Analysis and Design of Monolithic Radio Frequency Linear Power Amplifiers

by

Burcin Baytekin

B.S. (University of Southern California) 1997
M.S. (University of California, Berkeley) 1999

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in

Engineering - Electrical Engineering and Computer Science in the

GRADUATE DIVISION of the UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor Robert G. Meyer, Chair
Professor Ali M. Niknejad
Professor John A. Strain

Spring 2004
The dissertation of Burcin Baytekin is approved:

Chair

Date

Date

Date

University of California at Berkeley

Spring 2004
Analysis and Design of Monolithic Radio Frequency Linear Power Amplifiers

Copyright Spring 2004

by

Burcin Baytekin
Abstract

Analysis and Design of Monolithic Radio Frequency Linear Power Amplifiers

by

Burcin Baytekin

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Science

University of California at Berkeley

Professor Robert G. Meyer, Chair

Linear power amplifiers (PAs) are becoming widely used in modern wireless communication systems. The envelope of the signals in these systems is typically not constant, so that the PA design must address the issue of device nonlinearity in order to limit the amount of spectral regrowth, which can cause unacceptable levels of interference in the adjacent channels.

This research focuses on a new computational method for efficiently analyzing the relationship between spectral regrowth and physical distortion mechanisms in radio frequency power amplifiers. It utilizes a Volterra series model whose coefficients are computed from basic SPICE parameters. The analysis uses a decomposition of the Volterra kernels into simpler subsystems in order to greatly reduce the computation time.

The new computational approach was applied to the design of PAs using bipolar transistors as the active element. A series-based model was developed for representing the
increase in active-device forward transit time at high collector current densities.

The design of a number of single-stage SiGe power amplifiers is also described. Bias techniques are described which allow the average PA current to increase at high signal levels and thus reduce distortion-inducing gain compression. The amplifiers were tested using the IEEE802.11b and IS-95 modulation schemes at different carrier frequencies and these results are compared with the theoretical analysis.

Professor Robert G. Meyer
Dissertation Committee Chair
To my parents,

Sevim and Ali Baytekin,

without their support I could not be where I am today
# Contents

List of Figures

List of Tables

I Power Amplifier Fundamentals

1 Introduction

1.1 Motivation ........................................... 2
1.2 Previous Work ...................................... 4
1.3 Research Goals .................................... 7
1.4 Outline ............................................ 9

2 Power Amplifier Design

2.1 Important PA Specifications .......................... 11
2.2 Classes of Operation ................................ 14
   2.2.1 Class A Amplification .......................... 15
   2.2.2 Class B Amplification .......................... 18
   2.2.3 Class AB and C Modes .......................... 20
2.3 Supply Voltage, Breakdown and Efficiency ...... 23
2.4 Output Matching .................................... 25
2.5 Input and Interstage Matching ........................ 28
2.6 Thermal management ................................ 30
2.7 Packaging .......................................... 33
2.8 Stability ........................................... 37
2.9 Electrostatic Discharge .............................. 39

3 Bipolar Transistor Technologies .................... 43

3.1 Gallium-Arsenide Heterojunction Bipolar Transistors .............. 43
3.2 Silicon-Germanium Bipolar Transistors .......................... 45
3.3 Thermal Runaway .................................... 47
II Analysis 49

4 Volterra Series 50
4.1 Calculating the Volterra Kernels of a Circuit 51

5 Computational Approach 56
5.1 Time and Frequency-Domain Calculations 56
5.2 Decomposition of the Volterra Kernels 58
5.3 Third-Order Volterra Subsystems 60
  5.3.1 Second-Order Interaction Subsystem 60
  5.3.2 Pure Third-Order Subsystem 62
5.4 Neglecting Even-Order Terms in Volterra Series 63

6 Device Modeling 66
6.1 Modeling the Variations in Forward Transit Time 70

III Application 75

7 Implementation 76
7.1 Specifications 76
7.2 Details of the Circuit 77
7.3 Kernel Calculations 82
7.4 Simulator 86

8 Results 89
8.1 Analysis versus Measurements 89
8.2 Observations 93

9 Conclusion 96
9.1 Future Research 98

Bibliography 100

A Programs for Implementing the Computational Approach 107
A.1 Generating the Input Signal 107
A.2 Spectral Regrowth Calculations 112
  A.2.1 Entering the Design and Model Parameters 117
  A.2.2 Scaling the Input Signal Waveform 122
  A.2.3 Calculating the Input and Output Loss 123
  A.2.4 Calculating the Third-Order Response 125
  A.2.5 First-Order Response Functions 128
  A.2.6 Second-Order Transfer Functions 133
  A.2.7 Third-Order Transfer Functions 137
List of Figures

1.1 A typical wireless transmitter ........................................... 3
1.2 Intermodulation distortion .............................................. 5

2.1 Simplified power amplifier schematic ................................ 14
2.2 Current and voltage waveforms for Class A amplifier ............... 16
2.3 The maximum collector efficiency of an ideal Class A amplifier versus the load resistance $R_L$ normalized to the optimum load resistance $R_{L_{\text{opt}}}$ .... 18
2.4 Current and voltage waveforms for Class B amplifier ............... 19
2.5 The efficiency of ideal Class A and Class B amplifiers versus output power level normalized to the maximum available power. ................... 21
2.6 The maximum efficiency of an ideal power amplifier versus the conduction angle. .................................................. 22
2.7 L-section output matching networks (a) Low-pass (b) High-pass . 27
2.8 L-section input or interstage matching networks (a) Low-pass (b) High-pass . 29
2.9 Band-pass matching network .......................................... 29
2.10 Heat Dissipation Paths for IC Packages ............................. 31
2.11 Thermal Resistance Equivalent Circuit ................................ 32
2.12 A flip-chip package ................................................... 35
2.13 Cross section of a flip-chip package with active and inactive solder bumps . 36
2.14 General two-port amplifier model .................................... 38
2.15 Simplified human body model for electrostatic discharge .......... 39
2.16 Grounded-base-npn clamp .......................................... 40
2.17 ESD current paths .................................................. 41

4.1 (a) Nonlinear transconductor model based on power series (b) The equivalent model of the same transconductor for calculating Volterra kernels .... 52

5.1 Third-order Volterra subsystems (a) Pure third-order subsystem (b) Second-order interaction subsystems ........................................ 59
5.2 Compressed spectrum .................................................. 59

6.1 Nonlinear bipolar transistor model .................................... 67
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2</td>
<td>(a) Electric field and (b) space-charge density in the collector-base junction including the effects of mobile electrons due to collector current</td>
<td>71</td>
</tr>
<tr>
<td>7.1</td>
<td>IEEE802.11b transmit spectrum mask</td>
<td>77</td>
</tr>
<tr>
<td>7.2</td>
<td>Die photo</td>
<td>78</td>
</tr>
<tr>
<td>7.3</td>
<td>Simplified schematic of the power amplifier</td>
<td>79</td>
</tr>
<tr>
<td>7.4</td>
<td>Conventional local-bias circuit</td>
<td>81</td>
</tr>
<tr>
<td>7.5</td>
<td>Local-bias circuit without base resistor</td>
<td>81</td>
</tr>
<tr>
<td>7.6</td>
<td>Norton equivalent of the schematic of the power amplifier in figure 7.3</td>
<td>83</td>
</tr>
<tr>
<td>7.7</td>
<td>The schematic for calculating the first-order response of the power amplifier presented in figure 7.6</td>
<td>84</td>
</tr>
<tr>
<td>7.8</td>
<td>The schematic for calculating the second and third-order responses of the power amplifier in figure 7.6</td>
<td>85</td>
</tr>
<tr>
<td>7.9</td>
<td>Simulation method</td>
<td>87</td>
</tr>
<tr>
<td>7.10</td>
<td>Input and output power spectral density calculated by simulations</td>
<td>88</td>
</tr>
<tr>
<td>8.1</td>
<td>Ratio of the first sidelobe and the mainlobe versus output power level for ( m = 78, V_{CC} = 3.3 \text{ V}, I_{cQ} = 196 \text{ mA}, f_c = 2.45 \text{ GHz} )</td>
<td>90</td>
</tr>
<tr>
<td>8.2</td>
<td>Ratio of the first sidelobe and the mainlobe versus output power level for ( m = 104, V_{CC} = 3.3 \text{ V}, I_{cQ} = 196 \text{ mA}, f_c = 2.4 \text{ GHz} )</td>
<td>91</td>
</tr>
<tr>
<td>8.3</td>
<td>Ratio of the first sidelobe and the mainlobe versus output power level for ( m = 78, V_{CC} = 3.5 \text{ V}, I_{cQ} = 176 \text{ mA}, f_c = 2.0 \text{ GHz} )</td>
<td>92</td>
</tr>
<tr>
<td>8.4</td>
<td>Ratio of the first sidelobe and the mainlobe versus current consumption at ( P_{out} = 24 \text{ dBm} ) for ( m = 78, V_{CC} = 3.5 \text{ V}, f_c = 2.0 \text{ GHz} )</td>
<td>92</td>
</tr>
<tr>
<td>8.5</td>
<td>ACPR versus output power level for IS-95 modulation, ( m = 78, V_{CC} = 3.5 \text{ V}, I_{cQ} = 176 \text{ mA}, f_c = 2.0 \text{ GHz} )</td>
<td>93</td>
</tr>
<tr>
<td>8.6</td>
<td>Electric field and the variation of length of the collector-base space charge region for (a) Low collector dopant density (b) High collector dopant density (c) Shallow buried layer</td>
<td>94</td>
</tr>
</tbody>
</table>
List of Tables

3.1 Properties of Si and GaAs at 300 K [35] .......................... 44
Acknowledgements

I am grateful for the support and guidance of my advisor, Prof. Robert G. Meyer, who has introduced me to the world of communication circuits. His deep knowledge about analog circuit design and his enthusiastic teaching style has been a source of inspiration. I would also like to thank Prof. Ali M. Niknejad and Prof. John A. Strain for carefully reviewing my thesis. I am also thankful for Prof. Bernhard E. Boser for taking part in my qualifying exam committee.

My six years at Berkeley gave me the opportunity to interact with many great faculty members. I thank Prof. Schwarz for teaching the microwave circuits course. I thank Prof. Boser once again for his invaluable analog and mixed signal classes. Thanks to Prof. C. Hu for teaching about CMOS devices so clearly and thanks to Prof. Tse for developing his insightful course about wireless communication systems.

I would like to thank Maxim Integrated Products for the assistance with manufacturing, packaging and testing the power amplifiers described in this thesis. I am also grateful to Joel King for invaluable technical discussions about power amplifier design.

I have learned a lot during my summer internships. I would like to thank Mehmet Soyuer of IBM T.J. Watson Research Center for giving me the opportunity to be part of a distinguished research laboratory. I am also grateful to have Julian Tham, who is an impressive circuit designer and a great supervisor, as a colleague during three summers at two different companies, Rockwell Semiconductor Systems and Maxim.

My stay in Berkeley was enhanced by many friends. I am fortunate to be close friends with M. Cenk Cavusoglu, whose “1-800-MATLAB” was a great source of assistance.
I would also like to thank Tunc Simsek, who shared an apartment with me for five years and amazed me more than a few times when he easily showed me a way out of my linear algebra problems. I am grateful to Steven Rose for demonstrating what would happen to me if I did not start cutting back on my daily Coke intake, while we were sharing an apartment in Sunnyvale for three months. I thank Harun Bayraktar for sharing the load of the two graduate level mathematics classes we took together. I also thank the other members of Prof. Meyer’s research group, Sang Won Son, Manolis Terrovitis, John Wetherell, Konstantin Kouznetsov, Henry Jen, Hakan Dogan and Axel Berny for their technical advice and friendly conversations. I would also like to thank Adam Eldredge, Turgut “Turi” S. Aytur, Luns Tee, Boris Murmann, Vladimir Petkov, Ken Wojciechowski and Naratip Wongkomet for making life at 550 Cory Hall more colorful.

The last two years of my life became much more enjoyable thanks to Melike Gunalp and her love. I hope she will be part of my life for years to come.

I would also like to express my most special gratitude to my parents for their never-ending love and support. They made it possible for me to become who I am today.
Part I

Power Amplifier Fundamentals
Chapter 1

Introduction

1.1 Motivation

The convenience of cell phones is attracting growing number of consumers worldwide. They are already so popular that they come to be seen as fashion accessories by many. As the cost of wireless transceivers are decreasing, they find their way into more consumer electronic products. In the meantime, wireless local area networks (WLAN) are becoming very popular among those who do not want to give up the mobility of a laptop computer while accessing the Internet. The ease of implementation of WLANs compared to a wired solution is leading companies to utilize them in office buildings as well. Wireless communication systems are becoming viable alternatives for service providers trying to avoid the cost of wiring the so called “last mile” into urban homes or upgrading their wired systems in rural locations with very low density housing.

The growing demand for wireless communications require an efficient use of the limited wireless spectrum. At first this led the service providers to switch from analog mo-
bile phone systems to time division multiple access (TDMA) based digital communication systems, such as “Groupe Speciale Mobile” (GSM) cellular telephone standard. Then, the decreasing price and size, as well as the increasing performance of digital signal processors allowed computationally more demanding, but more spectrally efficient wireless transmission schemes, which use code division multiple access (CDMA) or orthogonal frequency division multiplex (OFDM) signals.

The rising popularity of these wireless communication systems is fueling interest in wireless transceiver circuits. One of the important circuit blocks in a wireless transmitter is the power amplifier (PA). This is the last amplification stage before the antenna, as shown in figure 1.1. A power amplifier has to supply all of the radiated power at the transmitter antenna, as well as the power lost through the passive elements, such as radio frequency (RF) filters or passive duplexers. This makes the PA efficiency the dominant factor in the total power dissipation of the radio transmitter, which is especially significant for mobile
communication applications.

At the same time, linearity is becoming probably the most stringent specification, as more wireless systems start to utilize CDMA and OFDM. The envelope of the signals in these systems are not constant, so they require the use of linear PAs in order to prevent the radiated signals from being distorted. Still, numerous sources of nonlinearity in a PA may cause an excessive amount of spectral regrowth, the generation of undesired signals in frequencies outside of the wireless channel in use. If unchecked, spectral regrowth may cause unacceptable levels of interference in the adjacent channels. Nonlinear distortion may also cause errors in the transmitted data symbol.

1.2 Previous Work

The trade-off between efficiency and linearity leads PA designers to search for an optimum device operating point. Therefore, a good understanding of the effects of the transistor components and PA design parameters on linearity is essential. Linearity has long been analyzed by using the third-order intermodulation distortion ($IM_3$) as a metric, where the input is composed of two sinusoidal waveforms. The output of the circuit consists of the amplified version of these sine waves together with the nonlinear distortion products. $IM_3$ is defined as the ratio between the amplitude of the nonlinear products whose frequency is closest to the desired output signals, as shown in figure 1.2. However, wireless transmission standards generally specify the minimum acceptable linearity in terms of the transmit spectrum mask, adjacent channel power ratio (ACPR) or error vector magnitude (EVM). These specifications require an accurate analysis or simulation to take into account
the properly modulated input signal, instead of just two sinusoidal waveforms.

The phenomenon of spectral regrowth has been analyzed in recent publications, but these treatments employ empirical methods that require fitting a curve to some measured or simulated characteristics [22]. The most common one is the AM-AM transfer characteristics, which consist of a set of output power versus input power measurements using a single tone sinusoidal waveform. Another one is the AM-PM characteristics, which consist of the amount of phaseshift through the PA versus input power level.

One of these treatments is from Chen et al.. They have used the AM-AM and AM-PM measurement data for predicting spectral regrowth and power gain characteristics of a PA operated with signals modulated with the IS-95 cellular phone standard [7]. The power amplifier is represented by a complex baseband equivalent model. The input is expressed as

\[ x(t) = a(t)e^{j\phi(t)}e^{j2\pi f_c t} \]  \hspace{1cm} (1.1)

and the output is given by

\[ y(t) = Re \left[ G(a(t))e^{j(\phi(t)+P(a(t)))}e^{j2\pi f_c t} \right] \]  \hspace{1cm} (1.2)
where \( f_c \) is the carrier frequency, \( G(a) \) is the single tone AM-AM gain and \( P(a) \) is the AM-PM phaseshift function. The memory effects in the PA are neglected in this model. The time-domain output is calculated for each input sample and discrete fourier transform (DFT) is used to obtain the output spectral density. Pinsky uses a very similar method for the IS-54 signals by using cubic interpolation between adjacent data points to calculate \( G(a) \) and \( P(a) \) functions [39]. Struble et al. apply linear interpolation instead and perform a similar analysis for the IS-95 and personal handy phone (PHS) signals [47].

Leke and Kenney have used AM-AM and AM-PM measurements to compute the coefficients of a third-order Volterra series [28]. However, they assume that the Volterra kernels are constant across frequency, which reduces the Volterra series to a complex power series and neglects memory effects.

Wu et al. propose a closed-form equation to calculate spectral regrowth from the third-order intercept point (\( IP_3 \)) of a circuit [52]. However, they use a real third-order baseband equivalent power series to model an amplifier, which ignores AM-PM effects. They also assume that the CDMA input signal can be represented by a single real random variable, while the actual baseband equivalent signal has two complex variables, one each for I and Q channels. Furthermore, their approach assumes that the CDMA waveform has constant power spectral density inside the channel, although the signal spectrum is shaped by a low pass filter specified by the IS-95 standard. Besides, the input waveform is assumed to be Gaussian, but the actual signal is based on the transmission of a set of discrete symbols with equal probability.

Gard et al. have improved on the approach of Wu et al. by representing the input
by two complex input waveforms and removing the flat spectrum requirement [11]. The spectrum is shaped by the FIR filter whose coefficients are determined by the IS-95 standard. However, they still assume that the input random variables have Gaussian distribution and neglect memory effects. First, they fit a 13th order complex polynomial to the AM-AM and AM-PM measurements. They use this baseband equivalent series model and an algebraic formula for calculating the coefficients of an instantaneous power series model [46]. The power spectral density of the output waveform is then calculated from the moments of the input complex Gaussian random variables. Zhou has proposed an improvement by using cumulants (generalizations of the higher order moments) in order to remove the Gaussian distribution assumption [53].

Although all of these methods claim varying degrees of success in predicting spectral regrowth by using single tone measurement data, none of them attempt to investigate the relationship between spectral regrowth and transistor components or other circuit elements. Therefore, they do not provide much insight for designers trying to optimize power amplifiers.

1.3 Research Goals

In addition to the lack of theoretical analysis, PA designers are also hampered by CAD tools poorly suited for simulating spectral regrowth. The input signal of a typical communication system does not take on deterministic values, but is composed of a signal modulated by random data and a specified modulation scheme, requiring the generation of large number of bits for a proper simulation. Furthermore, the carrier frequency in wireless
communication systems can be 2 to 4 orders of magnitude larger than the bandwidth of the information-carrying signal or the envelope. Therefore, even if just a few samples per each carrier period is used, a computation in the time domain requires too many samples per bit of data for practical circuit simulations.

The lack of analysis and proper CAD tools leads designers to using rule-of-thumb methods for finding the optimum device operating point. These methods usually involve single or two-tone test simulations, although there is no simple relationship between the results of these tests and ACPR type specifications. Thus, it is important to investigate the relationship between the sources of nonlinearity in a PA and spectral regrowth.

