the transistor becomes. This thermo-electrical positive feedback
may result in severe device destruction. To prevent the thermal issue, some emitter ballasting resistance $R_E$ is always necessary in high-power HBT amplifiers. From [65], a lower bound on $R_E$ is given by

$$R_E \geq \frac{kT}{qI_C} \left[ (0.05I_C)\theta_{th} V_C - 1 \right]$$  \hspace{1cm} (2.18)$$

where $I_C$ and $V_C$ are the dc collector bias current and voltage respectively, and $\theta_{th} = \Delta T/I_C V_C$ is the thermal resistance. Assuming that $I_C = 110 mA$ and $V_C = 3 V$, $R_E \geq 1.2 \ \Omega$. Note that $R_E$ is inserted into the emitters of individual transistors, instead of the common path of all transistors. The details are demonstrated in Figure 2.29.

![Emitter ballasting resistor configuration](image)

Figure 2.29 Emitter ballasting resistor configuration.

### 2.4.3 Driver Stage Design

Compared to the output stage, the design of the driver stage is straightforward.

To meet the specification of the power gain for a 2-stage PA (typically > 21dB), the
driver stage needs to provide enough power gain (> 13dB), since the power gain of the output stage is rather low (< 8dB). Furthermore, the driver stage should provide a linearly-amplified signal to the input of the output stage. This requires the driver stage to be biased nearly in Class A.

![Simplified schematic of the driver stage including the inter-stage matching (CVB: constant voltage biasing).](image)

The simplified schematic of the driver stage is shown in Figure 2.30. Assuming that the maximum output power at the output stage ($P_{\text{max, out}}$) is 27dBm and the power gain of the output stage ($G_{\text{out}}$) is 8dB, the maximum output power of the driver stage ($P_{\text{max,drv}} = P_{\text{max, out}} - G_{\text{out}}$) is 19dBm. Since the maximum output power ratio between the driver stage and the output stage is equal to $P_{\text{max,drv}}/P_{\text{max, out}} \approx 0.16$ and the device number of the output stage is 100, the device number of the driver stage is found to be 20. To bias the driver stage nearly in Class A, the quiescent bias collector current of the driver stage is set to be 30mA.
2.4.4 Matching Networks

The inter-stage matching network consists of $L_{is}$ and $C_{is}$ as depicted in Figure 2.30. The equivalent circuit of the inter-stage matching network is shown in Figure 2.31, where $Z_{out\_drv}$ is the equivalent output impedance of the driver stage and $Z_{in\_out}$ is the equivalent input impedance of the output stage. Based on simulations, $Z_{out\_drv} = 32.8 - j24.5\,\Omega$. Note that $Z_{in\_out}$ has two different values (2.77 - j0.76$\,\Omega$ and 3.32 - j3.89$\,\Omega$ at 1.95GHz) for high-power and low-power operations respectively.

As discussed previously in Section 2.2.2, the inter-stage matching was tuned to maintain a small gain difference between high-power and low-power operations. The optimal parameters are found to be $L_{is} = 1.25\,\text{nH}$ and $C_{is} = 9\,\text{pF}$, when $G_{T\_hiP} = 21.1\,\text{dB}$ and $G_{T\_lowP} = 19.7\,\text{dB}$ from layout simulations including extracted parasitics. The corresponding gain variation is less than 1.5$\,\text{dB}$.

To achieve the desired inter-stage matching, the effect of bond-wire ($L_{bondwire}$) connecting $L_{is}$ to external $V_{cc}$ must be eliminated in practical implementations. Two
Effective termination methods are shown in Figure 2.32, where the internal on-chip termination capacitor \( C_{\text{gnd1}} = 40 \text{pF} \) and the external termination capacitor \( C_{\text{gnd2}} = \frac{1}{\omega_{RF}^2 \cdot L_{\text{bondwire}}} \).

![Figure 2.32 Inter-stage matching terminations. (a) Internal termination. (b) External termination.](image)

The OMN is the same as that in SDCB and the IMN is implemented with \( L_{mi} = 1.2 \text{nH} \) and \( C_{mi} = 7.3 \text{pF} \). The final circuit schematic of the two-stage PA with SDDB is shown in Figure 2.33.
Figure 2.33 Final schematic of the two-stage PA with SDDB.
2.5 Layout and Fabrication

In RF designs, circuit performance is always affected to some extent by layout parasitics when implementing the schematic in the layout. Since RF power amplifiers require high output power with high power gain, they are more sensitive to layout parasitics. As a result, layout practice is critical in RF PA design. This section begins with some basics of the IBM SiGe BiCMOS 6HP process, which is followed by layout details of the power transistor cell. Subsequently, some practical fabrication issues, which are required for successful fabrications, are discussed. Finally, the complete layouts for the designs of SDCB and SDDB will be presented.

2.5.1 IBM SiGe BiCMOS 6HP Technology

In IBM SiGe BiCMOS 6HP technology, there are different options for metal usage, resistors, capacitors, and inductors, which are selected with the following considerations.

- **Technology Metal Options**

IBM SiGe BiCMOS 6HP is a 0.25μm lithography technology, which supports 4, 5 or 6 levels of metal. To utilize the top metal (AM), the option of 6 levels was chosen for our designs. The details of metal levels are shown in Table 2.3, where the bottom metal (M1) is the thinnest metal level and upper metal levels are thicker with lower sheet resistance and higher current-carrying capability. Based on their electric characteristics, AM or MT are always used for RF signal paths to reduce the series
parasitic resistance of the routing line. For less parasitic-sensitive routing paths such as dc bias lines, lower metal levels may be used to conserve layout area.

Table 2.3 Characteristics of metal levels in IBM SiGe BiCMOS 6HP technology.

<table>
<thead>
<tr>
<th>Metal Level</th>
<th>Min. line/space (um)</th>
<th>Thickness (um)</th>
<th>Rs (Ω)</th>
<th>EM* (mA/um)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0.32/0.32</td>
<td>0.40</td>
<td>0.127</td>
<td>0.65</td>
</tr>
<tr>
<td>M2-M4</td>
<td>0.40/0.40</td>
<td>0.54</td>
<td>0.078</td>
<td>1.03</td>
</tr>
<tr>
<td>MT</td>
<td>0.40/0.40</td>
<td>0.66</td>
<td>0.050</td>
<td>1.48</td>
</tr>
<tr>
<td>AM</td>
<td>4.00/5.00</td>
<td>4.00</td>
<td>0.00725</td>
<td>6.17</td>
</tr>
</tbody>
</table>

*Electromigration current limits at 100C, 100K power-on-hours and 100% duty cycle.

- **Resistors**

There are two types of resistors in IBM SiGe BiCMOS 6HP technology: polysilicon and silicon resistors. As shown in Table 2.4, polysilicon resistors are typically preferred over silicon resistors because of their low temperature and voltage sensitivities. As a result, polysilicon resistors were utilized in our designs, except for the emitter ballasting resistors, whose resistance is quite small. For polysilicon resistors, oppcres (high-resistance poly resistor) and oppcres (low-resistance poly resistor) are placed over substrate with the deep trench (DT) option for capacitance reduction; nspcpres (high-resistance poly resistor) and nspcres (low-resistance poly resistor) are placed over the NS groundplane for noise shielding. The cross-sections of oppcres and oppcres resistors are shown in Figure 2.34 for illustration. Since
noise is not a big concern in PA design, \textit{oppcpres} and \textit{oppcres} were utilized in our designs to reduce the parasitic capacitance.

\textbf{Table 2.4 Resistor comparison in IBM SiGe BiCMOS 6HP technology.}

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Rs ((\Omega/))</th>
<th>Rs Tol (%)</th>
<th>TCR (%/C)</th>
<th>VCR (%/V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>oppcpres nspcpres</td>
<td>p- polysilicon</td>
<td>3600</td>
<td>25</td>
<td>-0.235</td>
<td>0.065</td>
</tr>
<tr>
<td>oppcres nscpres</td>
<td>p+ polysilicon</td>
<td>210</td>
<td>20</td>
<td>0.02</td>
<td>~0</td>
</tr>
<tr>
<td>oppdres</td>
<td>Si p+ resistor</td>
<td>100</td>
<td>10</td>
<td>0.155</td>
<td>0.065</td>
</tr>
<tr>
<td>opndres</td>
<td>Si n+ resistor</td>
<td>63</td>
<td>10</td>
<td>0.15</td>
<td>0.115</td>
</tr>
</tbody>
</table>

\textbf{Figure 2.34 Cross-sections of oppcpres and oppcres resistors.}

- Capacitors

IBM SiGe BiCMOS 6HP provides two types of capacitors: \textit{MOSCAP33} (poly to silicon) and \textit{MIM} (metal-insulator-metal) capacitors. The cross-sections of \textit{MOSCAP33} and \textit{MIM} capacitors are shown in Figure 2.35 for illustration.
**MOSCAP33** capacitors provide a higher per-unit-area capacitance (3.1fF/µm²) than **MIM** capacitors (0.7fF/µm² for a single mask and 1.4fF/µm² for 2 stacked masks). The tolerance on the per-unit-area capacitance is 15% for both types of capacitors. Since **MOSCAP33** capacitors exhibit voltage dependent characteristics due to depletion effects, **MIM** capacitors were used in our designs to achieve less voltage-dependent circuits.

![Cross-sections of MOSCAP33 and MIM capacitors.](image)

**Figure 2.35 Cross-sections of MOSCAP33 and MIM capacitors.**

- **Inductors**

  In IBM SiGe BiCMOS 6HP, AM octagonal inductors are typically utilized for RF designs. There are two options for groundplane in AM octagonal inductors: DT and PC (polysilicon). As shown in Figure 2.36, the PC groundplane has four segments with slits to reduce eddy currents. Furthermore, it has less substrate losses compared to the DT lattice. Therefore, it can achieve higher Q and therefore was selected in our designs.
2.5.2 Power Transistor Cell

BiCMOS6HP supports two types of NPN transistors: high $f_T$ ($npnrf$) and high breakdown ($npnhb$) NPNs. The detailed comparison between these two NPN transistors is shown in Table 2.5. To accommodate higher voltage swing at the output, high breakdown NPNs were employed in our designs. The layout topology of a single transistor is shown in Figure 2.37, where each emitter has 2 stripes and each stripe is $24 \mu m \times 0.44 \mu m$. 
Table 2.5 Parameter comparison of NPN transistors.

<table>
<thead>
<tr>
<th>Parameter*</th>
<th>high ( f_T )</th>
<th>high breakdown</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_T ) (GHz)</td>
<td>47</td>
<td>27</td>
</tr>
<tr>
<td>Peak ( f_T ) current (mA/\mu m^2)</td>
<td>1.5</td>
<td>0.45</td>
</tr>
<tr>
<td>( BV_{CBO} ) (V)</td>
<td>10.5</td>
<td>14.0</td>
</tr>
<tr>
<td>( BV_{CEO} ) (V)</td>
<td>3.0</td>
<td>5.2</td>
</tr>
</tbody>
</table>

*The emitter width is 0.44 \( \mu m \).

Figure 2.37 Layout topology of single transistor.

There are two fundamental considerations in the layout of the whole power cell. To save chip area and cost, it is always desired to layout the whole block of transistors in the form of square. At the same time, the routing line from the input to the output of every single transistor should be identical to minimize the phase shift of individual output signals; otherwise the power gain of the PA will be degraded. The layout of the high-power group with 100 transistors is analyzed here for illustration. Assume that there are \( N \) rows and \( M \) columns, where \( M = 100/N \). Two typical cases are demonstrated in Figure 2.38.
In case (a), there is almost no phase difference between individual RF signal paths of transistors \( npn1 \) and \( npn100 \). Unfortunately, all transistors are aligned along a straight line, which results a large die size. In contrast, the layout in case (b) is formed close to a square. However, there is certain phase difference between individual transistors (e.g. \( npn1 \) and \( npn21 \)), which will degrade the power gain of the power amplifier. To calculate the gain loss, the phase delay of individual transistors are estimated by
\( \phi_d = \omega_{RF} t_d \)  

(2.19)

where \( \omega_{RF} \) is the RF signal frequency and \( t_d \) is the signal-traveling time delay of individual transistors, which refers to the reference transistor \( npn21 \). Furthermore, 

\[ t_d = \sqrt{L_pC_p} \]

where \( L_p \) and \( C_p \) are the parasitic inductance and capacitance respectively of individual signal paths, which are obtained by subtracting the reference signal path from individual signal paths.

Subsequently, the gain loss can be found by comparing the amplitude of the composite output signal with all phase delays of individual signals to its counterpart of the ideal output signal without any phase delay of individual signals. For instance, assuming that there are two individual signals \( A \cos(\omega_{RF}t) \) (the reference signal) and \( A \cos(\omega_{RF}t - \phi_d) \) (the delayed signal), the composite and ideal output signals are \( 2A \cos(\phi_d/2) \cos(\omega_{RF}t - \phi_d/2) \) and \( 2A \cos(\omega_{RF}t) \) respectively. Hence, the gain loss is given by

\[
G_{loss} = 10 \log \left( \frac{(2A \cos(\phi_d/2))^2}{(2A)^2} \right) \quad (2.20)
\]

Similarly, the total gain loss in case (b) is found to be roughly 0.32dB. Considering its area-saving feature, the topology in case (b) was selected in our designs.
2.5.3 Fabrication Requirements

To avoid device destruction during the silicon fabrication and experimental operation, two simple but critical protection structures must be inserted into the layout, as illustrated in the following.

- **Floating Gate Tie Down**

Thin gate oxides are subjected to the charging damage during wafer processing, as shown in Figure 2.39. If the gate is charged to a sufficient potential, the gate oxide will break down. The gate oxide damage results in a degradation of device reliability (for instance, transistor transconductance and threshold voltage shift over time). This issue can be eliminated by providing an alternative discharge path from the gate to the substrate. A diode to substrate is an effective means to prevent charge build up across the gate dielectric.

![Figure 2.39 Gate tie down diagram.](image-url)
- Electro Static Discharge

All RF pins need an Electro Static Discharge (ESD) protection. Effective ESD protection confines the discharge current to certain paths that are designed to handle the current levels without damage. Without degrading the circuit performance, a double-diode structure was utilized to protect all RF I/O pins in our designs, as shown in Figure 2.40(a). Figure 2.40(b) shows a stacked diode configuration, which is used on the PA output, where the signal voltage may be higher than the power supply $V_{CC}$.

![Diagram of Bipolar-based diode ESD structure](image)

Figure 2.40 Bipolar-based diode ESD structure. (a) typical configuration. (b) stacked configuration.

2.5.4 Final Layouts

The final layout of a single-stage PA with SDCB is shown in Figure 2.41. The die area is 0.9mm x 1.2mm. Its experimental results will be shown in Chapter 5.
The final layout of a two-stage PA with SDDB is shown in Figure 2.42. The die area is 1mm x 1.8mm. Its experimental results will be shown in Chapter 5.
2.6 Summary

With the advancement of wireless communication systems, stringent efficiency requirements are posed on RF handset power amplifiers. The research on switched dynamic biasing techniques intends to achieve high average power efficiency and keep power gain constant for WCDMA handset power amplifiers. The prototype power amplifiers have been implemented in an advanced SiGe BiCMOS technology to demonstrate the effectiveness of the proposed techniques.

This chapter, in part or in full, is a reprint of the materials in 2004 IEEE Radio Frequency Integrated Circuits Symposium, 2005 IEEE Radio Frequency Integrated Circuits Symposium, and 2005 IEEE Transactions on Microwave Theory and Techniques. The author of this dissertation was the primary researcher in these publications.
CHAPTER 3
Linearity Analysis of SiGe HBT Power Amplifiers

With the advancement of wireless communications systems, more stringent requirements are imposed on the linearity as well as efficiency, because multi-carrier modulations are typically employed in these systems for both high data rate and efficient spectrum utilization [26]. A multi-carrier transmit signal results in a high peak-to-average power ratio (PAR) [25], which is defined as the ratio between the peak envelope power and the overall mean power. Assuming a 1Ω resistive load, the PAR is given by

\[
PAR = \frac{v_{pk}^2}{\sum_{k=1}^{N} v_k^2}
\]

in which \( v_{pk} \) is the peak voltage amplitude of the multi-carrier envelope signal and \( v_k \) denotes the voltage amplitude of \( N \) individual carriers. Even if the voltage amplitudes of individual carriers are small, the peak voltage amplitude in the composite signal can be high when the individual carriers of the transmit signal add constructively [67]. Consequently, the corresponding PAR is high. Therefore, it is necessary to design
linear transmitters in these systems to accommodate high PAR signals. Since the power amplifier is the main nonlinear device in the transmitter, it is crucial to design high-linearity power amplifiers in advanced wireless handsets.

In this chapter, the linearity of SiGe HBT power amplifiers will be analyzed. To begin with, different linearity analysis methods will be reviewed. This is to illustrate the necessity of employing power-dependent coefficient Volterra (PDCV) analysis in high-power amplifiers. Then, we will introduce the principle of PDCV analysis. Subsequently, a SiGe HBT power amplifier will be analyzed with PDCV to identify the dominant sources of nonlinearity and provide directions to design linear SiGe HBT high-power amplifiers. Furthermore, a SiGe HBT amplifier with envelope signal injection will be analyzed with PDCV as a linearization technique for RF amplifiers. Both simulation and experimental results will be presented for verification, which is followed by concluding remarks.

### 3.1 Linearity Analysis Methodologies

Analysis techniques of nonlinear circuits include time domain, harmonic balance, power series, and Volterra series. Time domain methods utilize numerical integration and calculate the instantaneous value of the output from the instantaneous value of the input. In general, time domain methods are adopted in computer-aided-design programs such as SPICE [68]. However, the lack of the frequency domain information and the long simulation time constrain their further applications in RF and microwave circuits [69].
The harmonic balance method [69] combines time domain analysis of the nonlinear devices with frequency domain analysis of the linear elements. The currents into the linear elements are calculated by means of a frequency domain linear analysis. The currents into the nonlinear devices are calculated in the time domain and then transformed to the frequency domain with a generalized Fourier transform technique. Since most of the analysis is performed in the frequency domain, the effects of individual tones can be distinguished readily. Consequently, harmonic balance is well suited for analyzing RF and microwave circuits, which are most naturally handled in the frequency domain. Considering the complexity of the Fourier transformation, however, harmonic balance is generally used in commercial computer analysis programs, not for hand analysis.

Power series and Volterra series are the two methods widely used in the nonlinearity hand analysis, since their frequency domain nature offers great promise for multi-tone analysis and useful insights into the nonlinearity mechanism of microwave circuits. The details of these two methods will be discussed as follows.

3.1.1 Power Series

A power series is the simplest method for the nonlinearity analysis of microwave circuits. In 1969 Sea [70] proposed to solve the output variable \( y \) of nonlinear systems in terms of the input variable \( x \) as

\[
y = \sum_{k=0}^{\infty} a_k x^k
\]
where $a_k$ is real and $x$ is the input and equal to a sum of $N$ sinusoids, i.e.

$$x = \sum_{i=1}^{N} A_i \cos(\omega t) \quad (3.3)$$

For a nonlinear system with two input tones $A_1 \cos(\omega t)$ and $A_2 \cos(\omega t)$, the output can be derived from (3.2) as

$$y = a_1[A_1 \cos(\omega t) + A_2 \cos(\omega t)] + a_2[A_1 \cos(\omega t) + A_2 \cos(\omega t)]^2 + a_1[A_1 \cos(\omega t) + A_2 \cos(\omega t)]^3 + ... \quad (3.4)$$

After the series expansion, it is straightforward to obtain that the coefficient at the frequency $2\omega_1 - \omega_2$ is $0.75a_3A_1^2A_2$. Consequently, the third-order intermodulation ratio (IMR$_3$)$^1$ is found to be $0.75a_3A_1A_2 / a_1$.

Clearly, the above approach only analyzed the amplitude nonlinearity (AM-AM), and there was no inclusion of phase nonlinearity (AM-PM). As a result, a power series is generally used in low frequency, low-power, or Class A amplifiers, in which phase nonlinearity is negligible. Even though some improved power series models including both AM-AM and AM-PM nonlinearity were proposed [71]-[72], the calculation complexity associated with those advanced models makes them unsuitable for hand analysis for nonlinear circuits.

$^1$ The output voltage ratio between the third-order intermodulation product (at $2\omega_1 - \omega_2$) and the fundamental tone (at $\omega_1$).
3.1.2 Volterra Series

Volterra series has been widely applied in the nonlinearity analysis of RF and microwave circuits [73]-[78], since it can address both amplitude and phase nonlinearities. In principle, the output time response \( y(t) \) of a nonlinear system with memory can be expressed as

\[
y(t) = \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(u_1, \ldots, u_n) \prod_{r=1}^{n} x(t-u_r) \]

\[
= \int_{-\infty}^{\infty} h_1(u_1) x(t-u_1) du_1
\]

\[
+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(u_1, u_2) x(t-u_1) x(t-u_2) du_1 du_2
\]

\[
+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(u_1, u_2, u_3) x(t-u_1) x(t-u_2) x(t-u_3) du_1 du_2 du_3
\]

\[
+ \cdots
\]

where \( h_n(u_1, \ldots, u_n) \) is the n-th order impulse response of the system and a symmetrical function of \( u_i \), which means that \( h_n(u_1, \ldots, u_n) \) has the same value regardless of the permutation of \( u_1, \ldots, u_n \). The corresponding output spectrum can be calculated through Fourier transform as

\[
Y(f) = H_1(f)X(f)
\]

\[
+ \int_{-\infty}^{\infty} df_1 H_2(f_1, f - f_1) X(f_1) X(f - f_1)
\]

\[
+ \int_{-\infty}^{\infty} df_2 \int_{-\infty}^{\infty} df_1 H_3(f_1, f_2, f - f_1 - f_2) X(f_1) X(f_2) X(f - f_1 - f_2)
\]
in which \( H_n(f_1, f_2, \ldots, f_n) \) is the Fourier transform of \( h_n(u_1, \ldots, u_n) \) and the n-th order Volterra kernel. Therefore, the output spectrum is the sum of the contributions from each order of the Volterra kernel, as shown in Figure 3.1.

\[
+ \cdots
\]

Figure 3.1 Frequency domain model of Volterra series.

To keep the analysis tractable and preserve its accuracy, Volterra series is often truncated to third-order to analyze weakly nonlinear RF amplifiers (such as Class A PAs) [74]. On the other hand, high efficiency CDMA power amplifiers typically operate in the Class AB mode and exhibit strong nonlinearities. This renders the traditional Volterra analysis, truncated to third-order, inadequate and necessitates the inclusion of higher order terms, vastly increasing the complexity of the Volterra analysis. To keep the analysis manageable, and accurately predict the intermodulation
distortion, a power-dependent coefficient Volterra technique [14] will be used in our analysis.