In this thesis, a new method is proposed for predicting spectral regrowth in PAs. The method uses basic SPICE parameters, which are based on the active device physical mechanisms in order to make it applicable to transistors fabricated by different processes. First, a Volterra series model of the PA is calculated from the SPICE parameters, and spectral regrowth is then predicted by using modulated signals. The decomposition of the Volterra kernels into simpler subsystems as proposed in chapter 5 allows the combination of frequency and time-domain computations so that numerical results can be rapidly calculated. These results can assist circuit designers in understanding the effect of PA design parameters and transistor components on spectral regrowth and the trade-off between efficiency and linearity. This would allow them to determine their initial design parameters more accurately before they start their detailed simulations, helping them avoid time-consuming iterations. Identifying the transistor components contributing to distortion would also help semiconductor device designers optimize power transistors.
Efficiency and linearity are not the only factors in the design of PAs. Size and cost considerations are behind the drive to implement them in cheaper silicon-based technologies and to integrate them with the rest of the transmitter. Therefore, the PAs designed and fabricated for comparing the measurements with the theoretical results are implemented by using silicon-germanium (SiGe) bipolar transistors. In order to analyze the bipolar transistors more accurately, a series-based model is developed for representing the increase in active device forward transit time at high collector current densities.

1.4 Outline

Chapter 2 of this thesis describes important specifications of a power amplifier, introduces some of the classes of PA operation and outlines other key aspects of linear RF PA design.

Chapter 3 briefly presents the basic operating principles of Gallium-Arsenide (GaAs) heterojunction bipolar transistors (HBT) and compares their performance with Silicon (Si) BJTs when used in power amplifiers. Then it describes the performance improvements provided by Silicon-Germanium (SiGe) BJTs along with the thermal instability problems encountered in different bipolar transistor technologies.

Chapter 4 introduces nonlinear system models with an emphasis on Volterra series and describes a practical method of calculating the Volterra kernels of a circuit.

Chapter 5 presents the problems associated with computing the output of a circuit represented by a nonlinear model in response to a modulated RF signal. Afterwards, it describes the proposed method of decomposition of the Volterra kernels into simpler
subsystems in order to greatly reduce the computation time.

Chapter 6 explains how to generate device models so that the new computational approach can be applied to the design of PAs using bipolar transistors as the active element. It presents the proposed series-based model for representing the increase in active device forward transit time at high collector current densities as well.

Chapter 7 presents an implementation example, where a number of single stage SiGe power amplifiers have been designed and fabricated in order to compare the theoretical results with the measurements. First, it outlines important PA specifications of the IEEE802.11b and IS-95 modulation schemes. Then, it discusses the details of the circuit and describes how the Volterra kernels of this particular circuit are calculated. The operating principles of the program implementing the computational approach are explained at the end.

Chapter 8 presents the results of the analysis and measurements taken from the single-stage PAs using the IEEE802.11b and IS-95 modulation schemes at different carrier frequencies.

Chapter 9 concludes the thesis and identifies possible future applications of the new computational approach.

The appendix consists of the programs written for implementing the analysis method presented in the thesis.
Chapter 2

Power Amplifier Design

2.1 Important PA Specifications

It is important to outline the main PA specifications and their definitions before main aspects of PA design are reviewed. As the name power amplifier suggests, two important specifications are the maximum output power $P_{\text{out max}}$ and power gain, which is defined as

$$A_P = \frac{P_{\text{out}}}{P_{\text{in}}} \quad (2.1)$$

where $P_{\text{in}}$ is the average power delivered into the PA and $P_{\text{out}}$ is the average power delivered to the load. The computation of power requires both voltage and current, so the source and load (antenna) impedances have to be specified along with $P_{\text{out max}}$ and $A_P$. Most wireless communication systems use the standard impedance of 50 $\Omega$.

If the signal is a sinusoidal waveform and the power is delivered into a resistor,
then average power can be calculated by

\[ P = \frac{\hat{v}\hat{i}}{2} = \frac{\hat{v}^2}{2R} = \frac{\hat{i}^2 R}{2} \]  \hspace{1cm} (2.2)

where \( \hat{v} \) and \( \hat{i} \) are the amplitudes of voltage and current swing. During simulations, it is often necessary to monitor the power gain performance of the circuit between some of the internal nodes of the amplifier. The impedance at these nodes usually have a complex part, so (2.2) would not be accurate. Instead, the average power delivered at angular frequency \( \omega \) can be calculated by

\[ P = \frac{1}{2} \text{Re}\{V(j\omega)I(j\omega)^*\} = \frac{1}{2} |I(j\omega)|^2 \text{Re}\{Z(\omega)\} = \frac{|V(j\omega)|^2}{2 \text{Re}\{Z(\omega)\}} \]  \hspace{1cm} (2.3)

where \( ^* \) denotes complex conjugate, \( Z(\omega) \) is the impedance at angular frequency \( \omega \), \( V(j\omega) \) is the Fourier transform of the voltage waveform across the input port of the circuit block being probed and \( I(j\omega) \) is the Fourier transform of the current waveform flowing into that port.

If an accurate reproduction of a time-varying signal envelope is required, PA linearity becomes very important. Although linearity can be measured through the third-order intermodulation distortion \( IM_3 \), defined in the previous chapter, the digital communication systems usually employ other specifications. For example, transmit spectrum mask specifies the maximum power spectral density of the signal radiated from the antenna. Adjacent channel power ratio (ACPR) is defined as the ratio of the power transmitted in a band of frequency outside of the desired channel to the power inside the channel. Both of these are used to limit the interference in the adjacent channels, so that the performance of the communication system does not degrade. In addition, there are other specifications, such
as error vector magnitude (EVM) or rho ($\rho$), in order to make sure the transmitted signal is modulated accurately so that the errors are minimized when the receiver demodulates the signal [16].

As the current consumption of the PA is considerably more than the other circuit blocks in most transmitters, PA efficiency is very important. A simple measure of efficiency is called the collector or drain efficiency and they are defined as

$$\eta = \frac{P_{\text{out}}}{P_{\text{DC}}}$$

(2.4)

where $P_{\text{DC}}$ is the power drawn from the voltage supply.

Higher power gain reduces the required amount of input power supplied by the circuit driving the PA and allows lower current consumption for that driver. This increases the overall efficiency of the transmitter. Therefore, a good figure of merit for comparing the efficiency of different PAs should take power gain into account as well. This makes power-added efficiency, defined as

$$\text{PAE} = \frac{P_{\text{out}} - P_{\text{in}}}{P_{\text{DC}}}$$

(2.5)

a very useful metric.

The trade-off between efficiency and linearity requires a good understanding of PA operation. Therefore, in the next section, the classes of PA operation will be reviewed in order to illustrate this trade-off.
2.2 Classes of Operation

There are various classes of operation suitable for radio frequency power amplifiers. The switching PAs forming Classes D and E can be designed to provide amplification with very high efficiency, but they cannot accurately reproduce the envelope of the input signal without complicated linearization schemes [23, 48].

The basic RF amplifier circuit shown in figure 2.1 can be biased for operation in Class A, AB, B or C modes, which trade-off linearity for efficiency. When operated in Class A mode, the transistor $Q_1$ conducts current throughout the RF cycle, while it is turned off during half of the RF cycle in Class B mode. If the transistor is conducting during less than 50% of each cycle, it enters Class C mode. If it is conducting between 50 and 100% of each RF cycle, the amplifier is classified as Class AB.
2.2.1 Class A Amplification

There is no sharp dividing line between Class A PAs and small-signal amplifiers. They usually have very similar configurations, although the collector current swing and output voltage swing of a PA are comparable to the quiescent values in order to maximize efficiency.

In the amplifier of figure 2.1 the RF-choke is a large inductor which passes the bias current, but has high impedance at radio frequencies so that almost all of the signal flows into the load through the large AC coupling capacitor $C_c$. The parallel tuned circuit formed by $C_T$ and $L_T$ is not a necessary part of a Class A PA, but it can be used to filter out the high frequency harmonics.

The collector current and output voltage of typical a Class-A power amplifier during an RF cycle is shown in figure 2.2. Assuming the amplifier is perfectly linear and the input is sinusoidal, the collector current is given by

$$I_c(t) = I_{CQ} + \hat{i}_c \sin \omega t$$

(2.6)

where $I_{CQ}$ is the quiescent current and $\hat{i}_c$ is the amplitude of the collector current swing. The corresponding collector voltage is

$$v_c(t) = V_{CC} - \hat{v}_c \sin \omega t$$

(2.7)

where $V_{CC}$ is the supply voltage and $\hat{v}_c$ is the amplitude of the collector voltage swing.

The collector efficiency of this amplifier is

$$\eta = \frac{P_{out}}{P_{DC}} = \frac{\frac{1}{2} \hat{i}_c \hat{v}_c}{V_{CC} I_{CQ}}$$

(2.8)
Figure 2.2: Current and voltage waveforms for Class A amplifier
In order to prevent the transistor from going into saturation, the maximum $\hat{v}_c$ value should be kept smaller than $V_{CC} - V_{sat}$ where $V_{sat}$ is the saturation voltage of the transistor. At this operating condition if $\hat{\dot{i}}_c$ is equal to $I_{CQ}$, the efficiency reaches its maximum value and is given by

$$\eta_{max} = \frac{1}{2} \frac{V_{CC} - V_{sat}}{V_{CC}}$$

(2.9)

Therefore, the maximum efficiency of a Class A PA is 50% even if $V_{sat}$ is equal to zero.

The optimum load resistor required for achieving the highest efficiency is given by

$$R_{L_{opt}} = \frac{V_{CC} - V_{sat}}{I_{CQ}}$$

(2.10)

For a given supply voltage and bias current, the highest amount of power the amplifier can deliver to the load decreases if the actual $R_L$ is not equal to $R_{L_{opt}}$. If $R_L$ is smaller than $R_{L_{opt}}$ the output power is limited by the available current swing and the maximum output power is reduced to

$$P_{out_{max}} = \frac{I_{CQ}^2 R_L}{2}$$

(2.11)

If $R_L$ is larger than $R_{L_{opt}}$ the output power is limited by the available voltage swing and the highest output power drops to

$$P_{out_{max}} = \frac{(V_{CC} - V_{sat})^2}{2R_L}$$

(2.12)

As $P_{DC}$ in (2.8) does not depend on $R_L$, the collector efficiency decreases in either case. Figure 2.3 shows the highest achievable efficiency for different values of $R_L$ normalized to $R_{L_{opt}}$ assuming that $V_{sat}$ is negligible.

The efficiency will also be reduced when the amplifier is delivering less than $P_{out_{max}}$, as the power drawn from the battery stays constant regardless of output power.
Figure 2.3: The maximum collector efficiency of an ideal Class A amplifier versus the load resistance $R_L$ normalized to the optimum load resistance $R_{L_{opt}}$

2.2.2 Class B Amplification

When the amplifier of figure 2.1 is operating in Class B mode, the transistor is biased to have zero quiescent current and it conducts during only 50% of the RF cycle. Thus, the collector current of an ideal Class B PA with a sinusoidal input can be represented by

$$I_c(t) = \begin{cases} \hat{i}_c \sin \omega t, & \text{if } 0 < \omega t < \pi \\ 0, & \text{otherwise} \end{cases}$$

(2.13)

as shown in figure 2.4.

The parallel tuned circuit formed by $C_T$ and $L_T$ is designed to resonate at the angular frequency $\omega$ so that the higher harmonics contained in the collector current are shunted to ground and the output voltage is composed of a sinusoidal waveform at the
fundamental frequency.

If the collector current waveform is represented by a Fourier series, the amplitude of the collector current at the fundamental frequency $\omega$ is given by

$$ (i_\omega) = \frac{\hat{i}_c}{2} $$

and the output power can be calculated as

$$ P_{out} = \frac{\hat{i}_c^2 R_L}{8} $$

Average collector current can be calculated as

$$ I_{cave} = \frac{1}{T} \int_0^T I_c(t) \, dt = \frac{\omega}{2\pi} \int_0^{\pi} \hat{i}_c \sin \omega t \, dt = \frac{\hat{i}_c}{\pi} $$
The maximum output voltage swing is still $V_{CC} - V_{sat}$ which is achieved when

$$\left(\hat{i}_o\right) = \frac{V_{CC} - V_{sat}}{R_L}$$

Thus, the maximum efficiency of an amplifier operating in Class B mode is

$$\eta_{max} = \frac{1}{2} \frac{V_{sat} V_{CC} - V_{sat}}{V_{CC}} = \frac{\pi}{4} \frac{V_{CC} - V_{sat}}{V_{CC}}$$

which is the same for any $R_L$. This is a big advantage for Class B PAs over Class A amplifiers. The maximum efficiency of a Class B PA is about 78.5% if $V_{sat}$ is negligible.

The current consumption of a Class B amplifier in (2.16) is reduced when it is delivering less than $P_{out_{\text{max}}} = (V_{CC} - V_{sat})^2/2R_L$ while it stays the same for an amplifier operating in Class A mode [31]. Therefore, the efficiency of a Class B amplifier does not fall as rapidly as that of a Class A amplifier when the output power level is decreased. Figure 2.5 shows the relationship between the efficiency and $P_{out}$ normalized to $P_{out_{\text{max}}}$ for Class A and B amplifiers assuming that $V_{sat}$ is negligible.

### 2.2.3 Class AB and C Modes

The ideal collector current through the transistor in figure 2.1 can be modeled by

$$I_c(t) = \begin{cases} 
I_Q + \hat{i}_c \cos \omega t, & \text{if } -y/2 < \omega t < y/2 \\
0, & \text{otherwise}
\end{cases}$$

where $y$ is the conduction angle, which is $2\pi$ for a Class A amplifier, $\pi$ for Class B, between $\pi$ and $2\pi$ for Class AB and less than $\pi$ for Class C. $I_Q$ is positive and more than $\hat{i}_c$ for Class A, positive and less than $\hat{i}_c$ for Class AB, zero for Class B, negative and less than $\hat{i}_c$. 
Figure 2.5: The efficiency of ideal Class A and Class B amplifiers versus output power level normalized to the maximum available power.

for Class C [26]. It can also be expressed as

\[ I_Q = -\hat{i}_c \cos \frac{y}{2} \]  \hspace{1cm} (2.20)

The Fourier series representation of the collector current in (2.19) is [14]

\[ I_c(t) = I_{ave} + \sum_{n=1}^{\infty} i_n \cos n\omega t \]  \hspace{1cm} (2.21)

where

\[ I_{ave} = \frac{\hat{i}_c}{\pi} \left( \sin \frac{y}{2} - \frac{y}{2} \cos \frac{y}{2} \right) \]  \hspace{1cm} (2.22)

and

\[ i_n = \frac{\hat{i}_c}{2\pi} \left( \frac{2}{n+1} \sin(n+1)\frac{y}{2} + \frac{2}{n-1} \sin(n-1)\frac{y}{2} \right) + \frac{2}{n\pi} \hat{i}_c \cos \frac{y}{2} \sin \frac{ny}{2} \]  \hspace{1cm} (2.23)

As explained in the case of Class B amplification, the parallel tuned circuit allows only the fundamental component of the collector current to reach the load and filters out
Figure 2.6: The maximum efficiency of an ideal power amplifier versus the conduction angle. The rest. Therefore, a sinusoidal waveform is obtained at the load, whose amplitude is given by $n = 1$ case in (2.23).

The maximum collector efficiency of the amplifier can be calculated from equation 2.22 and 2.23 as

$$
\eta_{\text{max}} = \frac{V_{CC} - V_{\text{sat}}}{V_{CC}} \frac{y - \sin y}{4 \left( \sin \frac{y}{2} - \frac{y}{2} \cos \frac{y}{2} \right)}
$$

Figure 2.6 shows the maximum collector efficiency versus the conduction angle if $V_{\text{sat}}$ is negligible and load resistance is optimized for each conduction angle. It illustrates that the PA efficiency can be increased by simply reducing the conduction angle. However, this also increases the amount of clipping and the relationship between the amplitudes of the input and the output sinusoidal waveforms becomes more nonlinear. In addition, the gain of the amplifier is reduced as the conduction angle is decreased.
This overview of the classes of PA operation assumes that the transistor switches rapidly between operating regions with linear gain and without any output current. Although the transition between these two regions is not as sharp for an actual transistor, this analysis is still useful to illustrate the trade-off between efficiency and linearity.

In an actual amplifier, the exponential nature of a bipolar transistor or the square-law nature of a field-effect transistor may require additional means, such as a feedback or feedforward system, for linearizing the transfer function of the amplifier, even if it is operating in Class A mode. The most popular method is employing series feedback through a resistor or inductor tied between the ground and the emitter or source node of the transistor.

### 2.3 Supply Voltage, Breakdown and Efficiency

The analysis in the previous section shows that the maximum efficiency of a PA operating in Class A, B or C mode depends on the ratio

\[ \eta_v = \frac{V_{CC} - V_{sat}}{V_{CC}} \]  

which is a common factor in (2.9), (2.18) and (2.24). In order to increase the efficiency of the digital and some of the analog circuit blocks, there has been a trend towards reducing the supply voltages. However, as the supply voltage is reduced, the efficiency of a power amplifier degrades, as \( V_{sat} \) stays almost constant and \( \eta_v \) becomes smaller.

If the same analysis based on the ideal transistor is followed, it can be shown that the highest voltage drop across the collector and emitter nodes of a bipolar transistor is \( 2V_{CC} - V_{sat} \) when the output voltage swing is maximized. This value needs to be kept smaller than the breakdown voltage of the transistor. When the collector voltage approaches the
breakdown voltage, collector current starts increasing due to avalanche multiplication in the reverse-biased collector-base junction, distorting the relationship between the input signal and output current. Although avalanche breakdown by itself is not destructive [35], once the collector voltage exceeds the breakdown voltage, collector current increases significantly and the junction temperatures may reach harmful levels for the device. Therefore, the breakdown voltage places an upper limit on efficiency as well.

In designing a bipolar transistor, there is a trade-off between the speed (reflected in unity gain frequency $f_T$) and the breakdown voltage $BV_{CEO}$. The $f_TBV_{CEO}$ product is usually assumed to be constant for a given semiconductor material and it is commonly referred to as the Johnson limit [19]. This limit is derived using some crude assumptions and reevaluations of it have been published [37], but it is still useful in illustrating the trade-off between the speed and breakdown, as well as comparing the properties of different semiconductors, such as the treatment in Chapter 3.

The $f_TBV_{CEO}$ constant indicates that when $f_T$ is increased so that the bipolar device can operate at higher frequencies, the breakdown voltage is degraded by the same amount. This puts an upper limit on the maximum voltage swing across a transistor and forces the designers to either use a lower supply voltage or cascode devices, both of which reduce the efficiency.

Designing the PA so that the transistor is prevented from breakdown only during normal operating conditions is not sufficient. The impedance of the antenna differs from its nominal value during operation, in which case the transmitter may try to prevent the output power from falling and force the output voltage swing to increase through a rise
in input power. Therefore, a circuit which monitors the peak voltage at the collector and automatically reduces the gain or the input signal amplitude once that voltage exceeds safety limits may be necessary [43].

2.4 Output Matching

If a small-signal analysis is used, it can be shown that maximum power is delivered when the load is conjugately matched to the source impedance of a generator [40], which requires the load impedance seen by the PA to be the complex conjugate of the output impedance of the PA. However, this condition does not necessarily yield the highest efficiency, because of the limited voltage and current swing available. The maximum collector efficiency is attained when the load seen by the amplifier is

$$R_{L_{opt}} = \frac{(V_{cc} - V_{sat})^2}{2P_{out}}$$ (2.26)

If the load is smaller than $R_{L_{opt}}$, the efficiency is reduced as outlined in section 2.2.1. If $R_L$ is larger than $R_{L_{opt}}$, the output power is limited by the available voltage swing.

It should also be noted that the designer needs to be careful when picking the value of $P_{out}$ in (2.26). The maximum output power specified in the wireless transmission standards usually refer to the highest average power level. The output power level varies along with the envelope over short periods of time while the average power level stays the same. Thus, the PA has to be capable of supplying more power than the specified highest (average) power level. The difference between the average and peak instantaneous power is reflected in the peak-to-average ratio (PAR) of the modulation scheme. A sine wave has a PAR of 3 dB, so $P_{out}$ should be calculated in units of dBm by adding PAR (expressed in...
terms of dB) to the highest average output power level (in terms of dBm) and subtracting 3 dB. As long as the amplifier clips the signal only when the envelope takes on a very large, but rare value, ACPR or transmission accuracy does not degrade by a considerable amount. Therefore, some system simulations should be done before the critical PAR value is determined.