3.2 Power-Dependent Coefficient Volterra Analysis

In the power-dependent coefficient Volterra (PDCV) analysis, the nonlinear elements are linearized and approximated in finite orders. For instance, the ac charge across a nonlinear capacitor is approximated in a third-order series as

\[ Q_c = C_1 v_{\text{contr}} + K_{2c_1} v_{\text{contr}}^2 + K_{3c_1} v_{\text{contr}}^3 \]  \hspace{1cm} (3.8)

where \( v_{\text{contr}} \) is the small-signal controlling voltage over the capacitor and \( C_1, \ K_{2c_1}, \) and \( K_{3c_1} \) are the first, second, and third-order nonlinearity coefficients respectively.

The total equivalent capacitor can be derived as

\[ C_{\text{tot}} = \frac{dQ_c}{dv_{\text{contr}}} \]  \hspace{1cm} (3.9)

\[ = C_1 + 2K_{2c_1} v_{\text{contr}} + 3K_{3c_1} v_{\text{contr}}^2 \]  \hspace{1cm} (3.10)

Unlike the conventional Volterra analysis, all nonlinearity coefficients in the PDCV analysis vary in response to changes in the RF signal power. As a result, the PDCV analysis can analyze strongly nonlinear circuits within limited orders (e.g. third-order in the above nonlinear capacitor). In contrary, the conventional Volterra analysis assumes that all nonlinearity coefficients are fixed under certain bias condition. To model a strong nonlinearity, the series order in the conventional
Volterra analysis has to be very high, rendering the analysis too complex and intractable. In general, a higher order nonlinearity in the Volterra analysis can be approximated by a power dependent lower order nonlinearity in the PDCV analysis. This approximation will bring certain discrepancy into the analysis and hence the validity of the analysis has to be verified with simulation and experimental results.

Figure 3.2 illustrates the general operation flow diagram of the PDCV analysis. The Volterra calculation begins with the proper selection of nonlinearity sources. Too few nonlinearity sources render the accuracy of the calculation poor, whereas too many nonlinearity sources make the calculation too complex to perform. Therefore, a tradeoff needs to be made between the complexity and the accuracy of the calculation.

![Flow diagram of PDCV analysis.](image)

After selecting the proper nonlinearity sources, the nonlinearity coefficients will be extracted. Note that these nonlinearity coefficients depend on RF signal power as
well as quiescent bias point. Therefore, their values have to be determined under large-signal conditions, as illustrated in the later examples.

With the extracted nonlinearity coefficients, the Volterra responses of the circuit can be derived using the “nonlinear currents” method [79]-[80]. Compared to the conventional Volterra calculation, the “nonlinear currents” method has the advantage of avoiding the complicated calculation of Volterra kernels. To verify its effectiveness, Volterra calculation results will be compared to simulation and experimental results. After those results match each other, the PDCV model can be safely employed to provide helpful insights into the nonlinearity mechanisms of the circuit.

### 3.3 Analysis of SiGe HBT High-Power Amplifiers

The SiGe HBT power amplifier depicted in Figure 2.15 will be analyzed with the PDCV technique. By substituting the small-signal equivalent circuit of the HBT transistor in Figure 2.16 into the circuit, the equivalent circuit of the power amplifier is shown in Figure 3.3, where the effect of $r_\pi$ can be ignored at RF because the impedance of $r_\pi$ is much larger than that of $C_\pi$. 
Figure 3.3 Equivalent circuit of the SiGe HBT power amplifier.

Based on the above equivalent circuit, the nonlinear circuit of the SiGe HBT for Volterra calculation is derived in Figure 3.4, where $Z_S$ represents the source impedance (including the effects of the input matching network, bias network and base resistance of transistors), $Z_L$ is the load impedance (including the effects of the output matching network and output capacitance of transistors), and $Z_E$ is the emitter impedance (including the bond-wire inductance and the emitter parasitic resistance).

Figure 3.4 Nonlinear circuit of SiGe HBT for Volterra series calculation.
In our analysis, $C_{\pi}$ and $i_C$ are the main sources of nonlinearity and the remaining elements are assumed to be linear. The nonlinear elements depend on both quiescent bias point and RF signal power. Therefore, their nonlinearity coefficients have to be determined under large-signal conditions. Our simulations showed that this is an adequate model to predict the nonlinearities in our power amplifier.

### 3.3.1 Nonlinearity Coefficient Extraction

The nonlinear base-emitter capacitance ($C_{\pi}$) was obtained from the simulations under large-signal conditions\(^2\), as shown in Figure 3.5. Nonlinearity coefficients were obtained through the curve fitting in MATLAB. When the fitting curve with minimum order fits the simulation data with the satisfactory accuracy\(^3\), the corresponding order is taken as the order of nonlinearity coefficients.

---

\(^2\) The effect of the linear base-collector capacitance $C_{bc}$ is removed by varying the collector bias voltage in the simulation to emulate the collector voltage fluctuation when the RF input signal changes.

\(^3\) If the theoretical analysis with the presumed minimum-order nonlinearity matches both simulation and experiment results within a variation of 3dB [81], the analysis will be taken with the satisfactory accuracy. Otherwise, the analysis will be iterated with a higher order nonlinearity.
Figure 3.5 Simulation data of $C_π$ and its third-order and fifth-order fitting curve.

As shown in Figure 3.5, both the third-order fitting curve and the fifth-order fitting curve match the simulation data closely over the whole excursion of the base-emitter voltage. To simplify our analysis, third-order is first assumed to be the nonlinearity order of $C_π$. The validity of this assumption will be verified through the Volterra calculation results in comparison with both simulation and experimental counterparts. Hence, the nonlinear $C_π$ can be expressed as

$$C_π = C_1 + 2K_{2C_π}v_{be} + 3K_{3C_π}v_{be}$$  \hspace{1cm} (3.11)

The RF input signal determines the excursion range of the base-emitter voltage. Nonlinearity coefficients $C_1$, $K_{2C_π}$ and $K_{3C_π}$ are decided by the RF input signal as
well as the quiescent bias condition. The resulting coefficients\(^4\) are plotted in Figure 3.6, where \(K_{3C_x}\) changes more dramatically than \(C_1\) over the whole range of the output power. Therefore, \(K_{3C_x}\) is the major coefficient characterizing the nonlinearity behavior of \(C_x\), which will be illustrated in the later Volterra calculation.

![Figure 3.6 Nonlinearity coefficients \(C_1\) and \(K_{3C_x}\) \(I_{CQ} = 110\)mA.](image)

Based on the first-order device approximation \([79]\),

\[
C_1 = \tau_p g_m \tag{3.12}
\]

\[
K_{3C_x} = \tau_p g_m / 6V_T^2 \tag{3.13}
\]

where \(\tau_p\) is the transit time (\(\cong 10\)ps) and \(V_T\) is the thermal voltage. As the output power approaches 0dBm, the extracted nonlinearity coefficient \(C_1\) is very close to its

\(^4\) To simplify the Volterra analysis, even-order harmonics are assumed to be terminated at the base of the transistors. The validity of this assumption is to be verified by experiment. Therefore, \(K_{3C_x}\) is not
counterpart in (3.12), i.e. $4.4 \times 10^{-11}$ F; nevertheless there is a discrepancy between the extracted $K_{3C_e}$ and its counterpart in (3.13). Since we are using the limited third-order fitting curve to match the simulation data over the whole power range, there are some discrepancies between these two curves, as shown in Figure 3.5. This consequently results in the $K_{3C_e}$ discrepancy. The validity of the extracted coefficients will be verified later through the experimental verification.

Similarly, the large-signal nonlinear collector current $i_C$ can be obtained from the simulations. As shown in Figure 3.7, since the third-order fitting curve of the collector current versus base-emitter voltage matches the simulation data closely over the whole excursion of the base-emitter voltage, third-order is initially assumed to be the nonlinearity order of $i_C$. The validity of this assumption will be verified through the Volterra calculation results in comparison with both simulation and experimental counterparts. Hence, the nonlinear $i_C$ can be expressed as

$$i_C = g_m v_{be} + K_{2g_m} v_{be}^2 + K_{3g_m} v_{be}^3$$

(3.14)

in which $g_m$, $K_{2g_m}$, and $K_{3g_m}$ can be determined from the excursion of the base-emitter voltage together with the quiescent bias condition.

---

5 The criterion of the satisfactory accuracy for the nonlinear collector current $i_C$ is the same as that for the nonlinear base-emitter capacitance $C_e$. 
Figure 3.7 Simulation data of $i_C$ and its third-order curve.

The resulting coefficients are shown in Figure 3.8, where $g_m$ remains nearly constant and $K_{3g_m}$ changes dramatically over the whole range of the output power. Therefore, $K_{3g_m}$ is the major coefficient characterizing the nonlinearity behavior of $i_C$, which will be illustrated in the later Volterra calculation.

Figure 3.8 Nonlinearity coefficients $g_m$ and $K_{3g_m}$. $I_{CQ} = 110mA$. 
Based on the first-order device approximation [79],

\[ g_m = \frac{I_C}{V_T} \quad (3.15) \]

\[ K_{3g_m} = \frac{g_m}{6V_T^2} \quad (3.16) \]

where \( I_C \) is the collector bias current. As the output power approaches 0dBm, the extracted nonlinearity coefficient \( g_m \) is very close to its counterpart in (3.15), i.e. 4.4A/V; nevertheless there is a discrepancy between the extracted \( K_{3g_m} \) and its counterpart in (3.16). Since we are using the limited third-order fitting curve to match the simulation data over the whole power range, there are some discrepancies between these two curves, as shown in Figure 3.7. This consequently results in the \( K_{3g_m} \) discrepancy. The validity of the extracted coefficients will be verified later through the experimental verification.

### 3.3.2 Volterra Calculation of Nonlinear Responses

After extracting all the nonlinearity coefficients, the Volterra responses of the circuit can be derived using the “nonlinear currents” method [79]-[80]. The output third-order intermodulation ratio (IMR_3) is the objective of our analysis, as defined in terms of the ratio between the third-order intermodulation voltage and the fundamental voltage at the output node:

\[ \text{IMR}_3 = \frac{v_o(2\omega_1 - \omega_2)}{v_o(\omega_1)} \quad (3.17) \]
Figure 3.9 Block diagram of IMR₃ calculation.

The block diagram for calculating IMR₃ is demonstrated in Figure 3.9. In principle, higher order Volterra responses are derived iteratively with the formulas of nonlinear currents composed of lower order responses. The derivation details will be illustrated in the following description.

The fundamental responses of the collector voltage and the base-emitter voltage are derived through the small-signal analysis in Figure 3.10. Hence,

\[
v_c(\omega_t) = \frac{(-g_m + j\omega_t C_{bc} - j\omega^3_t C_c C_{bc} L_e)Z_L}{\det(\omega_t)} v_s
\]

(3.18)

\[
= -\frac{g_m Z_L}{\det(\omega_t)} v_s
\]

(3.19)
Figure 3.10 First-order equivalent circuit of SiGe HBT power amplifier.

\[ v_{be}(\omega_i) = \frac{1 + j\omega C_{be} Z_L}{\text{det}(\omega_i)} v_S \]  
(3.20)

\[ = A_{vbe}(\omega_i) \cdot v_S \]  
(3.21)

where

\[ \text{det}(\omega_i) = 1 - \omega^2 (C_1 C_{be} Z_E Z_L + C_1 C_{be} Z_E Z_S + C_1 C_{be} Z_L Z_S) \]
\[ + g_m Z_E + j\omega [C_1 Z_E + C_{be} Z_L + C_1 Z_S + C_{be} Z_S] \]
\[ + C_{be} g_m Z_E Z_L + C_{be} g_m Z_E Z_S + C_{be} g_m Z_L Z_S ] \]  
(3.22)

in which \(Z_E, Z_S\) and \(Z_L\) are evaluated at the frequency \(\omega_i\).

The response of the base-emitter voltage at \(-\omega_2\) is given by

\[ v_{be}(-\omega_2) = v_{be}^*(\omega_2) = \frac{1 - j\omega C_{be} Z_L^*}{\text{det}^*(\omega_2)} v_S \]  
(3.23)

\[ = A_{vbe}(-\omega_2) \cdot v_S \]  
(3.24)

in which \(Z_E, Z_S\) and \(Z_L\) are evaluated at the frequency \(\omega_2\).
Since $Z_S = 0$ at even-order harmonic frequencies (including $\omega_1 - \omega_2$ and $2\omega_1$), there is no need to compute the second-order responses. Now we can calculate the third-order intermodulation product at $2\omega_1 - \omega_2$. The equivalent circuit for its computation is shown in Figure 3.11, where $\tilde{i}_{nlC_e,2\omega_1-\omega_2}$ and $\tilde{i}_{nlg_m,2\omega_1-\omega_2}$ are the third-order nonlinear currents at $2\omega_1 - \omega_2$ from the two nonlinearity sources respectively.

Based on the formulas of nonlinear currents in [79], we find

$$
\tilde{i}_{nlC_e,2\omega_1-\omega_2} = j(2\omega_1 - \omega_2) \frac{3}{4} K_{3C_e} v_{be}^2(\omega_1) v_{be}(-\omega_2) \quad (3.25)
$$

$$
\equiv j \omega_1 \cdot \frac{3}{4} K_{3C_e} A_{vbe}^2(\omega_1) A_{vbe}(-\omega_2) \cdot v_S^3 \quad (3.26)
$$

![Figure 3.11 Equivalent circuit for the computation of the third-order intermodulation product $2\omega_1 - \omega_2$.]

$$
\tilde{i}_{nlg_m,2\omega_1-\omega_2} = \frac{3}{4} K_{3g_m} v_{be}^2(\omega_1) v_{be}(-\omega_2) \quad (3.27)
$$

$$
\equiv \frac{3}{4} K_{3g_m} A_{vbe}^2(\omega_1) A_{vbe}(-\omega_2) \cdot v_S^3 \quad (3.28)
$$
By applying Kirchoff’s law in the circuit of Figure 3.11, we find that

\[ v_c(2\omega_1 - \omega_2) = \frac{Z_L}{\det(2\omega_1 - \omega_2)}. \]  

(3.29)

\[ \{i_{mc_{x,2\omega_1-\omega_2}}[-j(2\omega_1 - \omega_2)C_{bc}Z_S + g_m(Z_E + Z_S)] \]
\[ -i_{mg_{m,2\omega_1-\omega_2}}[1 + j(2\omega_1 - \omega_2)C_{bc}Z_S + j(2\omega_1 - \omega_2)C_1(Z_E + Z_S)] \}

in which \( Z_E, Z_S \) and \( Z_L \) are evaluated at the frequency \( 2\omega_1 - \omega_2 \).

Since \( \omega_1 >> \omega_1 - \omega_2 \) and \( 2\omega_1 - \omega_2 \equiv \omega_1^6 \), \( Z_E, Z_S \) and \( Z_L \) evaluated at the frequency \( 2\omega_1 - \omega_2 \) are approximately equal to \( Z_E, Z_S \) and \( Z_L \) evaluated at the frequency \( \omega_1 \) respectively. Then \( \det(2\omega_1 - \omega_2) \equiv \det(\omega_1) \).

Substituting the expressions of the nonlinear currents (3.26) and (3.28) into (3.29) yields the final intermodulation product:

\[ v_c(2\omega_1 - \omega_2) \]
\[ \equiv \frac{Z_L}{\det(\omega_1)} \frac{3}{4} A_{cbe}(\omega_1) A_{cbe}(-\omega_2) \]
\[ \{ j\omega_1 K_{3Cz} [-j\omega_1C_{bc}Z_S + g_m(Z_E + Z_S)] \]
\[ - K_{3g_{m}}[1 + j\omega_1C_{bc}Z_S + j\omega_1C_1(Z_E + Z_S)] \} \cdot v_S^3 \]

Therefore,

\[ \text{In the simulation and experiment, } \omega_1 = 1955MHz \text{ and } \omega_2 = 1950MHz. \]
\[ \text{IMR}_3 = \frac{v_c(2\omega_1 - \omega_3)}{v_c(\omega_1)} \]

\[ \equiv \frac{3}{4} A_{\text{vbe}}^2 \left( \omega_1 \right) A_{\text{vbe}} \left( -\omega_2 \right) \]

\[ \left\{ K_{3C_e} \left[ j \omega_1 Z_E + j \omega_1 Z_s + \omega_1^2 C_{bc} Z_s / g_m \right] \right. \]

\[ - K_{3g_m} / g_m \left[ 1 + j \omega_1 C_1 Z_E + j \omega_1 (C_1 + C_{bc}) Z_s \right] \left. \right\} v_S^2 \]

(3.31)

Since the output matching network is a linear passive network, (3.31) is also valid for the ratio of the third-order intermodulation product and the fundamental at the output load.

### 3.3.3 Experimental Verification

In order to verify our theoretical analysis, experimental measurements of output IMR\(_3\) were taken in the test bench shown in Figure 3.12. The single-stage power amplifier shown in the figure was fabricated in IBM SiGe BiCMOS 6HP. The bipolar transistors in the power amplifier are composed of 100 devices, each with an emitter area of 48\(\mu\)m \(\times\) 0.44\(\mu\)m. The quiescent bias current at the collector is 110mA.

Figure 3.12 Test bench used for experimental verification of linearity analysis.
Figure 3.13 compares the IMR₃ from SPECTRE simulation, Volterra calculation based on the equation (3.31), and measurement for the circuit of Figure 3.12. The good agreement throughout the entire range of output powers from 0dBm to +26dBm verifies the effectiveness of PDCV analysis. In addition, the comparison validates the completeness of our model shown in Figure 3.4 and the corresponding extraction methodology described in Section 3.3.1.

![Figure 3.13 IMR₃ comparison between SPECTRE simulation, Volterra calculation, and measurement for the circuit of Figure 3.12.](image)

Our analysis not only accurately predicts the IMR₃, but also provides insight into the individual contributions from the main nonlinear sources. Figure 3.14 displays the calculated amplitude and phase of IMR₃ in three cases:

1) with the effect of the nonlinearity of $C_\pi$ only ($K_{3g_m} = 0$);

2) with the effect of the nonlinearity of $i_C$ only ($K_{3C_\pi} = 0$);

3) with the effects of the nonlinearities of $C_\pi$ and $i_C$ together.
Figure 3.14 Individual contributions to amplitude and phase of output IMR$_3$.$^7$

We observed that the nonlinearity of $C_\pi$ and the nonlinearity of $i_C$ are very close in amplitude over the whole power range, but they are opposite in phase, resulting in the well-known intermodulation cancellation effect [81]. In principle, this third-order nonlinearity cancellation is caused by the high-frequency limit of the third-order Volterra kernel [82], where the output distortion at $2\omega_1 - \omega_2$ rolls off with the
fundamental frequency \((\omega_1 \approx \omega_2)\) at -60dB/decade. At low input powers the nonlinearity of \(i_C\) dominates the magnitude and the phase of output IMR\(_3\). The \(C_\pi\) nonlinearity dominates at high input powers so that the output IMR\(_3\) is determined by the \(C_\pi\) nonlinearity. The overall linearity of the SiGe HBT PA depends on the degree to which this cancellation is achieved [82]. Furthermore, the linearity is also limited due to considerations of power gain and efficiency.

### 3.3.4 Effect of Emitter Bond-wire Inductance

The effect of bond-wires connected to the emitter of the power amplifier is critical in the design of the power amplifier. Without considering its effect, the power gain and the linearity cannot be accurately predicted. Therefore, it is necessary to include its effect in our linearity analysis.

Expression (3.31) was compared to SPECTRE simulations for different values of the bond-wire inductance, as shown in Figure 3.15. From the figure, we observed that the linearity of the power amplifier is improved with the increase of the bond-wire inductance \(L_e\). Obviously, it may not be possible to pursue the best linearity by continuously increasing the value of \(L_e\), due to the reductions in gain and PAE.

\[\text{7 Note that the jump in the phase of output IMR}_3 \text{ comes from the fact that: at low input powers the nonlinearity of } i_C \text{ dominates, so the phase of output IMR}_3 \text{ follows that of } i_C; \text{ whereas at high input powers the } C_\pi \text{ nonlinearity dominates, so the phase of output IMR}_3 \text{ follows that of } C_\pi.\]
Figure 3.15 IMR$_3$ with different values of $L_e$.  $I_{CQ} = 110\text{mA}$.

Figure 3.16 Comparison of simulated and calculated power gains with different values of $L_e$.

The effect of $L_e$ on power gain is demonstrated in Figure 3.16, where three different values of $L_e$ are compared. Calculation results are based on (2.5), which is the complete expression for power gain of the amplifier (including the effects of source impedance $Z_L$, load impedance $Z_s$ and bond-wire inductance $L_e$). All circuit
parameters are extracted at the corresponding bias point from SPECTRE simulation. Note that the power gain is very sensitive to the effect of $L_e$.

The effect of $L_e$ on power added efficiency (PAE) is illustrated in Figure 3.17, where PAEs at 26 dBm output power are compared with different values of $L_e$. The effect of emitter inductance on PAE is not as significant as its effect on power gain, but we can still observe that with an increase of $L_e$, the power gain decreases and so does the PAE. This can be seen from the definition of PAE:

$$PAE = \frac{P_{out} - P_{in}}{P_{dc}} = \frac{P_{out}}{P_{dc}} \left(1 - \frac{1}{G}\right)$$  \hspace{1cm} (3.32)

![Figure 3.17 Simulated power added efficiencies at 26 dBm output power with different values of $L_e$.](image)

For WCDMA power amplifiers, it is always desirable to have the power gain and PAE as high as possible, while satisfying the linearity requirement. For our power amplifier, we found that 80pH is the optimum value for bond-wire inductance $L_e$, considering the tradeoff between gain, linearity and efficiency.
3.4 Analysis of SiGe Dynamic Biased Amplifiers with Power Detector

To dynamically bias RF amplifiers, power detectors are typically used to sense the average input signal power and adjust the amplifier bias level accordingly. A SiGe dynamically-biased amplifier with a power detector [45] has been implemented to save the dc power at the low input power level. Its simplified schematic is shown in Figure 3.18, where the RF amplifier is differential, but its single-ended view is shown here for simplicity. At the output of the power detector, the extra dc current $\Delta I_{cq}$, which is proportional to the input power, will be applied to the RF amplifier ($Q_{amp}$) through the digitally-programmable bias current mirror network.

![Figure 3.18 Simplified schematic of the dynamically-biased RF amplifier.](image)

Besides the additional dc current, the power detector circuit also produces an envelope signal current as

$$i_{env}(t) = I_{env} \cos[(\omega_2 - \omega_1)t + \theta_{env}]$$

(3.33)
where $I_{env}$ and $\theta_{env}$ denote the amplitude and the phase of the envelope signal respectively.