Although $R_{L_{opt}}$ gives the highest collector efficiency, it may also lead to a significant loss of gain compared to the conjugate-matching case and reduce power added efficiency. Besides, some of the passive elements following the PA may also need to see a certain impedance for proper operation. Therefore, the desired load impedance may actually lie somewhere between the conjugate-matching and maximum-collector-efficiency condition. The designer has to find that optimum load by trial and error during simulations and testing. While implementing a PA for this work, the simulations have shown that $R_{L_{opt}}$ given by (2.26) is a good starting point.

The output terminal of the power amplifiers are connected to RF switches, duplexers or antennas. Although these parts can be custom made, the standard 50 Ω ones are usually preferred to reduce the cost. Therefore, the actual load $R_L$ needs to be transformed into the desired impedance $Z_{M}$ through a matching network.

Simplicity is a very attractive feature for a matching network. Although a high order design may provide matching over a wider bandwidth, using more parts increases cost and leads to higher insertion loss. Probably the simplest and the most appealing type is the L-section matching network, which uses two reactive elements. When $Re(Z_M) < R_L$, there are two possible such networks as shown in figure 2.7. The shunt element reduces the
real impedance to the desired level, while the series element is used to change the complex impedance.

Among the two matching networks, the one in figure 2.7(a) is more attractive as it functions as a low pass filter and filters out the higher harmonics. The series inductor can be replaced by a transmission line, whose length can be changed by sliding the shunt capacitor between the transmission line and a slice of exposed ground plane, allowing more precise tuning.

The off-chip element count can be reduced further if on-chip inductors and capacitors are used. However, they have lower Q values than off-chip elements, so they result in higher levels of insertion loss. A lossy inductor can be modeled as an ideal inductor with a series parasitic resistance $R_s = \omega L/Q$. This resistance is in series with the transformed load resistance, so smaller Q values cause more power loss and reduce the efficiency. A similar analysis can be made for the on-chip capacitors, but the inductor Q is the limiting factor for most circuits.
2.5 Input and Interstage Matching

If the power amplifier has one gain stage, the input impedance of the PA should be conjugately matched to the source impedance, which is usually 50 Ω, for maximum power delivery to the power transistor. Accuracy of the input matching network is not as critical as the output matching network and attenuation at the input does not degrade PAE as much as it does at the output. Therefore, the input matching network is a good candidate for realization with on-chip elements. If inductive emitter or source degeneration is used, the real part of the input impedance increases and a matching network with a lower Q can be used.

If more than one gain stage is used in the power amplifier, an interstage matching network becomes necessary. The desired load resistance for achieving maximum efficiency at the driver stage can be calculated according to (2.26) with the appropriate $P_{out}$ determined once the gain of the output stage is known. Then the input impedance of the output stage must be transformed into the desired load of the driver stage through a matching network.

The input impedance of an RF PA is likely to be low, so one of the L-section matching networks in figure 2.8 can be used to transform it into a higher impedance for input or interstage matching. The network in figure 2.8(a) has a low-pass frequency response, but requires an additional AC coupling capacitor. If it is used for interstage matching, the previous stage also requires an additional RF-choke. Thus, the matching network on the right in figure 2.8(b) is more attractive, despite its high-pass frequency response. If the inductor is tied to the supply voltage, an RF-choke is no longer necessary. It can actually be realized by a bondwire in the package and board traces of proper length, which are already
Figure 2.8: L-section input or interstage matching networks (a) Low-pass (b) High-pass

Figure 2.9: Band-pass matching network

necessary for supply connection. The capacitor value is likely to be small enough for an on-chip capacitor, so this network can be designed with almost no external components.

Another popular matching network is shown in figure 2.9. Although it requires more elements, it has a bandpass frequency response and it provides matching over a wider bandwidth.
2.6 Thermal management

As long as PAE of the amplifier is less than 100%, the electrical power dissipated in the circuit

\[ P_d = P_{DC} + P_{in} - P_{out} \] (2.27)

gets converted into heat. Unless the heat is radiated out of the package, the die temperature will quickly exceed the recommended operating conditions, the circuit performance will degrade and the device models will lose their accuracy. If the temperature is allowed to rise further, the transistors or the on-chip wiring can be damaged.

The general heat dissipation paths for IC packages are shown through the arrow marks in figure 2.10. Starting from the heat source (transistor junction), heat may transport through two paths. In the first path, heat transfers from the transistor junction through the molding compound and the top of the package by conduction and then to the air surrounding the package by convection. In the second path, heat flows from the junction of the device through the substrate of the die, through the exposed ground paddle or leadframe into the board ground plane by conduction and finally to the surrounding air by convection. The former path does not conduct much heat unless a special heat sink (a large metal structure with deep grooves for enlarging the surface area and increasing the rate of heat transfer) or cooling systems with air-flow are used. Neither method is suitable for handheld mobile communication applications, so the second path can be assumed to be the primary one for heat dissipation [45].

For illustrating how heat flows out of the package, the analogy of an electrical resistor network can be used. In the electrical resistor model, the electrical resistance
is defined as the potential difference (voltage) across the resistor divided by the current through that resistor. Thermal resistance is defined as the thermal potential difference (temperature) divided by the thermal current (dissipated heat) through the thermal resistor and it is expressed in units of °C/W. Thus, the difference between the device junction temperature and the ambient temperature can calculated as

\[ \Delta T_{ja} = T_j - T_a = P_d \theta_{ja} \tag{2.28} \]

where \( T_j \) is the junction temperature, \( T_a \) is the ambient temperature, \( P_d \) is the power dissipated in the circuit and \( \theta_{ja} \) is the thermal resistance. If the power dissipation is constant and heat is assumed to flow in one dimension, the thermal resistance of each element can be calculated from

\[ \theta = \frac{\ell}{\kappa A} \tag{2.29} \]

where \( \ell \) is the length, \( \kappa \) is the thermal conductivity (W/m °C) and \( A \) is the cross sectional area of the thermal conductor through which the heat is flowing.
A typical thermal resistor network for an IC is shown in figure 2.11. Once the values for each element is calculated, the equivalent thermal resistance $\theta_{ja}$ can be calculated following the steps similar to calculating the equivalent resistance of a typical electrical resistor network.

The designer has to make sure the junction temperature never exceeds the recommended limit $T_{\text{max}}$. Therefore, the power dissipated in the PA under the worst-case condition must be determined first. This is usually when the highest expected supply voltage and input signal swing is applied, the largest expected load is used and the circuit is operated at the highest ambient temperature the circuit is designed for. Under these conditions, $T_a + \Delta T_{ja}$ calculated by (2.28) has to be less than $T_{\text{max}}$.

A more extensive analysis of thermal resistance of the die or the package requires solutions to more complicated two- or three-dimensional differential equations governing
heat flow. Fortunately, most packaging companies provide experimental thermal resistance values. However, special attention must be paid to substrate structures with high thermal resistance, such as the deep trench isolations or buried-oxide layers employed in silicon-on-insulator (SOI) technologies. The thermal conductivity of silicon-dioxide is about one-hundredth that of silicon and these structures impede the heat from easily diffusing into the substrate [21]. This may significantly increase the thermal resistance of the die.

If the other circuit blocks sharing the same die are sensitive to temperature, designers should calculate the temperature transients based on the instantaneous power dissipation at the power transistors as well. In that case, solving the heat flow equations across time and space usually require special CAD tools such as ANSYS [2].

2.7 Packaging

Taking into account the parasitics of the package housing the semiconductor die is critical when designing circuits operating at high frequencies. The prominent among those are the package lead, which can easily have 3 nH of inductance, and bondwires, which usually have about 1 nH of inductance per millimeter. As each nanohenry of inductance gives about 16 Ω of reactance at 2.5 GHz, they may affect the operation of the circuit considerably.

The magnetic coupling between bondwires, reflected in the mutual inductance or coupling coefficient, may create undesired feedback paths, causing stability issues. The reduced isolation between different circuit blocks may reduce the maximum attenuation a variable gain amplifier (VGA) may provide or cause the operating frequency of a voltage
controlled oscillator (VCO) to change. This makes it difficult to integrate the PA with the rest of the transmitter. The electrical coupling, reflected in the capacitance between bondwires or pins, will add to the isolation problems.

As the current swing through a PA at a high power level is quite large, the ground node on the die no longer stays at zero volts, but starts bouncing considerably. This may affect the operation of the other circuit blocks and cause stability problems if these nodes are shared. One way to alleviate this problem is to assign a separate ground and supply pin for each stage of the power amplifier, as well as the main biasing circuitry. An on-chip voltage regulator for lower-power circuit blocks would further reduce isolation problems due to supply bounce [5]. If there is enough isolation, the parasitic inductance at the output pin can be incorporated into the output matching network or resonated out with an external series capacitor. However, even if the ground connection of the power amplifier is isolated from the others, the inductance between the emitter and the ground plane on the board may reduce the gain too much.

In order to alleviate problems associated with packaging parasitics and reduce the package size, “quad, flat-pack, no lead” (QFN) packages have been developed. As there is no lead, the parasitic inductance has been reduced by about two thirds. A package with an exposed ground paddle, which can be connected to the ground plane through several vias on the board, requires shorter bondwires, especially if the wafer is thinned after fabrication. The parasitic inductance between the die and the paddle can be reduced to about 0.5 nH. Using more than one bondwire for connection to the paddle (downbond) from each ground pad and using more pads reduce these values even more, but the mutual inductance between
the downbonds will present a lower limit and makes modeling more difficult.

Another option to reduce the inductance to ground is to create an on-chip ground plane. If a very large piece of metal is created around the PA and connected to the ground plane of the board through a large number of downbonds, the emitter degeneration inductance can be reduced significantly. However, accurate modeling of such structures is difficult and a large die area is required.

A flip-chip or chip-scale package offers significantly less parasitics over its QFN counterparts. A flip-chip package usually consists of a series of solder bumps or balls connecting the pads on the active side of the die (the side with the actual circuit elements) with the board as shown in figure 2.12. The solder bump has about 50 pH of inductance, but the total inductance to ground can be two or three times this value depending on how many vias are used between the top layer of the board and the ground plane.

The lower supply and ground bounce in a flip-chip package provides more isolation between different circuit blocks, increases the gain of the PA and allows a larger voltage swing at the collector, raising the efficiency. Smaller size can also be a big advantage for some applications.
Figure 2.13: Cross section of a flip-chip package with active and inactive solder bumps

The solder bumps provide electrical connections as well as mechanical strength to hold the chip. Therefore, some “dummy” or “inactive” solder balls may be necessary to increase the mechanical strength. However, unlike the bondwire pads, they do not have to be placed at the periphery of the chip. This may open up new possibilities in circuit design. Some of the connections can be made from the inside, reducing the need for long wiring to the edges which makes the circuit more prone to interference. However, solder bumps occupy a larger area than a typical bond pad, which may increase the die area too much if many outside connections are required.

The back side of flip-chip packages does not lie on an exposed ground plane, but faces the air, so almost all of the heat generated by the circuit flows out of the die through the solder bumps that are connected to the ground. As the cross section of these
solder bumps are inevitably smaller than the total area of the die, the thermal resistance increases. However, the difference may not be very large for silicon-on-insulator (SOI) processes, because of the low thermal conductivity of such substrates as mentioned in the previous section. There are also some flip-chip packages employing exposed ground planes for further reducing the inductance of the connection to the ground plane and decreasing the thermal resistance [1].

Prototyping with flip-chip packages is more difficult, as it is impossible to employ laser cutting or focused-ion-beam modifications of the chip under test once it is attached to the board. Every such modification requires the use of another die. The smaller size of the package also requires more accurate alignment on the board, which may require expensive upgrades in some manufacturing facilities.

### 2.8 Stability

It can be shown that oscillation is possible if either the input or output impedance of the amplifier shown in figure 2.14 has a negative real part [40]. Therefore, the designer needs to make sure the real part of both the input and output impedance of the amplifier are positive, which implies that $|\Gamma_{in}| < 1$ and $|\Gamma_{out}| < 1$ where

\[
\Gamma_{in} = \frac{Z_{in} - Z_o}{Z_{in} + Z_o}
\]

(2.30)

is the input return loss,

\[
\Gamma_{out} = \frac{Z_{out} - Z_o}{Z_{out} + Z_o}
\]

(2.31)

is the output return loss and $Z_o$ is the characteristic impedance of the system, usually 50 Ω. $|\Gamma_{in}|$ and $|\Gamma_{out}|$ depends on the load and source impedance values, so it is critical that the
amplifier is designed to be unconditionally stable (i.e. stable for all passive source and load impedances). Otherwise, oscillations may occur when the antenna impedance differs from its nominal value during operation.

If stability is analyzed by means of scattering matrix (S) parameters [40], it can be shown that for an amplifier to be unconditionally stable the necessary and sufficient conditions are

\[
K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|} > 1 \tag{2.32}
\]

and

\[
|\Delta| < 1 \tag{2.33}
\]

where

\[
\Delta = S_{11}S_{22} - S_{12}S_{21} \tag{2.34}
\]

This condition can easily be checked by a CAD tool capable of S-parameter analysis. If these conditions are not met, the methods used for increasing the isolation between circuit blocks outlined in the previous section can be tried. These methods improve stability, as they cause \(S_{12}\) to decrease and \(K\) to increase. The inductive degeneration, which
may be necessary to improve linearity, also improves stability as the gain and $S_{21}$ is reduced. This is also reflected in the higher positive input impedance of the amplifier. The coupling through biasing circuitry may also cause stability problems, so placing low pass filters in the local biasing circuits and careful layout is important.

### 2.9 Electrostatic Discharge

For many people, static electricity is no more than an annoyance in everyday life. However, integrated circuits are much more sensitive to electrostatic discharge (ESD) as the initial voltage across $C_{esd}$ in the simple human body model shown in figure 2.15 can be on the order of thousands of volts \[10\]. Unless special ESD protection circuits are used, ESD can easily degrade or destroy semiconductor devices.

An ESD protection structure turns on during an ESD event and forms a current discharge path for the huge ESD current and clamps the voltage of the external connection node to a sufficiently low level. The structure is designed to stay turned off during normal
A popular structure used for protection against ESD events between the supply and ground nodes is the grounded-base-npn clamp as shown in figure 2.16 [4]. When a negative ESD voltage is applied to the collector, it behaves as a forward-biased diode. In the case of positive ESD stress, the clamp is triggered when the collector voltage reaches the base-collector junction breakdown voltage. The output pins, connecting the collectors of the power transistors to the supply, can take advantage of the size of these transistors, which are usually much larger than any ESD structure. If the resistor in figure 2.16 is replaced by a diode which is reverse biased during normal operation, the power transistor itself could function as an ESD clamp. When a very large input signal swing is applied, the diode also prevents the base-emitter junction from becoming deeply reverse-biased which would degrade the transistor characteristics.

An ESD event may happen between an I/O pin and the supply or ground pins.
In these cases, two reverse-biased diodes, such as the ones shown in figure 2.17, can be used for protection. These diodes will discharge ESD current either by becoming forward biased or through reverse-bias breakdown. If these diodes are realized by grounded-gate MOS transistors instead of pn-junction diodes, more complicated protection circuits may become necessary to prevent them from going into the drain-breakdown [24].

An ESD stress may also be applied between two I/O pins as shown in figure 2.17. In this case, the pin-to-pin ESD current should be discharged through the upper ESD diode to the floating supply node, flow to the floating ground node through the clamp and leave the chip through the lower ESD diode at the second pin [25]. This situation requires the clamp to be well designed so that the ESD current does not get discharged through the internal circuit and cause damage.
As the devices used in ESD protection circuits are used under extremely high values of current and voltage, modeling and simulation of such structures are quite difficult [10]. Fortunately, most fabs provide ESD structure libraries which are already designed and tested. However, there are special circumstances to consider, such as the high voltage swings during normal operation of a PA, which may require the designer to spend some time to choose the right ESD protection. There are also some drawbacks associated with ESD structures, such as additional attenuation due to the parasitics of the ESD diodes. The additional coupling between different nodes and circuit blocks introduced by the ESD protection structures may also cause stability problems, so close attention must be paid during design and simulations.
Chapter 3

Bipolar Transistor Technologies

3.1 Gallium-Arsenide Heterojunction Bipolar Transistors

Despite the dominance of silicon (Si) based integrated circuits, most commercial power amplifiers are still implemented in Gallium-Arsenide (GaAs) heterojunction bipolar transistor (HBT) IC technology. The emitter of an HBT is modified for reducing the base current, which is dominated by hole injection into the base. It is given by

\[ I_b \approx I_{pE} = \frac{qA_E n_E^2 D_{pE}}{N_{dE} x_E} (e^{V_{BE}/V_T} - 1) \] (3.1)

where \( q \) is the charge of an electron, \( A_E \) is the emitter area, \( D_{pE} \) is the diffusion constant of holes in the emitter, \( N_{dE} \) is the donor dopant density in the emitter, \( x_E \) is the distance between the emitter contact and base-emitter junction and \( V_T \) is the thermal voltage which is equal to about 26 mV around room temperature [35]. The bandgap in the emitter of an HBT can be made several hundreds of millivolts higher than in the base by carefully depositing materials such as aluminum (Al) in the emitter. This reduces the amount of
<table>
<thead>
<tr>
<th>Property</th>
<th>Si</th>
<th>GaAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron Mobility (cm(^2)/V-sec)</td>
<td>1417</td>
<td>8800</td>
</tr>
<tr>
<td>Hole Mobility (cm(^2)/V-sec)</td>
<td>471</td>
<td>400</td>
</tr>
<tr>
<td>Saturated Electron Drift Velocity [15]</td>
<td>(10^7)</td>
<td>(2 \times 10^7)</td>
</tr>
<tr>
<td>Electric Field at Breakdown (V/cm)</td>
<td>(3 \times 10^5)</td>
<td>(3.5 \times 10^5)</td>
</tr>
<tr>
<td>Substrate Resistivity [27] (Ω cm)</td>
<td>(\approx 10^4)</td>
<td>(\approx 10^8)</td>
</tr>
<tr>
<td>Thermal conductivity (W/cm °C)</td>
<td>1.412</td>
<td>0.455</td>
</tr>
</tbody>
</table>

Table 3.1: Properties of Si and GaAs at 300 K [35]

The intrinsic carrier concentration \(n_i\) in the emitter and reduces the base current, but it does not affect the collector current. Thus, the emitter can be doped at a lower level and the current gain \(\beta\) of the transistor can still be kept high [41]. This results in a smaller base-emitter junction depletion capacitance \(C_{je}\).

An important figure of merit regarding the speed of bipolar transistors is the unity gain frequency \(f_T\) which is given by

\[
 f_T = \frac{1}{2\pi} \frac{g_m}{g_m \tau_F + C_{je} + C_\mu} \tag{3.2}
\]

where \(\tau_F\) is the forward transit time and \(C_\mu\) is the collector-base junction depletion capacitance [12]. The high electron mobility of GaAs in the base, as listed in table 3.1, reduces the base transit time, which is an important component of \(\tau_F\). Coupled with the smaller \(C_{je}\) this results in the \(f_T\) of a GaAs HBT being considerably larger than that of a Si BJT.

The electric field at breakdown in GaAs is slightly higher than in Si and the saturated electron drift velocity in GaAs is about twice that in Si. Therefore, the Johnson limit, which is covered in section 2.3, is doubled. Being able to withstand larger voltage swings without reducing the speed is a major advantage for a transistor utilized in power amplifiers. This is even more significant for nonlinear PAs, where the maximum voltage at
the collector is much larger than twice the supply voltage. Therefore, Si substrates have been slow to appear in power amplifiers.