Therefore, this additional envelope signal has to be included in the linearity analysis of the RF amplifier. The effects of the envelope signal injection on amplifier distortion will be analyzed using the PDCV technique. The results will explain the asymmetry in the third-order intermodulation distortion (IMD$_3$) of the RF amplifier and indicate an improved amplifier linearization technique with envelope signal injection.

### 3.4.1 Volterra Analysis of RF Amplifiers with Envelope Detector

Assume that the selection of nonlinearity sources and the extraction of nonlinearity coefficients are the same as those in Section 3.3. In the PDCV analysis of the dynamically-biased RF amplifier, there are three input tones: two fundamental tones ($\omega_1$ and $\omega_2$) and the envelope signal ($\omega_3 = \omega_2 - \omega_1$). As a result, the two-tone IMD$_3$ expression derived in (3.29) has to be modified to obtain the three-tone IMD$_3$ counterpart. Accordingly, the output IMD$_3$ voltage is re-written as

$$v_C(\omega) = \frac{Z_L}{\det(\omega)} \left\{ \tilde{\gamma}_{nlS,\omega} \left[ -j \omega C_{bc} Z_S + g_m (Z_E + Z_S) \right] ight. $$

$$- \left. \tilde{\gamma}_{nlB,\omega} \left[ 1 + j \omega C_{bc} Z_S + j \omega C_1 (Z_E + Z_S) \right] \right\}$$

(3.34)

where $\omega$ represents the third-order intermodulation frequency, i.e. $2\omega_1 - \omega_2$ at the lower side and $2\omega_2 - \omega_1$ at the upper side, assuming $\omega_1 < \omega_2$. 
The third-order nonlinear currents are given by

\[
\tilde{I}_{nlg_{an},2\omega_a-\omega_b} = K_{2g_{an}} \cdot \left[ \mathbb{1} + \mathbb{2} + \mathbb{3} + \mathbb{4} \right] + \frac{3K_{3g_{an}}}{4} \left[ \mathbb{5} + \mathbb{6} \right]
\]

\[
\tilde{I}_{nl_{ce},2\omega_a-\omega_b} = j(2\omega_a - \omega_b) \cdot \left[ K_{2C_{ce}} \cdot \left[ \mathbb{1} + \mathbb{2} + \mathbb{3} + \mathbb{4} \right] + \frac{3K_{3C_{ce}}}{4} \left[ \mathbb{5} + \mathbb{6} \right] \right]
\]

(3.35)

where \(\mathbb{1}-\mathbb{6}\) denote all possible combinations of the lower-order (fundamental and 2nd-order) terms that will give rise to the IMD3 products. Specifically, for IMD3 at \(2\omega_1 - \omega_2\), these six terms are:

\[
\begin{align*}
\mathbb{1} &= v_{be}(2\omega_1)v_{be}(-\omega_2) \\
\mathbb{2} &= v_{be}(\omega_2)v_{be}(-2\omega_3) \\
\mathbb{3} &= v_{be}(\omega_1)v_{be}(\omega_1-\omega_2) \\
\mathbb{4} &= v_{be}(\omega_2)v_{be}(-\omega_3) \\
\mathbb{5} &= v_{be}(-\omega_1)v_{be}(\omega_1) \\
\mathbb{6} &= v_{be}(\omega_2)v_{be}(-\omega_3)
\end{align*}
\]

(3.36)

Similarly, the corresponding terms for the IMD3 at \(2\omega_2 - \omega_1\) are given by

\[
\begin{align*}
\mathbb{1} &= v_{be}(-\omega_1)v_{be}(2\omega_2) \\
\mathbb{2} &= v_{be}(\omega_1)v_{be}(2\omega_3) \\
\mathbb{3} &= v_{be}(\omega_2)v_{be}(\omega_2-\omega_1) \\
\mathbb{4} &= v_{be}(\omega_2)v_{be}(2\omega_3) \\
\mathbb{5} &= v_{be}(-\omega_1)v_{be}(\omega_2) \\
\mathbb{6} &= v_{be}(\omega_2)v_{be}(\omega_3)
\end{align*}
\]

(3.37)

where the fundamental-order base-emitter voltage at \(\omega\) has been derived in (3.19) and the 2nd-order base-emitter voltage at \(\omega\) is found to be

\[
v_{be}(\omega) = \frac{-1}{\text{det}(\omega)} \left\{ \tilde{I}_{nlg_{an},\omega}\left[ Z_E + j\omega C_{bc}(Z_LZ_S + Z_LZ_E + Z_SZ_E) \right] \right. \\
+ \left. \tilde{I}_{nl_{ce},\omega}\left[ Z_S + Z_E + j\omega C_{ce}(Z_LZ_S + Z_LZ_E + Z_SZ_E) \right] \right\}
\]

(3.38)

For double frequency terms (such as \(2\omega_1\)), the nonlinear currents are given by
Similarly, for difference frequency terms (such as $\omega_2 - \omega_1$), the nonlinear currents are given by

\[
\begin{align*}
\tilde{i}_{n_{lg}, 2\omega_a} &= \frac{K_{2g_m}}{2} v_{be}^2(\omega_a) \\
\tilde{i}_{n_{c}, 2\omega_a} &= K_{2c} j\omega_a v_{be}^2(\omega_a)
\end{align*}
\] (3.39)

Finally, IMR$_3$ is determined by the ratio between the third-order and the fundamental output voltage given by (3.34) and (3.18) respectively.

### 3.4.2 Verification and Discussion

The effect of the phase of the injected envelope signal on the intermodulation distortion is shown in Figure 3.19, where an envelope signal of fixed amplitude and arbitrary phase is injected. The simulation was based on the circuit in Figure 3.18 at the maximum power level. The calculation results were derived from the PDCV analysis in the previous section. As shown in Figure 3.19, excellent agreement between simulation and calculation results verifies our theoretical analysis.

Furthermore, there are a number of observations from the figure. First, with the change of the phase of the envelope signal, asymmetry is observed between the lower-side IMD$_3$ ($2\omega_1 - \omega_2$) and the upper-side IMD$_3$ ($2\omega_2 - \omega_1$). Second, maximum IMD$_3$ cancellation can be achieved when the envelope signal is injected with the optimal phase relative to the RF inputs.
The above observations can be intuitively comprehended with the contributing terms of IMD$_3$ derived in (3.36) or (3.37). The terms denoted by ①③⑤ (or ①⑤⑦) are the conventional IMD$_3$ components, which are generated by two fundamental tones ($\omega_1$ and $\omega_2$) and independent of envelope-signal ($\omega_3$). The remaining terms ②④⑥ (or ②④⑥) are envelope-signal dependent. In the following analysis, terms ②④ (or ②④) will be ignored, because they are second-order of small envelope input.

The cause of IMD$_3$ asymmetry is illustrated in Figure 3.20, where vectors ③ and ④ denote the IMD$_3$ contribution by the injected envelope signal and vectors $\Sigma [①③⑤]$ and $\Sigma [①⑤⑦]$ represent the sum of the conventional IMD$_3$ components. Since the angle between vectors ④ and $\Sigma [①⑤⑦]$ is smaller than that between ③ and $\Sigma [①③⑤]$ (i.e. $\theta_1 < \theta_2$), the resultant IMD$_3$ vector at $2\omega_2 - \omega_1$ will be bigger than that
at $2\omega_1 - \omega_2$. Consequently, IMD$_3$ asymmetry is caused by the difference in the angle between individual components, such as $\theta_1$ and $\theta_2$ in Figure 3.20.

![Vector diagram showing IMD$_3$ asymmetry.](image)

Figure 3.20 Vector diagram showing IMD$_3$ asymmetry.

The optimization of IMD$_3$ cancellation is demonstrated in Figure 3.21. A $90^\circ$ phase shift is introduced to the envelope signal. As a result, both envelope-dependent vectors $\Theta$ and $\Phi$ rotate in opposite direction by the same angle ($90^\circ$), where they are at $180^\circ$ opposite to the vectors of $\Sigma[1\Theta\Phi]$ and $\Sigma[1\Phi\Theta]$. Consequently, the resultant IMD$_3$ vectors at both $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ are greatly reduced. This indicates that envelope signal injection can be optimized to be a linearization method for IMD$_3$ cancellation in RF amplifiers.
Besides simulation verification, the Volterra calculation results are also compared to the measured data. A dynamically-biased amplifier of Figure 3.18 was fabricated in IBM SiGe BiCMOS 6HP. Two input tones at 1.9475 and 1.9525GHz are applied and the envelope detector is enabled. The improvement of IMD$_3$ versus input power is plotted in Figure 3.22. Since good agreement is observed between the measured and calculation results, our Volterra analysis is further confirmed.
3.5 Summary

By using power-dependent coefficient Volterra analysis, a SiGe HBT power amplifier is analyzed to highlight the dominant sources of nonlinearity and provide directions to design a linear SiGe HBT high-power amplifier. Through the analysis of a dynamically-biased RF amplifier with envelope signal injection, the cause of IMD3 asymmetry has been clarified. Furthermore, the linearity of the amplifier can be improved by optimizing the phase and amplitude of the injected envelope signal. Both analyses are confirmed with simulation and experimental results.

This chapter, in part or in full, is a reprint of the materials in 2004 IEEE Radio and Wireless Conference (RAWCON), 2004 IEEE Custom Integrated Circuits Conference (CICC), and 2005 IEEE Journal of Solid-State Circuits. The author of this dissertation was the primary researcher and the first author of the RAWCON.
paper, and the primary researcher of the theoretical analysis and the second author of the CICC and JSSC papers.
CHAPTER 4
Linearity Improvement by Digital Predistortion

The power-dependent coefficient Volterra series discussed in the previous chapter represents an effective analytical method for calculating the linearity of high-power amplifiers. The insights into their nonlinearity mechanisms are provided using a simple and manageable hand analysis technique. In this chapter, we will focus on the linearity improvement of power amplifiers, which will increase the maximum output power satisfying the linearity requirements. Consequently, the peak power efficiency will be also boosted. The objective of this chapter is to find an effective linearization technique, which should be conveniently integrated, flexibly controlled, and economically implemented without sacrificing other performance, such as power gain and efficiency.

This chapter will begin with the comparison of different linearization techniques [83]-[120], such as feedback, feedforward, and predistortion. Based on the aforesaid objective, digital predistortion (DP) was chosen as our linearization method. Subsequently, the operation principle and hardware implementation of digital predistortion will be illustrated in detail. Experimental results will be presented to
demonstrate the effectiveness of digital predistortion, which is followed by concluding remarks.

4.1 Linearization Techniques

Different linearization techniques are compared in Table 1.1. According to their individual features, they are employed in different scenarios. For instance, feedforward techniques are effective, but very costly. Therefore, they are typically utilized in base-station power amplifiers, instead of handset applications. The details of these linearization techniques will be discussed in the following sections.

Table 4.1 Comparison of different PA linearization techniques.