Another advantage of a GaAs substrate is its high resistivity, which reduces the loss in the on-chip passives, such as planar inductors, and increases the quality factor Q. However, the thermal resistance of GaAs is also high, as most poor electrical conductors are also poor conductors of heat with few exceptions, such as diamond. This usually requires the GaAs wafers to be thinned to remove the dissipated heat more efficiently, but the brittle nature of GaAs makes these thin wafers even more prone to breakage. Together with high defect densities in GaAs wafers, this forces the diameter of GaAs wafers to be kept much smaller than that of Si wafers [9]. In addition, there is no high-quality grown oxide for GaAs [9]. These fabrication difficulties lower the yield and increase the cost, leading circuit designers to use Si based technologies as long as the better electrical performance of GaAs is not absolutely necessary.

3.2 Silicon-Germanium Bipolar Transistors

Silicon commands most of the global semiconductor market due to its numerous practical advantages. First of all, Si is a very abundant material and can be easily purified. It can be grown in large, virtually defect-free single crystals with a large diameter. It can be doped by n and p-type impurities in a well controlled manner and with an extremely high dynamic range \((10^{14} - 10^{22} \text{ cm}^{-3})\) [9]. \(\text{SiO}_2\) is a high-quality dielectric and can easily be grown on Si. It can be used for isolation, passivation or as an active layer, such as the gate-oxide in metal-oxide-semiconductor field-effect transistors (MOSFETs). In addition, Si has
very good thermal properties and mechanical strength. These reasons lead to cost-effective fabrication and explain the present success of Si.

The drive to integrate more circuit blocks on a single die leads device designers to look for ways to enhance the electrical performance of Si transistors so that they can be better utilized in RF circuits. One of the recent modifications to Si BJTs is depositing Germanium (Ge) into the base. The bandgap of Ge is about 0.66 eV, significantly lower than the bandgap voltage of 1.12 eV in Si [49]. By adding Ge into the base, the bandgap of the base region can be modified and some of the important device characteristics can be improved. SiGe BJTs allow the use of conventional silicon processing methods to be used in other parts of the wafer, making them very cost-effective.

In most SiGe BJTs, Ge changes from low concentration at the emitter-base junction to high concentration at the collector-base junction. This causes the bandgap to gradually become smaller towards the collector and produces an electric field in the neutral base region, which helps drive the electrons and reduces the base transit time significantly [49]. It can be argued that similar bandgap variation can be achieved by more heavily doping one side of the base with acceptors instead of Ge, but the electric field due to dopant variation tends to offset the electric field due to the heavy-doping effect [49].

The smaller bandgap voltage increases the intrinsic carrier concentration and the number of electrons in the base, which in turn increases the collector current. As the properties of the base region do not affect the amount of hole injection into the emitter and therefore the base current, the $\beta$ of the SiGe BJT is raised [49]. If large values of $\beta$ are not required, the base can be doped more heavily with acceptors for reducing the base
resistance, although this would reduce the electron density in the base and the collector current density [9].

The higher intrinsic carrier concentration in the base also makes it harder to deplete the base and increases the early voltage $V_A$ [49]. These improvements can be combined with low-resistance copper wires and low-permittivity dielectrics to reduce parasitics and achieve very high performance RF circuits. Silicon substrates also allow for high levels of integration, so SiGe BJTs have been making inroads in the field of power amplifiers.

### 3.3 Thermal Runaway

An important failure mechanism in power amplifiers is thermal runaway. In a Si BJT, the current gain increases with increasing temperature and since the temperature rises more in places where more current is flowing, even more current tends to flow through hot spots. This positive feedback action leads to severe localized heating and catastrophic failure [29]. Therefore, it is very important for PA designs to utilize a form of negative feedback for preventing thermal runaway. The most commonly used method is placing a resistor between the emitter and ground so that there is a 50 mV potential drop across the resistor at $P_{out_{max}}$. When the collector current increases, the voltage across the resistor increases as well and reduces the base-emitter voltage, which acts as a brake on the collector current.

In GaAs HBTs, thermal instability behaves differently. The current gain decreases with increasing temperature, so the gain collapses in hot spots. Therefore, the thermal instability in GaAs HBTs does not lead to catastrophic failure unless a constant current is
pushed through the collector [29]. Using a sufficiently large emitter or base ballast resistor
is a common method to prevent thermal instability in GaAs HBTs as well.
Part II

Analysis
Chapter 4

Volterra Series

If a nonlinear system does not have memory, the output can be expressed as a Taylor series

\[ y(x) = a_1 x + a_2 x^2 + a_3 x^3 + \cdots \]  

(4.1)

which models weakly nonlinear behavior reasonably well [18]. However, RF power amplifiers include circuit elements, such as capacitors and inductors, whose impedances vary with frequency. This variation introduces memory into the system, which may be modeled by means of a Volterra series [32].

\[
y(t) = \int_{-\infty}^{\infty} h(\tau) x(t - \tau) \, d\tau + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\tau_1, \tau_2) x(t - \tau_1) x(t - \tau_2) \, d\tau_1 \, d\tau_2 + \cdots
\]

\[+ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(\tau_1, \tau_2, \cdots, \tau_n) x(t - \tau_1) x(t - \tau_2) \cdots x(t - \tau_n) \, d\tau_1 \, d\tau_2 \cdots d\tau_n + \cdots\]

(4.2)

where the functions \( h(\tau_1, \tau_2, \cdots, \tau_n) \) are the Volterra kernels of the system [42].

For a causal system described by a Volterra series, \( h(\tau_1, \tau_2, \cdots, \tau_n) \) is zero if any
If \( \tau_j \) is less than zero. If that causal system lacks memory as well, then

\[
h(\tau_1, \tau_2, \cdots, \tau_n) = 0, \text{ for any } \tau_j \neq 0, \ j = 1, 2, \ldots, n
\]  

(4.3)

and (4.2) reduces to a Taylor series.

In narrowband systems, the distortion products due to even-order kernels fall into frequency bands well removed from the desired signal, a statement which will be revisited in the next chapter. Hence, even-order Volterra kernels can be neglected for the analysis of spectral regrowth.

4.1 Calculating the Volterra Kernels of a Circuit

The \( n \)-th-order frequency-domain Volterra kernel is defined as the multidimensional Fourier transform of the time-domain kernel

\[
H_n(f_1, f_2, \cdots, f_n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_n(\tau_1, \tau_2, \cdots, \tau_n) e^{-j2\pi(f_1 f_2 + \cdots + f_n)} d\tau_1 d\tau_2 \cdots d\tau_n
\]  

(4.4)

The frequency-domain kernels are usually more practical for circuit analysis and there are already established methods of calculating them [51]. They can be computed by first replacing each nonlinear component with a linear circuit element and a series of nonlinear generators. The linear response of the circuit can be found by ignoring the nonlinear generators, which turns the calculation into a familiar small-signal analysis. The second-order kernel is calculated by setting the linear independent source to zero and applying the second-order generators instead. Each one of these second-order generators are based on the linear response of the circuit and Taylor series coefficients of the input-output relationship of the nonlinear component. Higher-order responses are ignored at this point, as they will
result in signals of order higher than two when they are fed into the second-order nonlinearity and combined with another signal. The third-order kernel is calculated by applying the third-order generators, which are depend on the linear and second-order responses of the circuit. This procedure is repeated for each high-order kernel.

For example, the nonlinear transconductor represented by the Taylor series expansion

$$I(v_j) = K_{j0} + K_{j1}v_j + K_{j2}v_j^2 + K_{j3}v_j^3 + \cdots \quad (4.5)$$

as shown in figure 4.1(a) is replaced by the nonlinear current sources in figure 4.1(b). $K_{j0}$
of (4.5) is the DC term and $K_{j1}$ is used for calculating the linear response. The expression for the second-order current source is given by

$$I_{j2}(\omega_1, \omega_2) = K_{j2} H_{j1}(\omega_1) H_{j1}(\omega_2)$$ \hspace{1cm} (4.6)

where $H_{j1}$ is the linear transfer function of the voltage $v_j$ controlling the transconductor. The second-order response of the circuit is calculated by multiplying the expressions for the second-order generators by the linear transfer functions $H_T(\omega_1 + \omega_2)$.

The expression for the third-order current source is given by

$$I_{j3}(\omega_1, \omega_2, \omega_3) = K_{j3} H_{j1}(\omega_1) H_{j1}(\omega_2) H_{j1}(\omega_3)$$

$$+ \frac{2}{3} K_{j2} \left[ H_{j1}(\omega_1) H_{j2}(\omega_2, \omega_3) + H_{j1}(\omega_2) H_{j2}(\omega_1, \omega_3) + H_{j1}(\omega_3) H_{j2}(\omega_1, \omega_2) \right]$$ \hspace{1cm} (4.7)

where the latter component is the second-order interaction term due to cascaded nonlinearities or feedback. It is caused by the second-order nonlinearity acting upon a first and second-order signal at the voltage $v_j$ controlling the transconductor. $H_{j2}$ is the second-order transfer function of the voltage $v_j$ calculated by

$$H_{j2}(\omega_1, \omega_2) = \sum_{i=1}^{N} K_{i2} H_{i1}(\omega_1) H_{i1}(\omega_2) H_{T_{i,j}}(\omega_1 + \omega_2)$$ \hspace{1cm} (4.8)

where $N$ is the number of second-order nonlinear generators in the circuit and $H_{T_{i,j}}(\omega)$ is the linear transfer function between the nonlinear generator $i$ and the voltage $v_j$. The third-order kernels are found by multiplying the third-order generators by the linear transfer functions $H_T(\omega_1 + \omega_2 + \omega_3)$.

A voltage-dependent resistor can similarly be replaced by a linear resistor in parallel with nonlinear current sources whose expressions are the same as those of the sources for the transconductor.
A capacitor whose value varies with voltage across it, such as the depletion capacitance at a pn-junction, must be treated slightly differently. If the Taylor series expansion of the capacitance is

\[ C(v_j) = C_0 + C_1 v_j + C_2 v_j^2 + \cdots \]  

(4.9)

the current going through it can be expressed as

\[ i_c(t) = C(v_j) \frac{dv_j}{dt} \]
\[ i_c(t) = C_0 \frac{dv_j}{dt} + C_1 v_j \frac{dv_j}{dt} + C_2 v_j^2 \frac{dv_j}{dt} + \cdots \]
\[ i_c(t) = C_0 \frac{dv_j}{dt} + \frac{C_1}{2} \frac{dv_j^2}{dt} + \frac{C_2}{3} \frac{dv_j^3}{dt} + \cdots \]
\[ i_c(t) = C_0 \frac{dv_j}{dt} + K_{c2} \frac{dv_j^2}{dt} + \frac{K_{c3}}{3} \frac{dv_j^3}{dt} + \cdots \]  

(4.10)

The frequency-domain expressions for the nonlinear current sources in parallel to the linear capacitor include multiplication by angular frequency terms due to these derivatives with respect to time. Therefore

\[ I_{c2}(\omega_1, \omega_2) = (\omega_1 + \omega_2) K_{c2} H_{c1}(\omega_1) H_{c1}(\omega_2) \]  

(4.11)

and

\[ I_{c3}(\omega_1, \omega_2, \omega_3) = (\omega_1 + \omega_2 + \omega_3) K_{c3} H_{j1}(\omega_1) H_{j1}(\omega_2) H_{j1}(\omega_3) \]

\[ + \frac{2}{3} (\omega_1 + \omega_2 + \omega_3) K_{c2} \left[ H_{j1}(\omega_1) H_{j2}(\omega_2, \omega_3) \right. \]

\[ + \left. H_{j1}(\omega_2) H_{j2}(\omega_1, \omega_3) + H_{j1}(\omega_3) H_{j2}(\omega_1, \omega_2) \right] \]  

(4.12)

A nonlinear resistor whose value varies with current flowing through it or a nonlinear voltage or current controlled voltage source can be modeled by high-order voltage sources applied in series with the linear element.
Higher-order kernels can be calculated by following a similar procedure. The expressions for the nonlinear generators become more and more complicated, since they depend on lower-order kernels. A detailed derivation of this approach has been provided by Bussgang et al. by means of the probing method [6]. Similar results can also be obtained by using high-order convolutions and Fourier transforms, which are introduced in the next chapter.

Volterra kernels beyond the third are usually neglected in order to make a practical evaluation possible. Although inclusion of the higher order terms would result in a more accurate representation, the accuracy of these kernels depend on the accuracy of the derivatives of the nonlinear functions in the circuit, which are difficult to determine precisely. In practice, useful results are obtained neglecting terms beyond third order and this approach is followed in this thesis.
5.1 Time and Frequency-Domain Calculations

The input \( x(t) \) in (4.2) representing the signal in a typical communication system does not take on deterministic values, but is composed of a signal modulated by random data and a specified modulation scheme, requiring the generation of large number of bits for a proper simulation. Furthermore, the carrier frequency in wireless communication systems can be 2 to 4 orders of magnitude larger than the bandwidth of the information-carrying signal or the envelope. Therefore, even if just a few samples per each carrier period is used, a computation in the time domain requires too many samples per bit of data for practical circuit simulations.

Frequency-domain calculations require far fewer samples, because the computation can be limited to the frequency bands of interest in narrowband systems. Furthermore, taking the Fourier transform of the convolutions in the Volterra series reduces the order of the computation. For example, the first convolution integral in (4.2), representing the
linear portion of the system, is an $O(N^2)$ computation, while its Fourier transform results in a simple multiplication, which is $O(N)$, as shown in (5.1).

\[
y_1(t) = \int_{-\infty}^{\infty} h_1(\tau)x(t-\tau)\,d\tau \quad (5.1a)
\]

\[
Y_1(f) = H_1(f)X(f) \quad (5.1b)
\]

The three dimensional Fourier transform of the third-order Volterra operator enables a similar frequency-domain computation [3].

\[
y_3(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3)x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)\,d\tau_1\,d\tau_2\,d\tau_3 \quad (5.2a)
\]

\[
Y_3(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_3(\alpha, \beta - \alpha, f - \beta)X(\alpha)X(\beta - \alpha)X(f - \beta)\,d\alpha\,d\beta \quad (5.2b)
\]

Equation (5.2b) can be used for intermodulation distortion ($IM_3$) calculations, which involve only two tones. In this case, $X(f)$ is non-zero at only four frequency values, two on the positive axis ($f_1$ and $f_2$) and two on the negative axis ($-f_2$ and $-f_1$). Therefore, the double integral in (5.2b) reduces to a few multiplications and additions. This has allowed designers in the past to perform hand calculations based on the Volterra series [36].

As explained above, $X(f)$ is no longer a deterministic signal when an analysis of the spectral regrowth using modulated signals is desired. Even though frequency-domain computation reduces the required number of samples and the order of the computation compared to the time-domain approach, implementing (5.2b) directly would still result in an $O(N^3)$ algorithm. The resulting computation times are too long to be practical.
5.2 Decomposition of the Volterra Kernels

In the past, the long computation times have prevented the Volterra series from being utilized in a distortion analysis involving more than a few input tones. In order to dramatically reduce the amount of time it takes for the computations to be completed, a decomposition of the Volterra kernels is proposed in this treatment. The only source of memory in the PA circuit is the frequency dependence of the impedance of inductors or capacitors, whether their values are constant or voltage-dependent. Therefore, a closer examination shows that $H_3(\omega_1, \omega_2, \omega_3)$ for circuits can be broken down into parallel combinations of models resembling the ones shown in figures 5.1(a) and (b) without any approximations. The former figure shows the pure third-order subsystem, while the latter one represents the second-order interaction due to the feedback or cascaded nonlinearities in the circuit. This decomposition is still an exact representation of the original Volterra system as shown in the next section.

This decomposition allows the representation of the third-order nonlinear system with memory by a combination of some linear blocks with memory and nonlinear blocks without memory. The computations involving the linear blocks represented by filters can easily be done in the frequency-domain according to (5.1b). As the nonlinear blocks lack memory, cubing, squaring or multiplication of the signals can be done in the time-domain at a simulation carrier frequency $f_s$ much lower than the actual carrier frequency $f_o$. This allows the compression of the spectrum as shown in figure 5.2.

The compressed spectrum requires about 8 times more samples than the baseband equivalent version, but the order of the computation is reduced to $O(N \log N)$ limited
Figure 5.1: Third-order Volterra subsystems (a) Pure third-order subsystem (b) Second-order interaction subsystems

Figure 5.2: Compressed spectrum
by the inverse-FFT and FFT calculations required before and after the nonlinear blocks. Therefore, all of the computations can be completed in a very short time frame using a tool such as Matlab [30]. It is possible to reduce the processing time further by using a more direct programming language such as C if ease of implementation is not sought.

5.3 Third-Order Volterra Subsystems

5.3.1 Second-Order Interaction Subsystem

Analyzing the second-order interaction subsystem shown in figure 5.1(b) is easier when the upper path is considered first. If the input and output of the squarer are called \( U(f) \) and \( V(f) \) respectively, the following relationships can be derived

\[
U(f) = H_a(f)X(f)
\]

and

\[
V(f) = U(f) \ast U(f) = \int_{-\infty}^{\infty} U(\rho)U(f - \rho) \, d\rho
\]

\[
= \int_{-\infty}^{\infty} H_a(\rho)X(\rho)H_a(f - \rho)X(f - \rho) \, d\rho
\]

as a multiplication in the time-domain is equivalent to a convolution in the frequency-domain. Applying this property once again, the output of the multiplier \( Z(f) \) can be
calculated as

\[ Z(f) = (H_b(f)X(f)) \star (H_c(f)V(F)) \]

\[ = \int_{-\infty}^{\infty} H_b(f-F)X(f-F)H_c(F)V(F) \, dF \]

\[ = \int_{-\infty}^{\infty} H_b(f-F)X(f-F)H_c(F) \left[ \int_{-\infty}^{\infty} H_a(\rho)X(\rho)H_a(F-\rho)X(F-\rho) \, d\rho \right] \, dF \]

\[ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_c(F)H_b(f-F)H_a(\rho)H_a(F-\rho)X(\rho)X(F-\rho)X(f-F) \, d\rho \, dF \quad (5.5) \]

As \( Z(f) \) can also be defined as

\[ Z(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_Z(\alpha, \beta - \alpha, f - \beta)X(\alpha)X(\beta - \alpha)X(f - \beta) \, d\alpha \, d\beta \quad (5.6) \]

some simple change of variables allow \( H_Z(f_1, f_2, f_3) \) to be expressed as

\[ H_Z(f_1, f_2, f_3) = H_c(f_1 + f_2)H_b(f_3)H_a(f_1)H_a(f_2) \quad (5.7) \]

Unfortunately this kernel is not symmetric. In order to get a kernel which does not depend on the exact order of the variables \((f_1, f_2, f_3)\) equation (5.5) needs to be rewritten as

\[ Z(f) = \frac{1}{3} (Z(f) + Z(f) + Z(f)) \]

\[ = \frac{1}{3} \left( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_c(F)H_b(f-F)H_a(\rho)H_a(F-\rho)X(\rho)X(F-\rho)X(f-F) \, d\rho \, dF \right. \]

\[ + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_c(F)H_b(f-F)H_a(\rho)H_a(F-\rho)X(\rho)X(F-\rho)X(f-F) \, d\rho \, dF \]

\[ + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_c(F)H_b(f-F)H_a(\rho)H_a(F-\rho)X(\rho)X(F-\rho)X(f-F) \, d\rho \, dF \right) \quad (5.8) \]

The variables \( \rho \) and \( F \) in the first double integral of (5.8) can easily be replaced by \( \alpha \) and \( \beta \) respectively. The variables \( \rho \) and \( f-\rho \) in the second one can then be replaced by \( \alpha \) and \( \beta - \alpha \), respectively. Similarly, the variables \( f-\rho \) and \( f-F \) in the last double integral can
be replaced by $\beta$ and $\alpha$ respectively, making $d\beta = -d\rho$ and $d\alpha = -dF$. After making the appropriate changes in the limits of the integrals, (5.8) can be represented by

$$Z(f) = \frac{1}{3} \left( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_c(\beta) H_b(f - \beta) H_a(\alpha) H_a(\beta - \alpha) X(\alpha) X(\beta - \alpha) X(f - \beta) d\rho dF \\
+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_c(f - \beta + \alpha) H_b(\beta - \alpha) H_a(\alpha) H_a(f - \beta) X(\alpha) X(f - \beta) X(\beta - \alpha) d\rho dF \\
+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_c(f - \alpha) H_b(\alpha) H_a(f - \beta) H_a(\beta - \alpha) X(f - \beta) X(\beta - \alpha) X(\alpha) d\rho dF \right)$$

(5.9)

and (5.7) becomes

$$H_Z(f_1, f_2, f_3) = \frac{1}{3} (H_c(f_1 + f_2) H_b(f_3) H_a(f_1) H_a(f_2) + H_c(f_1 + f_3) H_b(f_2) H_a(f_1) H_a(f_3) \\
+ H_c(f_2 + f_3) H_b(f_1) H_a(f_2) H_a(f_3))$$

(5.10)

As the output of the second-order interaction subsystem $Y(f)$ is

$$Y(f) = H_d(f) Z(f)$$

(5.11)

$H_3(f_1, f_2, f_3)$ of this subsystem can be calculated as

$$H_3(f_1, f_2, f_3) = \frac{H_d(f_1 + f_2 + f_3)}{3} (H_c(f_1 + f_2) H_b(f_3) H_a(f_1) H_a(f_2) \\
+ H_c(f_1 + f_3) H_b(f_2) H_a(f_1) H_a(f_3) + H_c(f_2 + f_3) H_b(f_1) H_a(f_2) H_a(f_3))$$

(5.12)

5.3.2 Pure Third-Order Subsystem

A similar, but simpler analysis can be performed for the pure third order subsystem shown in figure 5.1(a). If the input and output of the cuber are called $U(f)$ and $V(f)$ respectively, $V(f)$ can be expressed as

$$V(f) = U(f) \ast U(f) \ast U(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(\alpha) U(\beta - \alpha) U(f - \beta) d\alpha d\beta$$

(5.13)
by using the property of the frequency convolution. Thus, \( Y(f) \) is given by

\[
Y(f) = H_b(f)V(f)
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_b(f)U(\alpha)U(\beta - \alpha)U(f - \beta) \, d\alpha \, d\beta
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_b(f)H_a(\alpha)H_a(\beta - \alpha)H_a(f - \beta)X(\alpha)X(\beta - \alpha)X(f - \beta) \, d\alpha \, d\beta
\]

As \( Y_3(f) \) is defined as

\[
Y_3(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_3(\beta, \alpha - \beta, f - \alpha)X(\alpha)X(\beta - \alpha)X(f - \beta) \, d\alpha \, d\beta
\]

(5.15)

some simple change of variables allow \( H_3(f_1, f_2, f_3) \) of this subsystem to be represented by

\[
H_3(f_1, f_2, f_3) = H_b(f_1 + f_2 + f_3)H_a(f_1)H_a(f_2)H_a(f_3)
\]

(5.16)

When the third-order Volterra kernels encountered during an analysis of a circuit is compared to (5.12) and (5.16), it can easily be shown that the original Volterra system can be decomposed into parallel combinations of the subsystems resembling the ones shown in figures 5.1(a) and (b) without any approximations.

5.4 Neglecting Even-Order Terms in Volterra Series

It can be shown that in narrowband systems, the distortion products due to even-order kernels fall into frequency bands well removed from the desired signal. The input \( x(t) \) to the system described by the Volterra series can be represented by an inverse Fourier transform.

\[
x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft} \, df
\]

(5.17)
The \( n \)-th order frequency domain Volterra kernel of that system can be expressed as the \( n \)-dimensional Fourier transform of the time-domain kernel, \( h_n(\tau_1, \cdots, \tau_n) \)

\[
H_n(f_1, \cdots, f_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \cdots, \tau_n) e^{-ij2\pi(f_1\tau_1 + \cdots + f_n\tau_n)} d\tau_1 \cdots d\tau_n \tag{5.18}
\]

If \( x(t - \tau_i) \) is replaced by (5.17) and \( \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \cdots, \tau_n) e^{-ij2\pi\sum_{i=1}^{n} \rho_i \tau_i} d\tau_1 \cdots d\tau_n \) is replaced by (5.18), the \( n \)-th term of the Volterra series can be rewritten as

\[
y_n(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \cdots, \tau_n) \left[ \prod_{i=1}^{n} x(t - \tau_i) \right] d\tau_1 \cdots d\tau_n
= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \cdots, \tau_n) \left[ \prod_{i=1}^{n} \int_{-\infty}^{\infty} X(\rho_i) e^{ij2\pi \rho_i (t - \tau_i)} d\rho_i \right] d\tau_1 \cdots d\tau_n
= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \cdots, \tau_n) e^{-ij2\pi\sum_{i=1}^{n} \rho_i \tau_i} d\tau_1 \cdots d\tau_n \right]
\times \left[ \prod_{i=1}^{n} X(\rho_i) \right] e^{ij2\pi\sum_{i=1}^{n} \rho_i t} d\rho_1 \cdots d\rho_n
= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_n(\rho_1, \cdots, \rho_n) \left[ \prod_{i=1}^{n} X(\rho_i) \right] e^{ij2\pi\sum_{i=1}^{n} \rho_i t} d\rho_1 \cdots d\rho_n \tag{5.19}
\]

Variable \( \rho_n \) can be replaced by defining \( f = \rho_1 + \cdots + \rho_{n-1} + \rho_n \). Please note that

\[
df = d\rho_n \text{ and } \rho_n = f - \sum_{i=1}^{n-1} \rho_i. \text{ Thus, equation (5.19) becomes}
\]

\[
y_n(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_n(\rho_1, \cdots, \rho_{n-1}, f - \sum_{i=1}^{n-1} \rho_i) X(\rho_1) \cdots X(\rho_{n-1}) X(f - \sum_{i=1}^{n-1} \rho_i)
\times e^{ij2\pi ft} d\rho_1 \cdots d\rho_{n-1} df \tag{5.20}
\]

If (5.20) is compared to the inverse Fourier transform equation

\[
y_n(t) = \int_{-\infty}^{\infty} Y_n(f) e^{ij2\pi ft} df \tag{5.21}
\]

The Fourier transform of the \( n \)-th term of the Volterra series can be calculated as

\[
Y_n(f) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_n(f_1, \cdots, f_{n-1}, f - \sum_{i=1}^{n-1} f_i) X(f_1) \cdots X(f_{n-1}) X(f - \sum_{i=1}^{n-1} f_i) df_1 \cdots df_{n-1} \tag{5.22}
\]
Let’s assume that a narrowband input signal is applied to this nonlinear system, such that the carrier frequency $f_o$ is much greater than the bandwidth of the signal $\Delta f$. Hence, $X(f_i)$ is nonzero only for $f_i \in \left[\left(-f_o - \frac{\Delta f}{2}, -f_o + \frac{\Delta f}{2}\right) \cup \left(f_o - \frac{\Delta f}{2}, f_o + \frac{\Delta f}{2}\right)\right]$ where $\Delta f \ll f_o$. Thus, the last term of the integral in (5.22) is non-zero only if

$$f - \sum_{i=1}^{n-1} f_i \in \left[\left(-f_o - \frac{\Delta f}{2}, -f_o + \frac{\Delta f}{2}\right) \cup \left(f_o - \frac{\Delta f}{2}, f_o + \frac{\Delta f}{2}\right)\right]$$

(5.23)

Therefore, $Y_n(f)$ is non-zero around $f_o$ only if $\sum_{i=1}^{n-1} f_i$ is about zero. This requires $\frac{n-1}{2}$ of $f_i$ terms to be in $\left(-f_o - \frac{\Delta f}{2}, -f_o + \frac{\Delta f}{2}\right)$ and the remaining $\frac{n-1}{2}$ terms to be in $\left(f_o - \frac{\Delta f}{2}, f_o + \frac{\Delta f}{2}\right)$. In order for $\frac{n-1}{2}$ to be an integer value, $n$ has to be odd.

As the distortion products due to even-order kernels fall into frequency bands well removed from the desired signal, they can easily be filtered out. Hence, even-order Volterra kernels can be neglected for the analysis of spectral regrowth. However, the even-order kernels of the circuit still need to be calculated as they contribute to the odd-order kernels as outlined in the previous chapter.
Chapter 6

Device Modeling

The new computational approach described in the previous chapter is applied to the design of PAs using bipolar transistors as the active element. The nonlinear bipolar transistor model used in the analysis is shown in figure 6.1. This model is applicable to Si and SiGe BJTs, as well as GaAs HBTs.

The output impedance seen by a PA is generally small compared to $r_o = 1/V_A$, where $V_A$ is the Early voltage, so $r_o$ is neglected in the model. The parasitic capacitance between collector and substrate can be assumed small in a silicon-on-insulator (SOI) process or can be assumed constant and combined with the package parasitics and output matching network in more conventional processes.

The parasitic collector resistance $r_c$ and emitter resistance $r_e$ are assumed constant. Although the value of $r_b$ is known to change at high base-current levels, its nonlinearity is usually negligible compared to the other nonlinear elements in the transistor [51], so $r_b$ is assumed constant as well.
The transconductance $g_m$ is nonlinear due to the exponential relationship between collector current and the base-emitter voltage given by

$$I_c = I_s e^{\frac{V_{BE}}{n_F V_T}}$$  \hspace{1cm} (6.1)

where $I_s$ is the transistor saturation current, $n_F$ is the forward emission coefficient, which is generally equal to about 1 and $V_T$ is the thermal voltage which is equal to about 26 mV around room temperature [35] and is calculated by

$$V_T = \frac{kT}{q}$$  \hspace{1cm} (6.2)

where $T$ is the temperature in Kelvins, $k$ is the Boltzmann’s constant, which is equal to $1.38 \times 10^{-23}$ J/K and $q$ is the charge of an electron, which is equal to $1.602 \times 10^{-19}$ C. The
Taylor series expansion of the collector current when $V_{bc} = V_{BE} + v_{be}$ is

$$I_c = I_c(V_{BE}) + \sum_{i=1}^{\infty} \frac{1}{n!} \frac{d^n I_c(V_{BE})}{dv_{be}^n} v_{be}^n$$

$$= I_c(V_{BE}) + \frac{dI_c(V_{BE})}{dv_{be}} v_{be} + \frac{1}{2} \frac{d^2 I_c(V_{BE})}{dv_{be}^2} v_{be}^2 + \frac{1}{6} \frac{d^3 I_c(V_{BE})}{dv_{be}^3} v_{be}^3 + \cdots$$

$$= I_C + g_m v_{be} + K_{gm2} v_{be}^2 + K_{gm3} v_{be}^3 + \cdots$$

where

$$g_m = \frac{I_s}{n_F V_T} = \frac{I_C}{n_F V_T}$$

$$K_{gm2} = \frac{I_C}{2n_F^2 V_T^2}$$

$$K_{gm3} = \frac{I_C}{6n_F^3 V_T^3}$$

The base current is

$$I_b = I_c/\beta$$

as long as the current levels are not very high or very small. Therefore, the base-emitter junction can be modeled as a nonlinear conductor, so that the base current is

$$I_b = I_B + \frac{g_m}{\beta} v_{be} + \frac{K_{gm2}}{\beta} v_{be}^2 + \frac{K_{gm3}}{\beta} v_{be}^3 + \cdots$$

The second term is the first-order relationship between $I_b$ and $v_{be}$ which is represented by the base-emitter resistance $r_n = \beta/g_m$ in the small-signal model of a bipolar transistor. The variation of $\beta$ due to changes in the base-current is assumed negligible compared to the exponential behavior of $I_c$, so $\beta$ is assumed constant.

Linear PA design requires that the power transistors are prevented from going into the saturation region, where the transistor operation is no longer weakly nonlinear. Therefore, it is assumed that the collector-base junction is never forward biased for the
analysis. A similar argument can be made for avoiding the device breakdown. Thus, the collector-base capacitor $C_\mu$ in figure 6.1 is composed of some constant parasitic capacitance and the collector-base junction depletion capacitance. The latter capacitance is the cause of nonlinearity in $C_\mu$ and the analysis shows that its contribution to distortion is considerable due to the large signal swings across the collector-base junction during the PA operation.

The relationship between the reverse-bias voltage and pn-junction depletion capacitance is given by

$$C_j = \frac{C_{j0}}{\left(1 + \frac{V_R}{\phi_j}\right)^{m_j}} \quad (6.9)$$

where $C_{j0}$ is the junction capacitance at zero bias, $V_r$ is the reverse-bias voltage across the pn-junction, $\phi_j$ is the built-in potential and $m_j$ is the junction grading coefficient, which is equal to 0.5 for an abrupt junction and about 0.33 for a linearly graded junction [51]. When the reverse-bias voltage is defined as $V_r = V_R + v_r$ the coefficients of the nonlinear current sources are calculated from (4.10) to be

$$C_J = \frac{C_{j0}}{\left(1 + \frac{V_R}{\phi_j}\right)^{m_j}} \quad (6.10)$$

$$K_{C2} = \frac{1}{2} \frac{C_J}{1 + \frac{V_R}{\phi_j}} \frac{m_j}{\phi_j} \quad (6.11)$$

$$K_{C3} = \frac{1}{6} \frac{C_J}{\left(1 + \frac{V_R}{\phi_j}\right)^2} \frac{m_j(m_j + 1)}{\phi_j^2} \quad (6.12)$$

where $V_R$ is the DC reverse-bias voltage across the capacitor.

The base-emitter capacitance $C_\pi$ in figure 6.1 is given by

$$C_\pi = C_b + C_{je} = g_m \tau_F + C_{je} \quad (6.13)$$

where $C_{je}$ is the base-emitter depletion capacitance, $C_b$ is the diffusion capacitance and $\tau_F$ is the forward transit time. Although $C_{je}$ is usually neglected or assumed constant, its
value can become comparable to $C_b$ and its nonlinearity can be significant when $V_{be}$ swing is large. Therefore, it is necessary to model the nonlinearity of $C_{je}$ for an accurate analysis.

The diffusion capacitance $C_b$ also varies with $V_{be}$ because of the variation in the transconductance $g_m$ and forward transit time $\tau_F$. The latter is usually assumed constant, but its value starts to increase and cause additional distortion at high collector current densities. In order to increase the accuracy of the analysis, the series based model outlined below has been developed.

### 6.1 Modeling the Variations in Forward Transit Time

The forward transit time $\tau_F$ has four components

$$\tau_F = \tau_E + \tau_{BE} + \tau_B + \tau_{BC}$$

(6.14)

where $\tau_E$ is the emitter transit time, $\tau_{BE}$ is the base-emitter depletion region transit time, $\tau_B$ is the base transit time and $\tau_{BC}$ is the base-collector depletion region transit time [49].

The first component of $\tau_F$ affected by the increasing current density is usually $\tau_{BC}$. When current is flowing through an npn transistor, the injected electrons are added to the negatively charged depletion region on the base side and subtracted from the positively charged depletion region on the collector side [34]. For a constant base-collector voltage, the area under the two electric field curves in figure 6.2 stay the same. Therefore, the depletion region of the collector becomes longer and the number of ionized donors rises. The depletion region of the base becomes shorter as well, but the effect is much less pronounced as the doping density in the base is several orders of magnitude higher than in the collector. Therefore, only the variation in $\tau_{BC}$ is modeled for this work.
Figure 6.2: (a) Electric field and (b) space-charge density in the collector-base junction including the effects of mobile electrons due to collector current.
If the current density is increased further until the mobile electron density exceeds the density of the positively charged donor ions, the electric field collapses and the neutral base width is effectively increased up to the heavily doped buried layer, which is referred to as Kirk effect, base push-out or base widening [49, 35, 41, 20]. However, at this point, the rise in $\tau_F$ is quite rapid and this operating region is not suitable for a linear PA. As the effective boundary of the collector region moves to the heavily doped buried layer, the device breakdown voltage drops significantly as well [29]. Thus, it is assumed that the current density is kept below the onset of the Kirk effect.

The base-collector depletion region transit time can be found by dividing the charge of these additional ions by the collector current.

$$\tau_{BC} = \frac{q N_c \Delta x}{J_c}$$  \hspace{1cm} (6.15)

where $N_c$ is the collector dopant density per volume and $J_c$ is the collector current density per cross sectional area. For the analysis, it is assumed that the electric field in the space-charge region is large enough to cause the electrons to move at the scattering-limited velocity $\vartheta_{\text{sat}}$ over the entire region. Thus, the additional mobile electron charge density is equal to $J_c/\vartheta_{\text{sat}}$. The area under the electric field curve in figure 6.2(a) is equal to the collector-base voltage $V_{cb}$ and the built-in potential $\phi_j$. As the doping density in the base is several orders of magnitude higher than in the collector, all of the voltage is assumed to be dropped on the collector side. Therefore,

$$V_{cb} + \phi_j = \frac{q N_c x_{d0}^2}{2 \epsilon_s} = \frac{(q N_c - J_c/\vartheta_{\text{sat}})}{2 \epsilon_s} (x_{d0} + \Delta x)^2$$  \hspace{1cm} (6.16)

where $x_{d0}$ is the width of the depletion region without current and $\epsilon_s$ is the permittivity of
the semiconductor. Thus,

$$\Delta x = x_{d0} \left( \sqrt{\frac{1}{1 - \frac{J_c}{qN_c\vartheta_{sat}}} - 1} \right) \quad (6.17)$$

and

$$\tau_{BC} = \frac{qx_{d0}N_c}{J_c} \left( \sqrt{\frac{1}{1 - \frac{J_c}{qN_c\vartheta_{sat}}} - 1} \right) \quad (6.18)$$