<table>
<thead>
<tr>
<th>Linearization Technique</th>
<th>Linearization Performance</th>
<th>Compensation Bandwidth</th>
<th>Cost</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback</td>
<td>moderate</td>
<td>narrow</td>
<td>moderate</td>
<td>stability; reduced gain</td>
</tr>
<tr>
<td>Feedforward</td>
<td>good</td>
<td>wide</td>
<td>high</td>
<td>not for handset PAs</td>
</tr>
<tr>
<td>Analog/RF Predistortion</td>
<td>low</td>
<td>wide</td>
<td>low</td>
<td>simplest form; reduced gain</td>
</tr>
<tr>
<td>Digital Predistortion</td>
<td>moderate</td>
<td>wide</td>
<td>moderate</td>
<td>easy to integrate and control</td>
</tr>
</tbody>
</table>

4.1.1 Feedback

The *negative* feedback improves the linearity of the PA by feeding back the sampled output signal to the input and accordingly modifying the amplitude and/or phase of the input signal to the PA. Judging by the frequency of the feedback signal, feedback linearization techniques can be classified into two groups: RF feedback and
low-frequency feedback. Furthermore, Cartesian and polar modulation feedback are two basic types of the low-frequency feedback. The details of three feedback techniques will be discussed as follows.

**RF Feedback**

The general scheme of RF feedback amplifiers is shown in Figure 4.1. Second-order harmonic feedback [84] is one representative example. This technique samples the second-order harmonics at the output of the nonlinear power amplifier and feeds it back to the input of the amplifier, where there are three feedback signals \( A_1 \cos(2\omega t + \phi) \), \( A_{22} \cos(2\omega t + \phi_2) \) and \( A_{12} \cos((\omega_1 + \omega_2) t + \phi_1) \) as well as two fundamental signals \( A_1 \cos(\omega_1 t) \) and \( A_2 \cos(\omega_2 t) \). Assuming that the transconductance is the dominant nonlinearity, the output current can be approximated by a third-order power series, as

\[
i_o = g_m v_{in} + g_{m2} v_{in}^2 + g_{m3} v_{in}^3
\] (4.1)

Nonlinearity of the amplifier causes interaction between the source signals and their feedback second harmonics. For the third-order intermodulation product (IMD3) at the frequency \( (2\omega_2 - \omega_1) \), we have

\[
A_1 A_{22} g_{m2} \cos(2\omega_2 t - \omega_1 t - \phi_2) + \frac{3A_1 A_{32}^2 g_{m3} \cos(2\omega_2 t - \omega_1 t)}{4}
\] (4.2)

in which the first term comes from the second-order intermodulation between \( A_1 \cos(\omega_1 t) \) and \( A_{22} \cos(2\omega_2 t + \phi_2) \), the second term is the consequence of the third-
order intermodulation between \( A_1 \cos(\omega_1 t) \) and \( A_2 \cos(\omega_2 t) \), and there is no term generated by \( A_{12} \cos((\omega_1 + \omega_2)t + \phi) \). By properly selecting the phase and amplitude of the feedback second-order harmonics (\( A_{22} \) and \( \phi_2 \)), it is possible to make these two terms cancel out each other. In principle, the complete cancellation of the IMD\(_3\) can be achieved. However, the phase and amplitude adjustment in the feedback path have to be very accurate in order to realize the ideal cancellation. Besides, the existence of band pass filter in the feedback path limits this technique to narrowband systems.

![Figure 4.1 Simplified schematic of RF feedback amplifiers.](image)

**Cartesian Modulation Feedback**

The Cartesian modulation feedback technique was first proposed by Petrovic in 1983 [85]. As shown in Figure 4.2, the sampled RF output signal is downconverted into a quadrature (I and Q) feedback signal, which is subtracted from the quadrature input signal to obtain the error signal. After passing through the loop filters, the error signal is upconverted into a complex RF signal. This complex signal is then fed into
the power amplifier. The phase shifter synchronizes the downconversion and upconversion paths. The attenuator is used to adjust the sampled output signal to the proper level suitable for the downconverter.

Figure 4.2 Schematic of Cartesian modulation feedback.

In essence, the Cartesian feedback system is a linearized transmitter rather than a linearized power amplifier. Cartesian transmitter systems [86]-[87] have been reported with more than 20dB suppression of the IMD3 and up to 500KHz modulation bandwidth. However, some practical issues need to be addressed in the Cartesian system design.

- **Synchronization.** The synchronization between the upconversion and downconversion paths has been difficult without manual trimming. To solve this historic problem, several automatic phase alignment techniques have been
proposed recently [88]-[89]. However, this further increases the complexity of the whole system.

- **Loop Stability.** The loop stability is affected by the loop delay [90], as well as by the characteristics of the nonlinear power amplifier [91]. A bigger loop delay will result in a smaller phase margin of the loop transfer function and therefore reduces the loop stability [90]. On the other hand, careful design of circuit components and parameters [91] is critical for stable feedback systems.

**Polar Modulation Feedback**

The polar feedback technique was also proposed by Petrovic [92]. The general diagram of a polar feedback system\(^1\) is shown in Figure 4.3. The sampled output RF signal is downconverted to an Intermediate Frequency (IF) signal by the local oscillator and the mixer. The IF signal is then decomposed into the polar form (phase and amplitude) with a limiter and a demodulator. The same topology is employed to extract the phase and amplitude of the input IF signal. The outputs of two demodulators are fed into an error amplifier to generate the error signal, which controls the amplitude of the RF signal. The phase control path is essentially a simple phase-locked loop (PLL), which includes a phase detector, a loop filter, a loop amplifier, and a voltage-controlled oscillator (VCO). Therefore, both the amplitude

---

\(^1\) Note that the polar loop feedback is also a transmitter linearization technique, rather than a power amplifier linearization technique.
and phase of the RF signal can be carefully controlled with independent feedback loops.

Figure 4.3 Schematic of polar modulation feedback.

The polar feedback system has a number of advantages. First, the RF path consists of a VCO, a modulation amplifier, and a power amplifier. Since there is no mixer, the issues related to the mixer (such as image-reject filtering) can be avoided. Second, high-efficiency power amplifiers can be employed in the RF path so that the efficiency of the whole system is improved. Several multi-band direct conversion transmitters in the polar loop topology [93]-[94] have been reported with the RF frequency up to 1900MHz and spurious emission more than 60dB below the desired RF signal.
On the other hand, several issues limit the application of the polar feedback technique in RF systems. The phase-locked loop may have difficulty in locking at low envelope levels or tracking drastic phase changes. Furthermore, its feedback bandwidth requirement is much higher than that of the Cartesian loop, considering the phase discontinuity inherent in multi-tone signals in the phase feedback loop [83].

4.1.2 Feedforward

The feedforward technique was invented by Black [95] nine years before the feedback concept [96]. The concept of feedforward systems is simple, but its hardware implementation is quite costly. Consequently, compared to its feedback counterpart, the feedforward technique is historically less popular. With more awareness of the limitations of conditional stability and loop bandwidth in the feedback systems, the feedforward technique has drawn more attention, especially in the wideband and multi-carrier systems [97]-[98].

Figure 4.4 shows the simplified scheme of feedforward systems. The RF input signal is split through a hybrid splitter into two paths: a main RF path and a signal cancellation path. The distortion generated by the main power amplifier is sampled and fed into a subtractor together with the delayed RF signal. By adjusting the attenuator, the RF signal is completely cancelled out at the output of the subtractor. The residual distortion products are linearly amplified by an error amplifier, and then combined with the distorted RF signal in the main RF path. Ideally, an amplified RF signal without distortion will be generated at the output of the injector coupler.
Figure 4.4 Schematic of feedforward system.

There are several practical considerations in the implementation of feedforward systems. The error amplifier [99] must be highly linear, which results in low-efficiency topology (e.g. Class A) and limits the total efficiency of the feedforward system [100]. Besides, the phase and amplitude imbalances between the main RF path and the signal cancellation path will dramatically degrade the linearization performance [101]. For instance, a 25dB distortion suppression requires either an amplitude error of 0.5dB or less or a phase error of 0.5 degree or less [99]. As a result, extra control must be employed to synchronize both the gain and phase. With the advancement of digital signal processor (DSP) technology, many digitally controlled feedforward amplifiers have been reported [102]-[103]. Furthermore, the implementation of the feedforward technique is so costly that it is employed only in the base-station and satellite systems [104]-[105].
4.1.3 Predistortion

Predistortion [83] is conceptually the simplest linearization technique for RF power amplifiers. Figure 4.5(a) illustrates the general form\(^2\) of the predistortion technique. Its compensation principle is shown in Figure 4.5(b). By creating a distortion characteristic complementary to the distortion characteristic of the nonlinear power amplifier and cascading them, linear amplification can be achieved between the input and output of the whole system. Based on the predistortion frequency, most predistortion systems can be categorized into two groups: analog and digital predistortion. Analog predistortion is typically implemented in the RF/IF frequency, whereas digital predistortion is generally achieved in the baseband domain.

![Figure 4.5 Schematic of predistortion systems. (a) general form. (b) compensation principle.](image)

\(^2\) Note that the predistortion system in Figure 4.5 is in the form of open loop. However, most practical predistortion systems employ a certain kind of feedback to adaptively compensate some external effects such as temperature changes.
Analog Predistortion

There are two types of analog predistortion: cubic and diode/FET predistorters. Cubic predistorters aim at eliminating the third-order distortion. Consequently, Cubic predistorters are particularly functional in traveling-wave-tube (TWT) amplifiers [106], in which the third-order distortion dominates the nonlinear characteristic.

A representative cubic predistorter is shown in Figure 4.6. The lower path generates the third-order nonlinearity that is the inverse of the counterpart in the nonlinear power amplifier. By combining the third-order nonlinearity and the delayed input signal, the predistorted signal is generated as the input to the nonlinear power amplifier. Note that two amplifiers in the lower path have to be linear and hence contribute a negligible distortion.

![Figure 4.6 Schematic of cubic predistorters.](#)

To effectively suppress the spectral regrowth in WCDMA systems, the odd-order intermodulation distortions should be cancelled. As a result, higher odd-order distortions need to be compensated as well as the third-order one. Several modified
cubic predistorters have been reported [107], by canceling both third- and fifth-order distortions. However, since it is generally implemented on the board level, cubic predistorters are employed in base stations, instead of handsets.

Compared to cubic predistortion, diode/FET predistorters [108]-[112] possess better simplicity and modularity. A diode/FET predistorter generates a nonlinear characteristic, which is the inverse of the transfer function of the nonlinear power amplifier, in both magnitude and phase. The nonlinear device to produce the needed transfer function can be a diode [108]-[111] or FET [112]. An example of parallel diodes [109] is shown here to illustrate the general principle of diode/FET predistorters.

![Figure 4.7 Parallel diode predistortion schematic and characteristics.](image)

As shown in Figure 4.7(a), the linearizer consists of a parallel diode with a bias feed resistance $R_b$ and a parallel capacitor $C_p$. The gain and phase transfer functions of the linearizer can be readily obtained as follows:
\begin{align*}
|S_{21}| &= \frac{2}{\sqrt{\left(2 + \frac{Z_o}{R}\right)^2 + \left(\omega(C_d + C_p)Z_o\right)^2}} \tag{4.3} \\
\angle S_{21} &= \tan^{-1}\left(\frac{\omega(C_d + C_p)Z_o}{2 + \frac{Z_o}{R}}\right) \tag{4.4}
\end{align*}

where, \( R = \frac{R_d}{R_b} = \frac{R_d \cdot R_b}{R_d + R_b} \).

As shown in Figure 4.7(b), with the increase of RF input power, \( R_d \) (i.e. \( R \)) increases (from \( R_{ds} \) to \( R_{di} \)). As a result, based on (4.3) and (4.4), the linearizer shows the positive gain and negative phase deviations, which are inverse to the gain and phase changes of the nonlinear power amplifiers respectively. From (4.3), the amplitude of the gain deviation of the linearizer can be further adjusted by varying \( C_p \). Experimental results showed a 20dB improvement of the IMD3 in an 18GHz-band PA at 15dBm output power.

There are a number of limitations in this predistortion technique. First, it works merely for specific nonlinearities and therefore it is not a general linearization method suitable for any kind of nonlinear power amplifiers. Second, although it achieves a positive gain deviation, its power gain is negative at all times, which degrades the efficiency of the power amplifier. Third, its linearization performance is moderate and the effective linearization range is rather limited. Consequently, it is not popular in the high-power amplifier linearization.


**Digital Predistortion**

Digital predistortion [83] is a linearization technique to predistort the baseband modulation signal in the digital signal processor (DSP) prior to the digital-to-analog converter (DAC), up-converter, and RF power amplifier. As shown in Figure 4.8, the sampled output power is down-converted and digitized into a baseband feedback signal and then fed into the DSP to predistort the input baseband signal. The amount of predistortion is controlled by a look-up table (LUT) that characterizes the nonlinearities of the RF power amplifier.

![Figure 4.8 Block diagram of digital predistortion.](image)

Different predistortion algorithms in the DSP have been proposed [113]-[119]. A straightforward solution is the *mapping* predistortion [114]-[115]. As shown in Figure 4.9(a), a two-dimensional table in a random access memory (RAM) contains complex predistorting signals. The mapping table accepts the input signal $S$ in the Cartesian form as the table address and outputs the corresponding correction value.
\( C(S) \) to predistort the signal \( S \). The predistorted signal \( \overline{S} \) is obtained with the sum of \( S \) and \( C(S) \). The mapping table can be adaptively adjusted by comparing the feedback signal \( R \) with the input signal \( S \). The major drawback of this approach is the size of the two-dimensional table (e.g. 2 Mwords in [114]), which results in the long convergence time\(^3\) (e.g. 10 seconds in [114]).

To overcome the size issue of the mapping predistortion, Faulkner proposed a polar predistortion [116], which contains two one-dimensional tables controlling gain and phase compensations respectively. As shown in Figure 4.9(b), the input signal is predistorted after the gain factor multiplication and phase rotation. In order to find the gain factor in the gain table, the Cartesian input signals need to be converted into the one-dimensional amplitude signal as the address in the gain table. The address of the phase table is obtained by multiplying the amplitude signal with the gain factor. The adaptive correction terms for the gain and phase tables are obtained by comparing the gain and phase differences respectively between the input signal and the sampled output signal. Compared with the mapping counterpart, the polar digital predistortion reduces the table size considerably. It was verified [116] that the size of the polar table is three orders of magnitude less than that of the mapping table, as shown in Figure 4.10.

\(^3\) The convergence time determines whether the algorithm is fast enough to be real-time, and also affects the linearization performance.
Figure 4.9 Different digital predistortion algorithms [113].
Figure 4.10 Look-Up-Table (LUT) size comparison between mapping and polar predistortion.

Cavers [117] presented another approach using a complex-gain table instead of two separate gain and phase tables. As shown in Figure 4.9(c), the input signal is predistorted by a complex multiplication between the input signal and the complex gain factor. The address of the complex gain factor is obtained by calculating the squared magnitude of the input signal. Compared with the polar predistortion, this technique reduces the number of operations in the predistorter. Compared to [114], the complex-gain approach [117] requires more than four orders of magnitude less memory and reduces convergence time by over three orders of magnitude. Because of its low memory and fast adaptation, the complex-gain predistortion has been popular in the linearization of nonlinear power amplifiers [118]-[119].

Compared with other linearization techniques, digital predistortion can effectively linearize handset power amplifiers with high integration and wide compensation bandwidth at moderate cost. At the same time, the rapid development
of DSP technology [120] renders digital predistortion in handset power amplifiers more promising and feasible. Therefore, it is employed as the linearization technique in the design with switched dual dynamic biasing\(^4\) in this dissertation. Furthermore, considering the aforesaid merits of complex-gain predistortion, it is taken as the predistortion algorithm in this dissertation.

### 4.2 Digital Predistortion Implementation

#### 4.2.1 Operation Principle

The concept of complex-gain digital predistortion was introduced in the previous section. Now we will derive its operating principle using the conceptual implementation diagram shown in Figure 4.11. Note that the quadrature up- and down-converters and the reconstruction and anti-aliasing filters are ignored for simplicity.

\(^4\) Since the linearity of the design with switched dynamic current biasing can satisfy the WCDMA specifications without any linearization technique, digital predistortion is not employed in the design with switched dynamic current biasing.
Figure 4.11 Conceptual implementation diagram of complex-gain predistortion [121].

It is straightforward to derive the output signal \( y(t) \) as

\[
y(t) = g(x(t) \cdot f(\| x(t) \|)) \tag{4.5}
\]

where \( x(t) \) is the input signal, \( g(\cdot) \) denotes the transfer function of the nonlinear power amplifier, and \( f(\cdot) \) represents the predistortion correction function. At time instant \( t_i \), the amplifier output \( y(t) \), is equal to the desired output \( h(x(t)) \). Hence,

\[
h(x(t_i)) = y(t_i) = g(x(t_i) \cdot f(\| x(t_i) \|)) \tag{4.6}
\]

For the ideal amplification with a constant gain \( G \),

\[
h(x(t_i)) = Gx(t_i) \tag{4.7}
\]

Combining (4.6) and (4.7), we have

\[
Gx(t_i) = g(x(t_i) \cdot f(\| x(t_i) \|)) \tag{4.8}
\]
Assume that $f_A(\omega)$ and $f_\theta(\omega)$ are the magnitude and phase components of $f(\omega)$.

Taking the magnitude of both sides of (4.8) yields:

\[
|Gx(t_i)| = |g(x(t_i) \cdot f(|x(t_i)|))| \tag{4.9}
\]

\[
G |x(t_i)| = g_A(|x(t_i)| \cdot f_A(|x(t_i)|)) \tag{4.10}
\]

To simplify the above expression, let $\alpha = |x(t_i)|$. Then,

\[
G\alpha = g_A(\alpha \cdot f_A(\alpha)) \tag{4.11}
\]

The magnitude correction of the predistortion function $f(\omega)$ can be found as:

\[
f_A(\alpha) = \frac{g_A^{-1}(G\alpha)}{\alpha} \tag{4.12}
\]

Consequently, the way to calculate the magnitude correction of the predistortion function is a procedure of rescaling the input magnitude. The principle of magnitude rescaling is illustrated in Figure 4.12. For any input power $\alpha$, the corresponding ideal output power is equal to $h(\alpha)$. The actual amplifier is capable of yielding the same output power, but at the scaled input power $f(\alpha)\alpha$. Therefore, $h(\alpha) = g(f(\alpha)\alpha)$. This inverse scaling is readily described in (4.12).
Figure 4.12 Predistortion magnitude rescaling.

The phase correction $f_\phi(\alpha)$ is simply the inverse of the phase error at the rescaled magnitude:

$$f_\phi(\alpha) = -g_\phi(\alpha f_\lambda(\alpha)) \quad (4.13)$$

### 4.2.2 Hardware Implementation

The hardware implementation diagram has been shown in Figure 4.8. For WCDMA handset power amplifiers, the digital predistortion system downconverts a portion of the 1.95GHz output signal to an analog IF of 140MHz, which is then converted to a 12-bit digital second IF of 26.88MHz at a sample rate of 107.52MHz and fed to the DSP. After comparison with the corresponding input signal during a “training” period, the resulting error amplitude and phase signal are used to adaptively predistort the subsequent input I/Q signals. The predistorted digital signal is converted to a 14-bit analog signal at an IF of 140MHz, and then upconverted to the final 1.95GHz output.
4.3 Experimental Results

As mentioned in the previous section, the digital predistortion system is employed in the design with switched dual dynamic biasing. The final test setup is shown in Figure 4.13. The test procedure is straightforward and therefore neglected here for simplicity.

![Test setup of the digital predistortion system.](image)

**Figure 4.13 Test setup of the digital predistortion system.**

With a WCDMA reverse-link signal, the linearity of the power amplifier is measured in terms of ACPR for two cases: with and without digital predistortion (DP). The measured results at the output power of 22.4dBm are shown in Figure 4.14, where the ACPR performance of the power amplifier is improved by 10dB with digital predistortion.
At the same time, the correction effectiveness of digital predistortion is also checked by comparing the AM/AM relationship between input and output envelope amplitudes. Figure 4.15 shows the AM/AM curve without digital predistortion, in which the nonlinearity exists over the whole power range. On the contrary, with digital predistortion, the output amplitude has been thoroughly linearized, as shown in Figure 4.16. Based on the above discussions, digital predistortion has been proven to be an effective linearization technique for handset power amplifiers and already applied in the design with switched dual dynamic biasing. The experimental results will be presented in detail in Chapter 5.
Figure 4.15 Input/output behavior without digital predistortion – measured data (dots), ideal performance (dashed line), curvefit (light line).

Figure 4.16 Input/output behavior with digital predistortion – measured data (dots), ideal performance (dashed line), curvefit (light line).
4.4 Summary

Based on the comparison of different linearization techniques, complex-gain digital predistortion is employed in the design with switched dual dynamic biasing described in the second chapter. The operation principle of digital predistortion has been derived rigorously. The effectiveness of digital predistortion has been verified through experimental results.

This chapter, in part or in full, is a reprint of the materials in 2005 IEEE Radio Frequency Integrated Circuits Symposium. The author of this dissertation was the primary researcher and the first author in this publication.
In this chapter, we will present experimental results of two prototype chips with switched dynamic current biasing (SDCB) and switched dual dynamic biasing (SDDB) respectively. We will begin with a complete description of the post-tapeout test procedure, which includes the following steps: die packaging, board generating and assembling, and test bench setup. Subsequently, the prototype chips will be tested individually and their measured performances will be compared to the existing dynamically-biased power amplifiers. Finally, concluding remarks will be given.

5.1 Post-Tapeout Procedure

There are two general approaches for measuring chips: probe and printed circuit board (PCB) testing. In a probe test, the chip die is first put on a probe station. Then microwave probes directly touch the die bondpads to provide the RF signal path and DC biases. A typical probe system is shown in Figure 5.1, where the probe station is mounted on an air floating table to ensure vibration-free measurements.

Compared to probe testing, PCB testing is a more complicated and costly process, which requires die packaging, PCB fabrication and assembling. As a result,
probe testing is preferred for testing prototype chips. However, since power amplifiers consume large dc currents, the thermal-mechanical effect may prevent the probe tips from stably touching the chip bondpads, which may consequently make the test unstable and unrepeatably. Furthermore, power amplifiers are large-signal circuits, where parasitics and low thermal resistance are crucial to the circuit performance. In order to get realistic and reliable measurements of the circuit performance, it was decided to use PCB testing for power amplifiers in this dissertation.

![A typical microwave probe system.](image)

**Figure 5.1 A typical microwave probe system.**

The details of PCB testing will be described in the following sections. The design with switched dual dynamic biasing (SDDB) will be taken as an example to illustrate the details of these steps. Since the post-tapeout procedure of the design with
switched dynamic current biasing (SDCB) is the same as that of the design with SDDB, it will be neglected for simplicity.