At low current densities $J_c/\vartheta_{sat}$ is much smaller than $qN_c$, so (6.18) gives the usual $\tau_{BC}$ equation

$$\tau_{BC} \approx \frac{x_{d0}}{2\vartheta_{sat}} \quad (6.19)$$

by using $(1 - x)^{-0.5} \approx 1 + 0.5x$ approximation [34]. In order to model the variation in the forward transit time at high current densities, $\tau_F$ can be expanded as a Taylor series

$$\tau_F(V_{BE} + v_{be}) = \tau_F(V_{BE}) + \tau'_{BC}(V_{BE})v_{be} + \frac{\tau''_{BC}(V_{BE})}{2}v_{be}^2 + \frac{\tau''''_{BC}(V_{BE})}{6}v_{be}^3 + \cdots$$

$$= \tau_F(V_{BE}) + K_{\tau_F1}v_{be} + K_{\tau_F2}v_{be}^2 + K_{\tau_F3}v_{be}^3 + \cdots \quad (6.20)$$

where

$$\tau_{BC}(V_{BE}) = \left[ \left( 1 - \frac{J_c}{qN_c\vartheta_{sat}} \right)^{-\frac{1}{2}} \right] \frac{x_{d0}qN_c}{J_c} \quad (6.21)$$

$$K_{\tau_F1} = \frac{x_{d0}}{2\vartheta_{sat}V_T} \left( 1 - \frac{J_c}{qN_c\vartheta_{sat}} \right)^{-\frac{3}{2}} - \frac{\tau_{BC}(V_{BE})}{V_T} \quad (6.22)$$

$$K_{\tau_F2} = \frac{3x_{d0}}{8\vartheta_{sat}V_T^2 qN_c\vartheta_{sat}V_T} \left( 1 - \frac{J_c}{qN_c\vartheta_{sat}} \right)^{-\frac{5}{2}} - \frac{K_{\tau_F1}}{2V_T} \quad (6.23)$$

$$K_{\tau_F3} = \frac{5x_{d0}}{16\vartheta_{sat}V_T^2} \left( \frac{J_c}{qN_c\vartheta_{sat}V_T^2} \right)^2 \left( 1 - \frac{J_c}{qN_c\vartheta_{sat}} \right)^{-\frac{7}{2}} + \frac{x_{d0}}{8\vartheta_{sat}V_T} \left( \frac{J_c}{qN_c\vartheta_{sat}V_T^2} \right)^2 \left( 1 - \frac{J_c}{qN_c\vartheta_{sat}} \right)^{-\frac{5}{2}} - \frac{K_{\tau_F2}}{3V_T} \quad (6.24)$$
Thus, the current in $C_b$ is given by

$$i_b = \frac{dQ_b}{dt} = \frac{d}{dt}(\tau_F I_c) = \tau_F \frac{dI_c}{dt} + I_c \frac{d\tau_F}{dt}$$

$$= \tau_F (g_m \frac{dv_{be}}{dt} + K_{g_{m2}} \frac{dv_{be}^2}{dt} + K_{g_{m3}} \frac{dv_{be}^3}{dt} + \cdots)$$

$$+ I_c (K_{\tau_F_1} \frac{dv_{be}}{dt} + K_{\tau_F_2} \frac{dv_{be}^2}{dt} + K_{\tau_F_3} \frac{dv_{be}^3}{dt} + \cdots)$$

(6.25)

(6.26)

The first part of (6.26) is the usual equation which governs the Volterra series representing the nonlinearity of the diffusion capacitance when $\tau_F$ is assumed constant. The second part is the series proposed for modeling the variations of $\tau_{BC}$. Equation (6.26) can be further simplified as

$$i_b = (g_m \tau_F + I_c K_{\tau_F_1}) \frac{dv_{be}}{dt} + (K_{g_{m2}} \tau_F + I_c K_{\tau_F_2}) \frac{dv_{be}^2}{dt} + (K_{g_{m3}} \tau_F + I_c K_{\tau_F_3}) \frac{dv_{be}^3}{dt} + \cdots$$

(6.27)

so that $\tau_F$ variation model does not increase the computational complexity, once the extra $K_{\tau_F}$ coefficients are calculated. However, it should be noted that the model requires a somewhat more intimate knowledge of the transistor than just SPICE parameters, as the collector area and the collector dopant density cannot be directly captured from a SPICE model.
Part III

Application
Chapter 7

Implementation

7.1 Specifications

In order to compare the results of the preceding analysis with measurements, a number of single stage PAs have been designed for the IEEE802.11b wireless LAN standard operating at 2.4 GHz in the ISM band. The standard specifies the maximum output power level to be 20 dBm at the antenna, but the PAs have been designed to supply 24 dBm in order to accommodate the losses through the passive elements before the signal reaches the antenna.

Quadrature phase-shift keying (QPSK) is the modulation scheme used in the high data-rate version of the IEEE802.11b standard [17]. The data pulses are shaped by a low-pass filter in order to reduce the spectral sidelobes. The exact pulse shape is not explicitly defined in the standard, but the transmit spectrum mask specifications require the spectral products in the adjacent sidelobe to be 30 dB below the main lobe, as shown in figure 7.1. The “unfiltered sin x/x” curve superimposed on the diagram shows the adjacent spectral
sidelobes, which would be only about 13 dB below the mainlobe if rectangular data pulses were used [8]. However, the spectral efficiency comes at the price of envelope variation, which requires a linear power amplifier in order to prevent distortion.

### 7.2 Details of the Circuit

The test PAs have been designed using low-cost SiGe bipolar transistor IC technology. A flip-chip package, called ultra chip-scale packaging (UCSP), has been utilized for reducing parasitics and package size. The use of UCSP enables the die to be attached directly to the printed-circuit board face-down, with the chip’s pads connecting to the PC board’s pads through individual solder balls without any underfill material or a laminate substrate. The solder balls are added onto the wafer before dicing.

The transistors are laid out in 8 columns as shown in the die photo in figure 7.2.
The base and collector connection of each column are wired such that there is 20 or 30 $\mu$m distance between the column itself and the point where it is connected to the rest of the circuit. This distance allows the base and collector connection in each column to be cut by a laser and the number of parallel transistors in the amplifier to be adjusted without damaging the rest of the circuit.

The simplified schematic of the common emitter PA is shown in figure 7.3. In order to reduce the number of nodes, the Norton equivalent of the signal source, input matching and local-bias circuit, as shown in the dashed box, has been used in the analysis.

The off-chip choke between the collector and VCC passes the bias current, but has high impedance at RF frequencies so that almost all of the signal flows into the antenna through the output matching network. $R_E$ includes the parasitic emitter resistance, as well
as the emitter degeneration resistance. The total value of $R_E$ is adjusted so that there is about 50 mV DC voltage drop across it, in order to prevent thermal runaway. $L_E$ is composed of the on-chip wiring inductance, package parasitics and the inductance of the vias on the board. The inductance of the structures outside of the die are calculated by means of a 3-D EM simulator and the inductance of the on-chip wiring is estimated by using the Greenhouse formula for a wire with rectangular cross-section

$$L = 2 \times 10^{-4} \ell \left( 0.50049 + \frac{w + t}{3\ell} + \ln \frac{2\ell}{w + t} \right)$$

(7.1)

where $L$ is the inductance in nH, $\ell$ is the length of the wire in $\mu$m, $w$ is the width and $t$ is the thickness of the wire [13]. This formula assumes that the magnetic permeability is 1 and the operation frequency is low enough such that skin-depth has negligible effects.
Both $R_E$ and $L_E$ improve linearity through series feedback, but an inductor is preferred, as it does not limit voltage headroom as a resistor does. Inductive degeneration also increases the real part of the input impedance so that the input (or interstage) matching network can be designed with a lower quality factor ($Q$). $L_E$ has been adjusted by changing the number of ground vias on the board so that there is enough improvement in spectral regrowth, but gain is not reduced too much in order to keep PAE high.

The series feedback used to improve the linearity of the PA usually causes the third-order coefficient of the system to have a sign opposite to that of the first-order coefficient. Thus, gain compression occurs at high power levels. One way to alleviate the gain compression problem is to allow the PA bias current to increase at high power levels through some modification of the biasing circuitry. The conventional biasing circuitry for a common emitter amplifier is usually based on a current mirror with a current helper as shown in figure 7.4 [33]. The ratio of the resistors tied to the base and emitter of $Q_1$ and $Q_2$ must be adjusted to make sure the voltage drop across them and, in turn, the voltage drop across the base-emitter junctions are the same. The base resistor and capacitor $C_F$ act as a filter to attenuate the input signal before reaching the bias circuit.

A large input signal swing increases the average value of the collector current over the quiescent value, due to the exponential nature of the bipolar transistor. However, this increases the base current and causes a bigger voltage drop across the base resistor, reducing the base-emitter voltage of the power transistor $Q_1$ and limiting the increase in the average collector current. Therefore, the biasing circuit shown in figure 7.5 is preferred, as the lack of a resistor at the base of the power transistor allows a bigger increase in average collector...
Figure 7.4: Conventional local-bias circuit

Figure 7.5: Local-bias circuit without base resistor
current at high power levels. A further advantage is that the lower impedance seen by
the base of $Q_1$ increases the device breakdown voltage. The values of the emitter and base
resistors of $Q_2$ need to be adjusted so that the voltage drop across the base-emitter junctions
of both transistors are equal to each other. The current mirror ratio depends on the actual
value of $\beta$, but DC simulations have shown that the change in the collector current of $Q_1$
due to the process variations is still small. $C_F$ also needs to be moved towards the base of
$Q_2$, so that the signal is not attenuated before reaching the power transistor.

Analysis shows that the amount of output voltage swing at a given output power
level is quite sensitive to the imaginary part of the impedance at the collector. A small
positive imaginary part at that node can easily increase the voltage swing across base-
collector junction and push the transistor into saturation, even if the output return loss
$S_{22}$ stays acceptable. Therefore a transmission line, instead of a discrete inductor, is used
in the output matching network, as described in section 2.4 for more precise tuning. The
typical length of the transmission line required for a PA in a 50 $\Omega$ system is usually short
enough to be realized on compact circuit boards.

7.3 Kernel Calculations

As mentioned in the previous section, the Norton equivalent of the signal source,
input matching and local-bias circuit has been used in the analysis in order to reduce the
number of nodes. The resulting schematic is shown in figure 7.6. Although the local-bias
circuit utilizes active devices, a frequency-dependent impedance based on AC simulations
has been used in order to simplify the analysis.
Figure 7.6: Norton equivalent of the schematic of the power amplifier in figure 7.3

The method outlined in chapter 4.1 has been followed for calculating the Volterra kernels of the amplifier. The schematic for calculating the first-order response is shown in figure 7.7. The load seen by the transistor $Z_L(\omega)$ consists of the actual load of 50 $\Omega$, the AC coupling capacitor, output matching network, RF-choke, the package parasitics due to the solder bump and pad, on-chip wiring parasitics and the parasitic collector resistor. $Z_A(\omega)$ is the parallel combination of the local-bias circuit and the source impedance seen by the transistor, which consists of the package and on-chip wiring parasitics, input matching network and the actual source resistance of 50 $\Omega$. The transfer function from the input current source $I_s$ to the output voltage across the load $Z_L$ gives the first-order response of the circuit. The nonlinear elements are located between the intrinsic base and emitter, the intrinsic collector and base, as well as the intrinsic collector and emitter nodes, so the
Figure 7.7: The schematic for calculating the first-order response of the power amplifier presented in figure 7.6

transfer functions from the input current source to the voltages at these nodes have also been calculated.

The schematic for calculating the second-order response is shown in figure 7.8. The linear independent source is set to zero, so the input current source becomes an open circuit. The nonlinear current source $I_{gm2}$ models the second-order nonlinearity of the exponential transconductance of the bipolar transistor according to (4.6) and (6.5). $I_{\mu2}$ models the second-order nonlinearity of the collector-base depletion capacitance according to (4.11) and (6.11). $I_{\pi2}$ models the nonlinearity of the base-emitter conductance, as well as the depletion and diffusion capacitance. The linear transfer functions from each nonlinear source to the voltages at the collector, base and emitter nodes are then multiplied by the expression for that second-order source in order to find the second-order response of the
Figure 7.8: The schematic for calculating the second and third-order responses of the power amplifier in figure 7.6

circuit at those nodes according to (4.8).

Figure 7.8 can be used to calculate the third-order kernel of the circuit by simply replacing the subscripts “2” with “3” and using the third-order expressions for the nonlinear sources.

The measurements of spectral regrowth made at certain output power levels need to be compared to the predictions from the analysis. Therefore, the power lost between the intrinsic collector and the equipment, which is modeled by the 50 Ω load, should be calculated as well. As there are no nonlinear generators in the path from the intrinsic collector to the equipment, this path does not make any additional contribution to spectral regrowth. Therefore, the power loss calculation is based on a series of simple linear current dividers. A similar calculation is made for finding the amount of power lost between the
actual source and the input of the transistor in order to predict the overall power gain more accurately.

The method of calculating the first and higher-order responses of the circuit is not very demanding, so these computations are done by hand. The resulting equations are introduced in the appendix, where the programs to implement these computations are also presented.

7.4 Simulator

The simulation method using the Volterra-series-based power-amplifier model and the decomposition of the Volterra kernels has been implemented in Matlab. The block diagram of the operating principles of the simulator is shown in figure 7.9. Baseband I and Q signals are generated from 64-times oversampled 1024-bit long stream of random data, which forms a rectangular waveform varying between 1 and -1. These rectangular pulses are then shaped by the baseband filter. As the details of that filter are not specified explicitly in the IEEE802.11b standard, the filtering is performed by the frequency-domain transfer function of the seventh-order Bessel filter of the arbitrary waveform generator used during the measurements.

Some measurements are taken by using waveforms modulated according to the IS-95 cell phone standard to further validate the computational approach developed in this thesis. Therefore, a program generating the IS-95 waveforms are also written. The baseband filter for IS-95 is a 48-tap finite impulse response (FIR) filter, whose coefficients are specified by the standard [50]. FIR filter requires 4-times oversampled data streams. Therefore, the
The filtered data stream is upsampled by 16 times (by using the “interp” command) after the FIR filter coefficients are convolved with the bit stream (by using the “conv” command). The general principles of upsampling can be found in [38]. IS-95 uses offset-QPSK modulation scheme, where the data stream at the baseband Q-channel is delayed by one half of a symbol duration. Therefore, the frequency-domain signal at the Q-channel is multiplied by $e^{-j2\pi fT/2}$ before the upsampling, where $T = 1/(1.2288 \times 10^6)$ is the symbol duration.

The baseband signals are then upconverted to the simulation carrier frequency $\omega_s$ for generating the input waveform. Afterwards, the input signal amplitude is adjusted according to the desired input power level and fed into the PA model, which includes the SPICE coefficients of the npn transistors and the design parameters, to generate the output waveform. The frequency spectrum of the linear filters inside the Volterra subsystems are divided into seven roughly equal length segments to represent the frequency bands around $\omega = 0, \omega = \pm \omega_c, \omega = \pm 2\omega_c$ and $\omega = \pm 3\omega_c$, where $\omega_c$ is the actual carrier frequency.
Figure 7.10: Input and output power spectral density calculated by simulations

The power spectral densities of the input and output waveforms shown in figure 7.10 are generated by this method. The amplitude of the adjacent sidelobes relative to the mainlobe is much larger in the output than in the input due to spectral regrowth.
Chapter 8

Results

8.1 Analysis versus Measurements

In order to validate the analysis methodology presented in the thesis, predicted values of spectral regrowth are compared to measurement results at different output power levels. Measurements have been taken using a range of values for the passive elements in the input and output matching networks, different bias currents and supply voltages, as well as different number of transistors, which can be changed by means of a laser cutter.

The measured ratio of the adjacent sidelobe and mainlobe versus output power levels are compared with the numerical predictions in figure 8.1. An IEEE802.11b modulated waveform at a carrier frequency of 2.45 GHz is applied to a power amplifier which consists of 78 output transistors in parallel. The PA is operated at a supply voltage of 3.3 V and a quiescent current of 196 mA. At high power levels the current density becomes large and an assumption of constant $\tau_F$ model results in underestimation of the spectral regrowth. The analysis including the $\tau_F$ model predicts the measured sidelobe growth to
Measurements with constant $\tau_F$

Figure 8.1: Ratio of the first sidelobe and the mainlobe versus output power level for $m = 78$, $V_{CC} = 3.3 \text{ V}$, $I_{Q} = 196 \text{ mA}$, $f_c = 2.45 \text{ GHz}$

within 1.6 dB or better accuracy, while only requiring minimal computation time.

The results of the analysis show that the increase in the base-collector depletion region transit time $\tau_{BC}$ at high collector current densities can easily become the dominant source of nonlinearity in a power amplifier. This problem can be alleviated by increasing the total collector area of the PA at the expense of increased parasitics and lower gain. The analysis predicts that using 104 parallel output transistors would reduce the contribution of $\tau_{BC}$ variation to negligible levels and improve spectral regrowth as shown in figure 8.2. A similar improvement is observed during the measurements as well.

A number of similar measurements have been taken from another PA with 78 parallel output transistors. The results of the measurements and analysis for this case can be seen in figure 8.3. The operating frequency in this case is 2 GHz, the quiescent current
Figure 8.2: Ratio of the first sidelobe and the mainlobe versus output power level for $m = 104$, $V_{CC} = 3.3$ V, $I_{cQ} = 196$ mA, $f_c = 2.4$ GHz

is 176 mA and the supply voltage is 3.5 V. The predicted spectral regrowth again differs by less than 1.5 dB compared to the measurements. Figure 8.4 shows the results for the same PA supplying 24 dBm output at different average current levels. The measurements and predictions agree to within 0.6 dB. The trend difference at very high currents is due to the onset of Kirk effect and consequent modeling inaccuracies, which result in the measured distortion to be somewhat below the predicted value.

The agreement between the measurements and analysis is similar when the same PA is operated with IS-95 waveforms at the same carrier frequency as shown in figure 8.5. The expected ACPR differs by less than 2 dB compared to the measurements.
Figure 8.3: Ratio of the first sidelobe and the mainlobe versus output power level for $m = 78$, $V_{CC} = 3.5 \text{ V}$, $I_{Q} = 176 \text{ mA}$, $f_c = 2.0 \text{ GHz}$

Figure 8.4: Ratio of the first sidelobe and the mainlobe versus current consumption at $P_{out} = 24 \text{ dBm}$ for $m = 78$, $V_{CC} = 3.5 \text{ V}$, $f_c = 2.0 \text{ GHz}$
Figure 8.5: ACPR versus output power level for IS-95 modulation, $m = 78$, $V_{CC} = 3.5$ V, $I_{cQ} = 176$ mA, $f_c = 2.0$ GHz

### 8.2 Observations

Measurements show that the spectral regrowth at 24 dBm output power level improves by up to 1 dB when the supply voltage is raised to 3.5 V. Some improvement can be still be observed even when the output is reduced to below 22 dBm. This indicates that the linearity of the PA is adversely affected when the collector-base junction becomes almost forward biased, even if the power transistor does not actually go into saturation. This might explain some of the discrepancy between the measurements and analysis. This observation underscores the need to minimize the voltage swing across the collector-base junction, which is very sensitive to the imaginary of the load seen by the collector according to simulations. A small positive imaginary part at that node easily increases the voltage swing and pushes
Figure 8.6: Electric field and the variation of length of the collector-base space charge region for (a) Low collector dopant density (b) High collector dopant density (c) Shallow buried layer

the transistor towards saturation. The results from the simulations are also supported by the measurements. It has been observed that the length of the transmission line used in the output matching network can be changed by an amount which barely affects the output return loss $S_{22}$, but increases the spectral regrowth considerably.

As mentioned in the previous section, the analysis show that the increase in $\tau_{BC}$ at high collector current densities can easily become the dominant source of nonlinearity in
a power amplifier. Therefore, a circuit designer must make sure the total collector area of the power transistor is large enough. The spectral regrowth can also be improved by some modifications to the power transistor, such as increasing the collector dopant density, but this increases the maximum electric field. Therefore, the device breakdown voltage may become too small. Instead, bringing the highly doped buried layer closer to the collector-base junction would still limit the $\tau_{BC}$ variation, as shown in figure 8.6, but it would prevent the maximum electric field and the device breakdown voltage from increasing as much. A similar approach is to gradually increase the collector dopant concentration away from the base-collector junction [20].

During measurements, the amount of spectral regrowth has frequently been observed to be more pronounced in the upper or lower sidebands. Although the Volterra series can be used to explain the asymmetry in spectral regrowth qualitatively [44], the asymmetry has been measured to be about twice (in dB scale) as large as the predicted amount. Accuracy of the predictions may improve if a nonlinear model of the local bias circuit is used and a more sophisticated circuit model is developed to represent the coupling between different parts of the PA through the circuit board.
Chapter 9

Conclusion

This thesis has presented an analysis of radio frequency linear power amplifiers with a particular focus on predicting spectral regrowth. Traditionally, this has not been a well characterized phenomenon, although it is one of the most important specifications in modern wireless communication systems. Lack of an adequate theory has made it difficult for circuit designers to find the optimum device operating point trade-off between efficiency and linearity.

Part I of the thesis reviewed previous research about spectral regrowth, followed by a review of important issues in power amplifier design. Part II started out by introducing Volterra series and explaining how to calculate the frequency-domain Volterra kernels of a circuit. Although Volterra series are quite useful for modeling nonlinearities with memory, obtaining numerical results from such a model by means of the traditional time or frequency-domain computations take too long to be practical for predicting spectral regrowth. Therefore, this thesis has developed a computational method based on decompo-
tion of Volterra kernels encountered during an analysis of a circuit into simpler subsystems. This allowed the representation of the high-order kernels by a combination of linear blocks with memory and nonlinear blocks without memory. The response of the linear blocks can be calculated in the frequency-domain, which is simply the multiplication of two arrays of complex numbers. As the nonlinear blocks lack memory, their responses were computed by means of time-domain calculations at an upconversion frequency much lower than the actual carrier frequency. This has reduced the required number of samples from the input signal waveform and increased the speed of the computation dramatically. The order of the overall computation is limited by the FFT and inverse-FFT calculations, performed between the linear and nonlinear blocks.

The new computational approach described was applied to the design of PAs using bipolar transistors as the active element. Therefore, the nonlinear bipolar transistor model used in the analysis has been presented. The differences between the traditional assumptions used for an analysis of a small-signal amplifier and additional requirements for a power amplifier were pointed out. One of the significant differences is the increase in the forward-transit time $\tau_F$ at high collector current densities. For a small-signal amplifier it may be adequate to assume $\tau_F$ to be constant, but the large current swings in a power transistor cause it to vary considerably during each signal swing at high output power levels. Therefore, a series-based model was developed to represent this variation for more accurate analysis.

Part III presented an implementation example. A number of single-stage power amplifiers were designed with SiGe bipolar transistors and flip-chip packaging. The local
bias circuit has been modified in order to let the power amplifier increase its average collector current at high signal levels and reduce gain compression. The low impedance of the modified local bias circuit also raises the breakdown voltage. The simulation method using the Volterra-series-based power-amplifier model and the decomposition of the Volterra kernels has been implemented in Matlab. The results from the analysis were compared to the measurements and it was shown that the analysis including the $\tau_F$ model predicted the measured sidelobe growth quite accurately, while only requiring minimal computation time. The analysis helped identify the transistor components and design parameters contributing most to distortion.

9.1 Future Research

The drive to reducing the cost of transceivers further is making the implementation of PAs in high-volume, low-cost silicon-based metal-oxide-semiconductor field-effect transistor (MOSFET) technology an attractive approach. Designing PAs in CMOS technology would also allow the integration of the baseband circuits and digital signal processors in a wireless communication system with the analog radio-frequency circuits on a single die. Therefore, an obvious extension of the present work is the application of this new computational approach to MOSFET PAs. The challenge here is representing the drain current as a function of the gate-source and drain-source voltages, in a continuous fashion and accurate up to the third derivative.

The computational approach presented in this thesis can also be applied to multi-stage PAs. However, increasing numbers of nodes and high-order generators increases the
complexity of the frequency response calculations presented in section 4.1. In this case, merging the present work with a more sophisticated CAD tool, which can calculate the frequency-dependent admittance matrix of a given circuit would be very helpful. This would enable designers to analyze different topologies without going through time-consuming hand calculations for finding the circuit kernels. This would also make it easier to model the board and substrate coupling effects by more complicated, but more accurate subcircuits. A more sophisticated model of the coupling between different parts of the circuit and a nonlinear model of the local bias circuit may better explain the reasons for spectral regrowth asymmetry, which was touched upon in chapter 8.
Bibliography


[14] Ravi Gupta, Brian M. Ballweber, and David J. Allstot. Design and optimization of


Appendix A

Programs for Implementing the Computational Approach

A.1 Generating the Input Signal

In this section, the program used for generating the input signal waveform is presented. It generates a symmetric power spectral density, where the ratio of the mainlobe to either one of the adjacent sidelobes is 40.34 dB. Once the waveform is generated, it is saved under the name “signal802.mat” so that it can be loaded by the nonlinear system simulations later.

```matlab
clear all
close all
format short
```
\[
\begin{align*}
k &= 8 \times 8 \quad \% \text{oversampling ratio} \\
\text{bits} &= 1024 \\
\text{offset} &= 0 \\
N &= k \times \text{bits} \quad \% \text{total number of samples}
\end{align*}
\]

% Generating the I-channel rectangular data pulses
get Ibits
for bitno = 1:bits
    start = (bitno - 1) \times k + 1 \\
    stop = bitno \times k \\
    \text{input}_I(\text{start:stop}) = \text{Ibits}(\text{bitno}+\text{offset}) \\
end
If = \text{fftshift}(\text{fft(input}_I)) \\

% Generating the Q-channel rectangular data pulses
get Qbits
for bitno = 1:bits
    start = (bitno - 1) \times k + 1 \\
    stop = bitno \times k \\
    \text{input}_Q(\text{start:stop}) = \text{Qbits}(\text{bitno}+\text{offset}) \\
end
Qf = \text{fftshift}(\text{fft(input}_Q)) ;
The programs “get_Ibits.m” and “get_Qbits.m” above are used to load 1024-bit long streams of data varying between 0.5 and -0.5 for generating the variables “Ibits” and “Qbits”. These are the data streams used in the arbitrary waveform generator. Similar streams of random data can also be generated by a MATLAB command such as “Ibits = (rand(bits,1)<0.5) - 0.5”.

% Calculating the frequency response of the 7th order analog Bessel filter for IEEE802.11b

[Z,P,K] = besself(7,1);

% The result of this MATLAB command is given below:

% Z = [] ;
% P = [ -0.49669172566723 + 1.00250850845442i
% -0.49669172566723 - 1.00250850845442i
% -0.75273554340932 + 0.65046963055226i
% -0.75273554340932 - 0.65046963055226i
% -0.88000293415234 + 0.32166527623077i
% -0.88000293415234 - 0.32166527623077i
% -0.91948715564903 ] ;
% K = 1 ;

[Nb,Db] = zp2tf(Z,P,K);
wstop = 0.581*k ; % Scaling the filter cut-off frequency
wstart = -1*wstop ;
wstep = 2*wstop/N ;
w = wstart:wstep:wstop-wstep ;
H_bessel = polyval(Nb,i*w)./polyval(Db,i*w) ;

% Baseband filtering
filtI = If.*H_bessel;
filtQ = Qf.*H_bessel;

% Upconversion
dt = 1/k ;
t = 0:dt:bits-dt ;
f1 = 9362/bits ; % Simulation upconversion frequency
It = ifft(fftshift(filtI)) ;
Qt = ifft(fftshift(filtQ)) ;
HPf = fftshift(fft( It.*cos(2*pi*f1*t) - Qt.*sin(2*pi*f1*t) )) ;
save signal_802 HPf

The frequency spectrum is then divided into seven roughly equal length segments to represent the frequency bands around $\omega = 0$, $\omega = \pm \omega_c$, $\omega = \pm 2\omega_c$ and $\omega = \pm 3\omega_c$, where $\omega_c$ is the actual carrier frequency. Afterwards, the normalized power spectral density is plotted and the ratio of the adjacent sidelobe to the mainlobe is calculated.
% Dividing the spectrum into 7 roughly equal length segments

fe = -k*11e6/2;

no_seg = 9362;

df = 2*abs(fe)/N;

fseg = ( -1*no_seg/2 + (0:(no_seg-1)) )*df;

fc = 2.45e9; % Actual carrier frequency

ftmp = [fseg-3*fc fseg-2*fc fseg-fc fseg fseg+fc fseg+2*fc fseg+3*fc ];

last = length(ftmp);

f0 = [ ftmp(1)-df ftmp ftmp(last)+df ];

ind1 = N/2 + no_seg/2 + 1;

ind2 = ind1 + no_seg - 1;

% Calculating the power spectral density

POWd = 0.5*(abs(HPf).^2)/N^2;

figure(1)

plot(10*log10(POWd),'-')

grid on

% Zooming into the close-in sidelobes

figure(2)
plot(f0,10*log10(POWd),'-')
axis([ f0(ind1) f0(ind2) -150 -10 ])
title('Power Spectral Density vs Frequency (k = 64 & 1024 bits)')
grid on

maxP = max(10*log10(POWd(ind1:ind2)));
dummy = find( f0-fc < 22e6 & f0-fc > 11e6 );
ind_adj1 = dummy(1);
ind_adj2 = dummy(length(dummy));

% Calculating the ratio of the mainlobe to the upper adjacent sidelobe
ratio_adj_up = max(dB10(POWd(ind_adj1:ind_adj2))) - maxP

dummy = find( f0-fc > -22e6 & f0-fc < -11e6 );
ind_adj1 = dummy(1);
ind_adj2 = dummy(length(dummy));

% Calculating the ratio of the mainlobe to the lower adjacent sidelobe
ratio_adj_down = max(10*log10(POWd(ind_adj1:ind_adj2))) - maxP

A.2 Spectral Regrowth Calculations

In this section, the program used for simulating the power amplifier and calculating the output power spectral density is presented. After the main simulation program is introduced, the programs called by the main one will be presented.
close all

clear all

% format long e

format short

global IQ

% Declaring the simulation parameters

Pin_desired = 17.0 ; % Desired input power in units of dBm

IQ = 0.25 ; % Collector current in amperes

fc = 2.45e9 ; % Actual carrier frequency

k = 8*8 ; % oversampling ratio

bits = 1024 ; % number of bits

N = k*bits ; % total number of points in DFT

% Dividing the spectrum into 7 roughly equal length segments

fe = -k*11e6/2;

no_seg = 9362 ;

df = 2*abs(fe)/N ;

fseg = ( -1*no_seg/2 + (0:(no_seg-1)) )*df ;

f_tmp = [fseg-3*fc fseg-2*fc fseg-fc fseg fseg+fc fseg+2*fc fseg+3*fc ] ;
last = length(f_tmp) ;

f0 = [ f_tmp(1)-df f_tmp f_tmp(last)+df ] ;

f0(N/2 + 1) = 1e-3 ;

% In order to avoid dividing by zero in some equations the zero hertz
% in the mid-point of the spectrum is replaced by 1 milihertz.

w0 = 2*pi*f0 ;

PA design and transistor model parameters are loaded through the program
“param_model.m” in the following part of the main simulation program. Afterwards, a
program can be called for estimating the rise in average collector current based on the
input voltage swing. Then, the program “param_model.m” should be called again, once the
variable “IQ” is updated to be equal to the average collector current $I_{Cave}$. Right now, the
program assumes that $I_{Cave}$ has already been estimated by other means, such as a more
sophisticated CAD tool.

Once the input signal waveform is loaded, the program “scale_signal.m” is called
to calculate the scaling coefficient “scale” and “inp” is multiplied by “scale” so that the
power applied to the power transistor is equal to “Pin_desired”. The scaled input signal is
stored as an array of complex numbers called “Xf”.

Then the first-order responses are calculated by using some of the linear transfer
functions. Calculating the third-order response of the Volterra system is more complicated,
so the program “third_response.m” is called.

param_model ;
load signal_802

inp = HPf;       % frequency-domain input current source signal waveform
scale_signal;

% Calculation of First-Order Responses
Vin1 = Hin1(w0).*Xf;
Iin1 = (Vin1 - HB1(w0).*Xf)/rb + Vin1./Zbias(w0);
Vout1 = Hout1(w0).*Xf;
% Calculation of Third-Order Responses
third_response

The frequency-domain output of the FFT is a signal of value \( N/2 \) times the amplitude of the time-domain signal. Therefore, the FFT output is scaled before the input or output power is calculated. Then, the program “calc_loss.m” calculates the amount of power loss before the signal reaches the actual load through the transistor parasitics, on-chip wiring and package parasitics as well as the output matching network, so that the spectral regrowth predictions at certain output power levels can be compared to the measurement results. The power loss through the input matching network, package and on-chip wiring parasitics is calculated as well, so that the predicted PA gain can be compared to the measured power gain.

% Input and Output Power
POW_in = 2*real(Vin1.*conj(Iin1))/N^2;
ind1 = N/2 + no_seg/2 + 1;
ind2 = ind1 + no_seg - 1;
Pin_norton = 10*log10( sum(POW_in(ind1:ind2))/1e-3 ) ;

Vout = Vout1 + Vout3;
Iout = Vout./ZLs;
POW_out = 2*real(Vout.*conj(Iout))/N^2;
Pout_inband = 10*log10( sum(POW_out(ind1:ind2))/1e-3 ) ;
calc_loss
Pin_source = Pin_norton + input_loss dB
Pout_load = Pout_inband - output_loss dB
Vgain = abs(Vout(N/2 + 2))/abs(Vin1(N/2 + 2))
Pgain = Pout_load - Pin_source

% Output power spectral density calculations
figure(1)
plot(f0,10*log10(POW_out),'-')
axis([ f0(ind1) f0(ind2) -150 -10 ])
title('Power Spectral Density vs. Frequency (k = 64 & 1024 bits )')
grid on

maxP = max(dB10(POW_out(ind1:ind2))) ;
dummy = find( f0-fc < 22e6 & f0-fc > 11e6 ) ;
A.2.1 Entering the Design and Model Parameters

The following program “param_model.m” is used to load most of the PA design and transistor model parameters. These parameters are changed substantially, so that the trade secrets of Maxim Integrated Products do not get disclosed.

```
global m beta Vt Is rb rc

global Re Le

global Rs Lin Cin

global RL Lm Cm Q ro

global gm rpi Cpi Cu

global Kcpi2 Kgm2 Kcu2 Kcpi3 Kgm3 Kcu3

% Design parameters

m = 90 ;       % number of parallel transistors
```
$V_{cc} = 3.5$;

temp = 40;  \quad \% \text{Temperature in C}

Rs = 50;  \quad \% \text{The source resistance}

if $fc \geq 2.4e9$
  \% Input match for $w=2.45\text{GHz}$
  Lin = 1.5e-9;
  Cin = 2.7e-12;
else
  \% Input match for $w=2\text{GHz}$
  Lin = 1.5e-9;
  Cin = 3.5e-12;
end

RL = 50;  \quad \% \text{The load resistance}

ro = 25/IQ ;  \quad \% \text{The output impedance}

Q = 25;

if $fc \geq 2.4e9$
  \% Output match for $w = 2.45\text{GHz}$
  Lm = 1e-9;
  Cm = 3.4e-12;
else
% Output match for w = 2GHz

Lm = 1.15e-9;

Cm = 4.5e-12;

end

Re = 22/m + 2.0/m + 0.02 + 12e-3 ;

% includes the thermal runaway resistor, re of the transistor,
% package parasitics (downbond or solder bump resistance) and
% resistance of the on-chip wiring

Le = 50e-12 + 85e-12 + 30e-12 ;

% includes package parasitics (downbond or solder bump inductance)
% the inductance of the ground vias and the on-chip wiring

% Model parameters

boltzman = 1.38062e-23;

qc = 1.602e-19; % Charge of an electron in coulombs

Vt = boltzman*(temp+273)/qc ;

beta = 100 ; % AC beta

Is = m*4e-17;

nF = 1.00;
TF = 5.0e-12;
cje = 1.5e-13*m;
vje = 0.9;
mje = 0.5;

Cjco = m*2.75e-14;
Vjc = 0.85;
mjc = 0.5;
Cu_res = m*0.3e-14;

rb = 10/m;
rc = 5.0/m;

% calculation of gm, rpi, Cpi and Cu
gm = IQ/(nF*Vt) ;
rpi = beta/gm ;
VBE = nF*Vt*log(IQ/(m*Is)) ;
Cpi = TF*gm + cje/(1-VBE/vje)^mje ;
VCB = Vcc - VBE - IQ*(Re+rc) ;
Cu = Cjco/(1+VCB/Vjc)^mjc + Cu_res ;

% Calculation of higher order current source coefficients
\begin{align*}
K_{gm2} &= \frac{gm}{2nFVT} \\
K_{gm3} &= \frac{gm}{6(nFVT)^2} \\
K_{cp12} &= TFK_{gm2} + 0.5\left(cje/(1-VBE/vje)^{mje}\right)mje/(vje-VBE) \\
K_{cp13} &= TFK_{gm3} + \left(cje/(1-VBE/vje)^{mje}\right)mje\left(mje+1\right)/(6(vje-VBE)^2) \\
K_{cu2} &= -0.5\left(Cjco/(1+VCB/Vjc)^{mjc}\right)mjc/(VCB+Vjc) \\
K_{cu3} &= \left(Cjco/(1+VCB/Vjc)^{mjc}\right)mjc\left(mjc+1\right)/(6(VCB+Vjc)^2) \\
\end{align*}

\% The collector-base space charge region transit time model

\texttt{vsat} = 1e5; \quad % saturated electron drift velocity (m/s)
\texttt{Nc} = 1e22; \quad % collector dopant concentration (1/m^3)

\% The cross-sectional area of the collector through which current flows

\texttt{Ac} = 1.5m*4*10e-6*0.48e-6 \\
\texttt{Jc} = IQ/Ac \\
\texttt{Jmax} = qc\times vsat\times Nc ;
\texttt{Jratio} = Jc/Jmax ;

\texttt{Es} = 11.7*8.854e-12; \quad % permittivity of Si
\texttt{Abc} = m*4*10e-6*0.48e-6 \\
\texttt{xd1} = Es\times Abc/Cu ; \quad % the depletion layer depth

\texttt{Tc} = ( (1-Jratio)^{(-0.5)} - 1 )\times xd1\times qc\times Nc/Jc ;
\texttt{KTF1} = (1-Jratio)^{(-1.5)}\times xd1/(2\times vsat\times Vt) - Tc/Vt ;
\[ KTF2 = \left(\frac{3}{8}\right) \left(\frac{x_d}{v_s a t V_t}\right) \left(\frac{J_{ratio}}{V_t}\right) (1 - J_{ratio})^{-2.5} - 0.5 \frac{KTF1}{V_t}; \]

\[ KTF3a = \left(\frac{5}{16}\right) \left(\frac{x_d}{v_s a t V_t}\right) \left(\left(\frac{J_{ratio}}{V_t}\right)^2\right) (1 - J_{ratio})^{-3.5}; \]

\[ KTF3b = \left(\frac{x_d}{8 v_s a t V_t}\right) \left(\frac{J_{ratio}}{V_t^2}\right) (1 - J_{ratio})^{-2.5}; \]

\[ KTF3 = KTF3a + KTF3b - \frac{KTF2}{3 V_t}; \]

\[ K_{cpi2} = K_{cpi2} + IQ \cdot KTF2; \]

\[ K_{cpi3} = K_{cpi3} + IQ \cdot KTF3; \]

### A.2.2 Scaling the Input Signal Waveform

The following program “scale_signal.m” is used to calculate the scaling coefficient “scale” which is multiplied by the input frequency spectrum, so that the power applied to the power transistor, including the biasing network and transistor parasitics, is equal to “\( P_{in \_desired} \).”

\[ Vin_{temp} = Hin1(w_0) \cdot inp; \]

\[ I_{in\_temp} = \frac{(Vin_{temp} - HB1(w_0) \cdot inp)}{rb} + \frac{Vin_{temp}}{Z_{bias}(w_0)}; \]

% Calculating the unscaled input power

\[ POW_{temp} = 2 \cdot \text{real}(Vin_{temp} \cdot \text{conj}(I_{in\_temp})) / N^2; \]

\[ \text{ind1} = N/2 + \text{no\_seg}/2 + 1; \]

\[ \text{ind2} = \text{ind1} + \text{no\_seg} - 1; \]

\[ Pin_{temp} = 10^\log10(\text{sum}(POW_{temp}(\text{ind1:ind2}))/1e-3); \]

\[ \text{scale} = 10^{((Pin_{\_desired} - Pin_{\_temp})/20)}; \]
Xf = scale*inp ;

### A.2.3 Calculating the Input and Output Loss

The following program “calc_loss.m” is used to calculate the amount of power loss before the signal reaches the actual load through the transistor parasitics, on-chip wiring and package parasitics, as well as the output matching network. First, the signal loss is computed through the current dividers formed by the pad, the matching capacitor “Cm” and the parasitics of “Cm” due to the finite “Q”. Then the power loss is calculated by using the current divider ratios and the ratio of the actual load resistance and the real part of the impedance seen by the intrinsic collector node.

The loss through the input matching network, package and on-chip wiring parasitics is calculated by following a similar strategy. “Pin_{desired}” is applied to the circuit which includes the base resistance “rb” of the transistor and the local biasing network. Therefore, the loss before the signal reaches that point is calculated by using current dividers.

\[
\%
\text{Calculating the power loss at the input}
\]

\[
w_{\text{mid}} = 2\pi f_{c}
\]

\[
X_{\text{pad}} = \frac{1}{i w_{\text{mid}} 500\text{e-15}}
\]

\[
Z_{\text{pack}} = i w_{\text{mid}} 50\text{e-12} + 20\text{e-3}
\]

\[
Z_{p} = i w_{\text{mid}} 272\text{e-12} + 132.5\text{e-3}
\]

\[
z_{\text{cap}} = \frac{1}{i w_{\text{mid}} C_{i}} + \frac{1}{2\pi \text{2.45e9}\text{C}_{i} Q}
\]
\(z_{\text{ind}} = 1.\left(1/(i*w_{\text{mid}}*L_{\text{in}}) + 1/(2*\pi*2.45e9*L_{\text{in}}*Q)\right)\);

\(Z_{\text{in}} = 1.\left(1/H_{\text{in}}(w_{\text{mid}}) - 1/Z_{\text{A}}(w_{\text{mid}})\right)\);

\(Z_{\text{inA}} = 1.\left(1/Z_{\text{in}} + 1/Z_{\text{bias}}(w_{\text{mid}})\right)\);

\(Z_{\text{inx}} = Z_{\text{p}} + Z_{\text{inA}}\);

\(z_{1} = z_{\text{cap}} + Z_{\text{pack}} + 1/(1/X_{\text{pad}} + 1/Z_{\text{inx}})\);

\(z_{\text{in}} = z_{\text{ind}} + 1/(z_{\text{inx}} + z_{\text{in}})\);

\(x_{1} = z_{\text{ind}}/(z_{\text{in}} + z_{\text{ind}})\);

\(x_{2} = X_{\text{pad}}/(X_{\text{pad}} + Z_{\text{inx}})\);

\(P_{\text{in\_ratio}} = (\text{abs}(x_{1}\times x_{2})^2)\times \text{real}(Z_{\text{inA}})\times \text{real}(z_{\text{in\_total}})\);

\(\text{input\_loss\_dB} = -10\times \log_{10}(P_{\text{in\_ratio}})\)

\% Calculating power loss at the output

\(R_{\text{Lp}} = 1/(1/R_{\text{L}} + 2*\pi*2.45e9*C_{\text{m}}/Q)\);

\(Z_{L_{\text{p}}} = R_{\text{Lp}} + 1/(i*w_{\text{mid}}*100e-12)\);

\(z_{1} = 1.\left(1/Z_{L_{\text{p}}} + 1/(i*w_{\text{mid}}*12e-9) + i*w_{\text{mid}}*C_{\text{m}} + i*w_{\text{mid}}*L_{\text{m}} + 2*\pi*2.45e9*L_{\text{m}}/Q\right)\);

\(z_{2} = 1.\left(1/X_{\text{pad}} + 1/z_{1}\right) + 45e^{-3} + \text{rc} + i*w_{\text{mid}}*130e^{-12}\);

\(x_{1} = X_{\text{pad}}/(X_{\text{pad}} + z_{1})\);
x2 = (1/(i*wmid*Cm))/( 1/(i*wmid*Cm) + ZLp ) ;

x3 = ( Q/(2*pi*2.45e9*Cm) )/( Q/(2*pi*2.45e9*Cm) + RL ) ;

P_ratio = (abs(x1*x2*x3)^2)*RL/real(z2) ;

output_loss_DB = -10*log10(P_ratio)

A.2.4 Calculating the Third-Order Response

The program “third_response.m” is used to calculate the third-order response of the PA based on the decomposition of the third-order Volterra kernel into simpler subsytems. The functions called by this program for calculating the impedance of some of the circuit elements, the transfer functions and expressions for the second and third-order expressions are presented in the next subsection.

% Calculating the impedances and coefficients

global ZLs Zpis Zus ZAs N1s N2s N3s N4s N5s

ZLs = ZL(w0) ;
Zpis = Zpi(w0) ;
Zus = Zu(w0) ;
ZAs = ZA(w0) ;
N1s = N1(w0) ;
N2s = N2(w0) ;
N3s = N3(w0) ;
N4s = N4(w0) ;
\[ N_5s = N_5(w_0) \]
\[ w_{HTpi\_outs} = w_{HTpi\_out}(w_0) \]
\[ H_{Trpi\_outs} = H_{Trpi\_out}(w_0) \]
\[ H_{Tgm\_outs} = H_{Tgm\_out}(w_0) \]
\[ w_{HTu\_outs} = w_{HTu\_out}(w_0) \]

% Pure third order

\[ V_{BE1} = H_{BE1}(w_0) \cdot X_f \]
\[ V_{BE1t} = \text{ifft} (\text{fftshift}(V_{BE1})) \]
\[ V_{BE3} = \text{fftshift} (\text{fft}(V_{BE1t} \cdot 3)) \]

\[ V_{CB1} = H_{CB1}(w_0) \cdot X_f \]
\[ V_{CB1t} = \text{ifft} (\text{fftshift}(V_{CB1})) \]
\[ V_{CB3} = \text{fftshift} (\text{fft}(V_{CB1t} \cdot 3)) \]

\[ V_{out\_pi3} = K_{cpi3} \cdot V_{BE3} \cdot w_{HTpi\_outs} \]
\[ V_{out\_rpi3} = (K_{gm3}/\beta) \cdot V_{BE3} \cdot H_{Trpi\_outs} \]
\[ V_{out\_gm3} = K_{gm3} \cdot V_{BE3} \cdot H_{Tgm\_outs} \]
\[ V_{out\_u3} = K_{cu3} \cdot V_{CB3} \cdot w_{HTu\_outs} \]

% Second order interaction at B-E junction

% between the signals from second-order sources
% and the first-order waveform
VBE2 = fftshift(fft(VBE1t.^2)) ;
VCB2 = fftshift(fft(VCB1t.^2)) ;

Vtmp1 = ifft(fftshift(VBE2.*HTgm_BE(w0))) ;
Vtmp2 = ifft(fftshift(VBE2.*wHTpi_BE(w0))) ;
Vtmp3 = ifft(fftshift(VBE2.*HTrpi_BE(w0))) ;
Vtmp4 = ifft(fftshift(VCB2.*wHTu_BE(w0))) ;
Vtmp5 = Vtmp1 + Vtmp2 + Vtmp3 + Vtmp4 ;

Vtmp6 = fftshift(fft(Vtmp5.*VBE1t)) ;
Vout_gm2 = 2*Kgm2*Vtmp6.*HTgm_outs ;
Vout_pi2 = 2*Kcpi2*Vtmp6.*wHTpi_outs ;
Vout_rpi2 = 2*(Kgm2/beta)*Vtmp6.*HTrpi_outs ;

% Second order interaction at C-B junction
% between the signals from second-order sources
% and the first-order waveform
Vtmp1 = ifft(fftshift(VBE2.*HTgm_CB(w0))) ;
Vtmp2 = ifft(fftshift(VBE2.*wHTpi_CB(w0))) ;
Vtmp3 = ifft(fftshift(VBE2.*HTrpi_CB(w0))) ;
Vtmp4 = ifft(fftshift(VCB2.*wHTu_CB(w0))) ;
Vtmp5 = Vtmp1 + Vtmp2 + Vtmp3 + Vtmp4;

Vtmp6 = fftshift(fft(Vtmp5.*VCB1t)); Vout_u2 = 2*Kcu2*Vtmp6.*wHTu_outs;
% Total
Vout3 = Vout_gm3 + Vout_rpi3 + Vout_pi3 + Vout_u3;
Vout3 = Vout3 + Vout_gm2 + Vout_pi2 + Vout_rpi2 + Vout_u2;

A.2.5 First-Order Response Functions

% Calculating ZA, the parallel combination of local biasing
% circuit, source resistance, input matching network,
% on-chip wiring and package parasitics.
function z = ZA(w)
  global Rs Lin Cin Q

  Zp = i*w*272e-12 + 132.5e-3;
  Xpad = 1./(i*w*500e-15);
  Zpack = i*w*50e-12 + 20e-3;

  Zs = 1./((1/Rs + 1./(i*w*Lin)) + 1/(2*pi*2.45e9*Lin*Q));

  zser = Zs + 1./(i*w*Cin) + 1./(2*pi*2.45e9*Cin*Q) + Zpack;
\[ z_1 = \frac{1}{(1/z_{ser} + 1/X_{pad}) + Z_p} \]
\[ z = \frac{1}{(1/z_1 + 1/Z_{bias}(w))} \]

% Calculating Zbias, impedance of local biasing circuit

function z = Zbias(w)

wa = abs(w);

fw0 = (wa < 2*pi*1e8);
fw1 = (wa > 2*pi*1.8e9).*((wa < 2*pi*2.65e9));
fw2 = (wa > 2*pi*3.6e9).*((wa < 2*pi*5.3e9));
fw3 = (wa > 2*pi*5.4e9).*((wa < 2*pi*8e9));

R = 80;
C = (1.83e-11)/(2*pi);
L = (0.6e-5)/(2*pi);

% z = 1/(i*w*C + 1/(i*w*L) + 1/R) ; at low frequency

z = (1/(i*w*C + 1/(i*w*L) + 1/R) + 11.5).*fw0 + (55-8i)*fw1 + (52-8.5i)*fw2 + (48-8i)*fw3;

% Calculating ZE, emitter degeneration

function z = ZE(w)
global Re Le

z = Re + i*w*Le;

% Calculating ZL, load seen by intrinsic collector of transistor
function z = ZL(w)
global rc RL Cm Lm Q ro

ZL = RL + 1./(i*w*100e-9);
Zlp = 1./( 1./ZL + 2*pi*2.45e9*Cm/Q ) ;
z1 = 1./(1./Zlp + 1./(i*w*12e-9) + i*w*Cm) + i*w*Lm + 2*pi*2.45e9*Lm/Q ;
z = 1./(i*w*500e-15 + 1./z1 ) + 45e-3 + rc + i*w*130e-12 ;

% Calculating Zpi, base-emitter impedance
function z = Zpi(w)
global rpi Cpi
z = rpi./( 1 + i*w*Cpi*rpi ) ;

% Calculating Zu, impedance of collector-base junction capacitance
function z = Zu(w)
global Cu
z = 1./(i*w*Cu) ;
function z = Zpar(w)
global gm
z = 1./( 1./Zpi(w) + 1./ZE(w) + gm );

function coef = M1(w)
global gm
coef = ZE(w).*(Zu(w) + ZL(w))./( ZL(w).*(gm*Zpar(w).*Zu(w) - ZE(w)) );

function coef = M2(w)
global rb
coef = rb*( 1./Zu(w) + M1(w).* 1/rb + 1./Zu(w) + Zpar(w)./( Zpi(w).*ZE(w) ) );

function H = Hout1(w)

% Calculating the impedance of the parallel combination of Zpi, ZE and 1/gm

% Calculating M1, a common calculation for first-order response functions

% Calculating M2, another common calculation for first-order response functions

% Hout1 is the 1st order transfer function between input current signal source and output voltage
global rb

H = 1./( M1(w)/rb - M2(w).*((1./ZA(w) + 1/rb) ) ) ;

% Hin1 is the 1st order transfer function
% between input current signal source and input voltage
function H = Hin1(w)
    H = -1*Hout1(w).*M2(w) ;

% HB1 is the 1st order transfer function
% between input current signal source and base voltage
function H = HB1(w)
    H = -1*Hout1(w).*M1(w) ;

% HBE1 is the 1st order transfer function between
% input current signal source and base-emitter voltage
function H = HBE1(w)
    H = HB1(w).*Zpar(w)./ZE(w) ;

% HCB1 is the 1st order transfer function between
% input current signal source and collector-base voltage
function H = HCB1(w)
    H = Hout1(w) - HB1(w) ;
A.2.6 Second-Order Transfer Functions

% Calculating N1, a common calculation
% for high-order transfer functions
function coef = N1(w)
    coef = 1 + ZL(w)./Zu(w) ;

% Calculating N2, another common calculation
% for high-order transfer functions
function coef = N2(w)
    global gm
    coef = ( 1./Zpi(w) + 1./ZE(w) + gm )./( gm*ZL(w)./Zu(w) + N1(w)./Zpi(w) ) ;

% Calculating N3, yet another common calculation
% for high-order transfer functions
function coef = N3(w)
    global gm rb
    coef = 1./Zu(w) + gm*ZL(w)./Zu(w) + N1(w).* ( 1./Zpi(w) + 1./(rb+ZA(w)) ) ;

% Calculating N4, one more common calculation
% for high-order transfer functions
function coef = N4(w)
global gm rb N1s N2s

coef = (1-N1s.*N2s)./Zpi(w) + gm*(1-N2s.*ZL(w)./Zu(w)) - N2s./Zu(w) - N1s.*N2s./(rb + ZA(w));

% Calculating N5, the last common calculation
% for high-order transfer functions
function coef = N5(w)

global gm rb

coef = gm*ZL(w)./Zu(w) + N1(w)./Zpi(w);

% HTgm_BE is the transfer function between
% gm non-linear generator and base-emitter voltage
function H = HTgm_BE(w)

global rb gm ZLs Zpis Zus ZAs N1s N2s N3s N4s N5s Kgm2

HB = (N2s.*ZLs./Zus - 1)./N4s;
HE = (ZLs./Zus + HB.*N3s)./N5s;
H = Kgm2*(HB - HE);

% HTgm_CB is the transfer function between
% gm non-linear generator and collector-base voltage
function H = HTgm_CB(w)

global rb gm ZLs Zpis Zus ZAs N1s N2s N3s N4s N5s Kgm2
HB = (N2s.*ZLs./Zus - 1)./N4s ;

HE = (ZLs./Zus + HB.*N3s)./N5s ;

HC = -1*ZLs.*( 1 + HB.*(1./Zpis + 1./(rb+ZAs) + gm)
- HE.*(1./Zpis + gm) ) ;

H = Kgm2*( HC - HB ) ;

% HTrpi_BE is the transfer function between
% rpi non-linear generator and base-emitter voltage
function H = HTrpi_BE(w)

global rb gm beta ZLs Zpis Zus ZAs N1s N2s N3s N4s N5s Kgm2

HB = (N1s.*N2s - 1)./N4s ;

HE = (N1s + HB.*N3s)./N5s ;

H = (Kgm2/beta)*( HB - HE ) ;

% HTrpi_CB is the transfer function between
% rpi non-linear generator and collector-base voltage
function H = HTrpi_CB(w)

global rb gm beta ZLs Zpis Zus ZAs N1s N2s N3s N4s N5s Kgm2

HB = (N1s.*N2s - 1)./N4s ;

HE = (N1s + HB.*N3s)./N5s ;

HC = -1*ZLs.*( 1 + HB.*(1./Zpis + 1./(rb+ZAs) + gm)
- HE.*(1./Zpis + gm) ) ;
H = (Kg2/beta)*( HC - HB ) ;

% wHTpi_BE is the transfer function between
% Cpi non-linear generator and base-emitter voltage
function H = wHTpi_BE(w)

global rb gm ZLs Zpis Zus ZAs N1s N2s N3s N4s N5s Kcpi2
HB = (N1s.*N2s - 1)./N4s ;
HE = (N1s + HB.*N3s)./N5s ;
H = Kcpi2*i*w.*( HB - HE ) ;

% wHTpi_CB is the transfer function between
% Cpi non-linear generator and collector-base voltage
function H = wHTpi_CB(w)

global rb gm ZLs Zpis Zus ZAs N1s N2s N3s N4s N5s Kcpi2
HB = (N1s.*N2s - 1)./N4s ;
HE = (N1s + HB.*N3s)./N5s ;
HC = -1*ZLs.*( 1 + HB.*(1./Zpis + 1./(rb+ZAs) + gm) - HE.*(1./Zpis + gm) ) ;
H = Kcpi2*w*i.*( HC - HB ) ;

% wHTu_BE is the transfer function between
% Cu non-linear generator and base-emitter voltage
function H = wHTu_BE(w)

global rb gm ZLs Zpis Zus ZAs N1s N2s N3s N4s N5s Kcu2

HB = -1*N2s./N4s ;
HE = ( -1 + HB.*N3s )./N5s ;
H = Kcu2*w*i.*( HB - HE ) ;

% wHTu_CB is the transfer function between
% Cu non-linear generator and collector-base voltage
function H = wHTu_CB(w)

global rb gm ZLs Zpis Zus ZAs N1s N2s N3s N4s N5s Kcu2

HB = -1*N2s./N4s ;
HE = ( -1 + HB.*N3s )./N5s ;
HC = -1*ZLs.*( HB.*(1./Zpis + 1./(rb+ZAs) + gm)
- HE.*(1./Zpis + gm) ) ;
H = Kcu2*w*i.*( HC - HB ) ;

A.2.7 Third-Order Transfer Functions

% HTgm_out is the transfer function
% between gm non-linear generator and output
function H = HTgm_out(w)

global rb gm ZLs Zpis Zus ZAs N1s N2s N3s N4s N5s
\[ \begin{align*}
HB &= \frac{(N2s \cdot ZLs / Zus - 1) \cdot N4s}{N4s} ; \\
HE &= \frac{(ZLs / Zus + HB \cdot N3s) \cdot N5s}{N5s} ; \\
H &= -1 \cdot ZLs \cdot (1 + HB \cdot (1/Zpis + 1/(rb+ZAs) + gm) \\
&\quad - HE \cdot (1/Zpis + gm)) ;
\end{align*} \]

% HTrpi_out is the transfer function
% between rpi non-linear generator and output
function H = HTrpi_out(w)

global rb gm ZLs Zpis Zus ZAs N1s N2s N3s N4s N5s
HB = \frac{(N1s \cdot N2s - 1) \cdot N4s}{N4s} ; \\
HE = \frac{(N1s + HB \cdot N3s) \cdot N5s}{N5s} ; \\
H = -1 \cdot ZLs \cdot \text{w} \cdot (1 + HB \cdot (1/Zpis + 1/(rb+ZAs) + gm) \\
&\quad - HE \cdot (1/Zpis + gm)) ;

% wHTpi_out is the transfer function
% between Cpi non-linear generator and output
function H = wHTpi_out(w)

global rb gm ZLs Zpis Zus ZAs N1s N2s N3s N4s N5s
HB = \frac{(N1s \cdot N2s - 1) \cdot N4s}{N4s} ; \\
HE = \frac{(N1s + HB \cdot N3s) \cdot N5s}{N5s} ; \\
H = -1 \cdot ZLs \cdot \text{w} \cdot (1 + HB \cdot (1/Zpis + 1/(rb+ZAs) + gm) \\
&\quad - HE \cdot (1/Zpis + gm)) ;
% wHTu_out is the transfer function
% between Cu non-linear generator and output

function H = wHTu_out(w)

global rb gm ZLs Zpis Zus ZAs N1s N2s N3s N4s N5s

HB = -1*N2s./N4s ;
HE = ( -1 + HB.*N3s )./N5s ;

H = -1*ZLs.*w*i.* ( HB.*(1./Zpis + 1./(rb+ZAs) + gm) )
- HE.*(1./Zpis + gm) );