- **Die Packaging**

  The Micro Lead Frame (MLF)\(^5\), belonging to the QFN (Quad Flat No-Lead) package family, is employed in our designs. As shown in Figure 5.2, the MLF is a plastic encapsulated package with a copper lead-frame substrate and uses perimeter lands on the bottom of the package to provide electrical contact to the printed circuit board. It offers the ExposedPad™ technology as a thermal enhancement by having the die attach paddle exposed on the bottom of the package surface to provide an efficient heat path when soldered directly to the board. This enhancement also enables a stable ground by use of down bonds or by electrical connection through a conductive die-attaching material [122]. Furthermore, its small body size and short bond-wires reduce the parasitic bond-wire inductance, which is critical to the performance of power amplifiers.

  The prototype two-stage power amplifier with SDDB was fabricated in the 0.25\(\mu\)m IBM BiCMOS 6HP process [123]. The die area is 1mm×1.8mm. A packaged die with bonding wires is shown in Figure 5.3. The chip is intentionally put to the right side of the package since reducing the output parasitic inductance is critical. All grounding down-bonds are made as short as possible to reduce parasitic inductors.

\(^5\) MLF is a licensed trademark of Amkor Technology.
Figure 5.2 Cross section of the Micro Lead Frame package.

Figure 5.3 Photograph of the packaged die of the prototype power amplifier with SDDB.

• Board Generating and Mounting

The next step is to make the PCB board, on which the packaged chip as well as other off-chip components will be mounted. Rogers 4003\(^6\) was selected as the board material, since it has more stable dielectric constant, much less insertion loss, more thermal conductivity and excellent power handling capacity. These electrical

\(^6\) Rogers 4003 is a licensed trademark of Rogers Corporation.
characteristics are essential for RF/microwave applications [124]. The photograph of the PCB board with the package mounted is shown in Figure 5.4.

![Photograph of the test PCB board with the package.](image)

**Figure 5.4 Photograph of the test PCB board with the package.**

- **Test Bench Setup**

  As shown in Figure 5.5, the test bench of the power amplifier with SDDB is composed of a signal generator, a spectrum analyzer, dc power supplies, and the device-under-test (DUT).
5.2 Measured Performance

Two prototype power amplifiers with SDCB and SDDB respectively have been fabricated and measured. Their design objective is to achieve far better average power efficiency with a constant power gain as well as satisfy the specifications of WCDMA Class 3 handset power amplifiers. The performance metrics include dc power consumption, power gain, power added efficiency, and ACPR.

5.2.1 Design with Switched Dynamic Current Biasing

The design with SDCB is implemented in a single-stage prototype power amplifier. Figure 5.6 compares the measured DC currents for different biasing approaches for the single-stage power amplifier, superimposed on a representative probability distribution function for the output power [46]. These approaches include
constant base voltage (CV) biasing with a fixed number of parallel transistors and SDCB. Using (2.1), average power efficiencies are calculated as 2.5% for CV biasing and 8.0% for SDCB – a substantial improvement with the new approach.

Figure 5.7 shows measured power added efficiencies with SDCB and CV approaches. The peak PAE is not as high as other III-V WCDMA amplifiers reported [125], but the average power efficiency is improved. Table 1.1 is a summary of average power efficiencies for different reported dynamic biasing techniques, including this SDCB work.

![Figure 5.6 Output power probability distribution P_e and measured DC current for different biasing techniques. The switch point from 100 devices to 20 devices occurs at Pout = 18 dBm. V_{cc}=3V.](image_url)
Figure 5.7 Measured power added efficiencies with constant voltage with fixed area (CV) and switched dynamic current biasing with varied area (SDCB) power amplifiers.

Table 5.1 Average power efficiency comparison of reported dynamic biasing techniques.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Fixed biasing efficiency</th>
<th>Dynamic biasing efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaAs MESFET PA with DVB(^1) [24]</td>
<td>3.89%</td>
<td>6.38%</td>
</tr>
<tr>
<td>AlGaAs/InGaAs MESFET PA with DVB [53](^2)</td>
<td>2.2%</td>
<td>11.4%</td>
</tr>
<tr>
<td>LDMOS PA with DVB [126](^2)</td>
<td>1.53%</td>
<td>6.78%</td>
</tr>
<tr>
<td>GaAs HBT PA with DCB(^3) [60]</td>
<td>2.50%</td>
<td>4.03%</td>
</tr>
<tr>
<td>SiGe HBT PA with SDCB in this work</td>
<td>2.50%</td>
<td>8.08%</td>
</tr>
</tbody>
</table>

\(^1\)Dynamic voltage biasing.
\(^2\)Output power distribution probability function used in [53] and [126] is different from that used in this work.
\(^3\)Dynamic current biasing.
Figure 5.8 compares the measured gain variation between SDCB with varied HBT area, DCB with fixed HBT area and CV with fixed HBT area. The gain change for SDCB with varied HBT area is less than 2dB, and is much more constant than DCB with fixed HBT area [59]-[60].

The linearity of the SDCB amplifier is measured under ACPR testing with a WCDMA signal. The corresponding simulated ACPR curve is not compared here because the simulation of ACPR in Cadence design environment is very time-consuming and the ACPR may be readily derived from IMR3 [127]-[128]. As shown in Figure 5.9, the circuit satisfies the 3GPP WCDMA Class 3 ACPR specification with 23.9 dBm channel output power.

Figure 5.8 Measured power gains with constant voltage with constant area (CV), SDCB with varied HBT area and DCB with fixed HBT area power amplifiers.
5.2.2 Design with Switched Dual Dynamic Biasing

The design with SDDB is implemented in a two-stage prototype power amplifier. Figure 5.10 compares the measured DC currents for different biasing approaches for the two-stage power amplifier, superimposed on a representative probability distribution function for the output power.
Figure 5.10 Output power probability distribution function $P_e$ and measured DC current comparison for different biasing techniques (CV: constant voltage biasing; SDDB: switched dual dynamic biasing). The switching point from high-power mode to low-power mode is 16 dBm.

Average power efficiencies are calculated from (2.1) to be: 1.9% for CV, 3.8% for SDCB, and 5.0% for SDDB, all in two-stage PAs. This verifies that SDDB does achieve the optimal average efficiency for WCDMA applications.

As discussed in the previous chapter, digital predistortion (DP) is employed in the SDDB amplifier to improve its linearity. The ACPR measurement at 5 MHz is shown in Figure 5.11. Note that it is only necessary to utilize the digital predistortion in the high-power mode, so DP is not applied in the low-power mode; this is the reason for the large discontinuity in ACPR in the “after DP” curve. The ACPR is improved by at least 8dB with digital predistortion, and the maximum output power satisfying the WCDMA linearity specification is improved from 22.4dBm to 26dBm. This satisfies the WCDMA Class 3 requirement of maximum output power.
Figure 5.11 Measured ACPRs of the SDDB power amplifier with digital predistortion (DP) (before DP and after DP).

Figure 5.12 shows the measured gain and power added efficiency (PAE) of the SDDB power amplifier with DP. The gain change for SDDB is less than 1.8dB, which is much more constant than the case if the power amplifier is operated with dynamic bias without changing the device size. Since the linear maximum output power increases by 3.6dB (from 22.4 to 26 dBm), the peak PAE is improved by 60% (from 17% to 27%).
5.3 Summary

The whole post-tapeout procedure has been illustrated through the design with switched dual dynamic biasing. The PCB testing is employed for realistic and repeatable measurements. Experimental results have shown that both prototype power amplifiers with dynamic biasing techniques can achieve far better average power efficiency, keep power gain almost constant, and satisfy the design specifications of WCDMA Class 3 handset power amplifiers described in the first chapter.

This chapter, in part or in full, is a reprint of the materials in 2004 IEEE Radio Frequency Integrated Circuits Symposium, 2005 IEEE Transactions on Microwave Theory and Techniques, and 2005 IEEE Radio Frequency Integrated Circuits Symposium. The author of this dissertation was the primary researcher and the first author in these publications.

Figure 5.12 Measured gain and power added efficiency (PAE) of the SDDB power amplifier with DP.
CHAPTER 6
Conclusions

For WCDMA handset power amplifiers, efficiency and linearity are the most critical design parameters. Moreover, average power efficiency, instead of peak power efficiency, is the key factor determining the talk time and battery life for portable wireless applications. In this dissertation, switched dynamic biasing techniques have been proposed to boost the average power efficiency. The linearity of high-power amplifiers has been analyzed through a varied-coefficient Volterra series. Digital predistortion has been employed to improve the linearity. This chapter will summarize the key research results achieved in this dissertation and suggest directions for future research.

6.1 Key Research Results

- **Efficiency Enhancement**

  Switched dynamic biasing techniques achieve far better average power efficiency than typical Class AB operation. To overcome the gain-varying issue in previous dynamic biasing techniques and reduce the burden on the power control in the CDMA operation, power gain is kept roughly constant by maintaining a constant
collector current density for each active device. Two prototype power amplifiers with switched dynamic current biasing and switched dual dynamic biasing respectively have been implemented in a 0.25\(\mu\)m SiGe BiCMOS process. Besides realizing better average power efficiency and constant power gain, both power amplifiers satisfy the specifications of WCDMA Class 3 handsets.

Switched dynamic current biasing improves the average power efficiency of a single-stage design from 2.5\% (with Class AB operation) to 8.0\%. The power amplifier satisfies the ACPR specification for uplink with 23.6dBm channel output power, and the corresponding peak PAE is 31\%.

Switched dual dynamic biasing boosts the average power efficiency of a two-stage design from 1.9\% (with Class AB operation) to 5.0\%. The measured maximum output power is 26dBm, and the peak PAE is 27\%.

The above experimental results have confirmed the effectiveness of proposed switched dynamic biasing techniques. Compared to other dynamic biasing techniques, they are more favorable because of their great integration, constant power gain, and better average power efficiency.

- **Linearity Analysis and Improvement**

A power-dependent coefficient Volterra technique has been utilized to analyze the linearity of power amplifiers. The nonlinear model of SiGe HBT power amplifiers has been derived and the analyzed model takes into account the effect of emitter bond-
wire inductance. The main sources of nonlinearity of SiGe HBT power amplifiers have been highlighted. Furthermore, a three-tone Volterra model for “Nonlinear Current Method” has been derived and applied to the analysis of an adaptive power amplifier with envelope signal injection. The comparison between analysis, simulation, and measurement data has validated the above models.

Digital predistortion features moderate linearization performance, wide compensation bandwidth, convenient integration, and economical implementation. A complex-gain digital predistortion system has been implemented to improve the linearity of the design with switched dual dynamic biasing. Experimental results demonstrated that the overall ACPR improvement with the digital pre-distortion is more than 8dB and the peak power added efficiency is improved by 60%.

### 6.2 Directions for Future Research

With the advancement of wireless communication systems, the need to accommodate multiple communication standards in a single handset becomes urgent. Since the power amplifier is the major block limiting the efficiency and linearity of the transmitter, the power amplifier must be wide-band as well as power-efficient and highly linear. This imposes more challenges on the power amplifier design. It would be interesting to investigate the effectiveness of switched dynamic biasing control and digital predistortion to support multiple-mode applications.

Envelope signal injection has been demonstrated to be an effective linearization technique in an adaptive low-power amplifier. However, in order to optimize both
high-side and low-side intermodulation products simultaneously, the phase of the feedback envelope signal has to be calibrated precisely in the silicon implementation. Therefore, it would be intriguing to examine the practicality and feasibility of envelope signal injection as an integrated and low-cost linearization solution in commercial power amplifiers.
References


[4] [http://www.3gpp.org](http://www.3gpp.org)


[46] P. Asbeck, University of California at San Diego, La Jolla, CA, private communication, April 2003.


