CMOS RF POWER AMPLIFIERS FOR MOBILE WIRELESS COMMUNICATIONS

A Dissertation
Presented to
The Academic Faculty

By

Kyu Hwan An

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
December 2009

CMOS RF POWER AMPLIFIERS FOR MOBILE WIRELESS COMMUNICATIONS

Approved by:

Dr. Joy Laskar, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Kevin T. Kornegay
School of Electrical & Computer Engineering
Georgia Institute of Technology

Dr. John D. Cressler
School of Electrical & Computer Engineering
Georgia Institute of Technology

Dr. Paul A. Kohl
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Dr. Emmanouil M. Tentzeris
School of Electrical & Computer Engineering
Georgia Institute of Technology

Date Approved: November 2, 2009

ACKNOWLEDGEMENT

First of all, I would like to appreciate the monumental support and the invaluable opportunity that my advisor, Prof. Joy Laskar, gave me to work in this research. Without his inspiration and teaching, I certainly would not have achieved this goal.

I would also like to thank Prof. John D. Cressler, Prof. Emmanouil M. Tentzeris, Prof. Kevin T. Kornegay, and Prof. Paul A. Kohl, for their time in reviewing my dissertation and serving as my defense committee members.

I am very grateful to Dr. Chang-Ho Lee for his great support and guidance throughout this study. Furthermore, I am also indebted to Dr. Kyutae Lim for his efforts to provide the best research environment for all students.

Dr. Minsik Ahn and Dr. Dong Ho Lee deserve a special acknowledgement for their comments, help in the design, and their tremendous contributions in this research. I would like to acknowledge Samsung Design Center engineers, Dr. Jae Joon Chang, Dr. Woonyun Kim, Dr. Wangmyong Woo, Dr. Changhyuck Cho, Dr. Yunseo Park, Dr. Seongmo Yim, Dr. Ki Seok Yang, Dr. Jeonghu Han, and Michael Kroger for their assistance.

The support from members of power amplifier group has been invaluable: Hyungwook Kim, Jeongwon Cha, Eungjung Kim, Jihwan Kim, Youngchang Yoon, Hamhee Jeon, Michael Oakley, Hyunwoong Kim, Yan-Yu Huang, Kun Seok Lee, and Kwanyeob Chae. I would like to express my special gratitude to Ockgoo Lee for years of insightful feedback in my research. I would also like to acknowledge all my mates at the Microwave Application Group, Sang Min Lee, Jong Min Park, Kwan-Woo Kim,

Taejoong Song, Sanghyun Woo, Joonhoi Hur, Jaehyouk Choi, Seungil Yoon, Michael Lee, Sungho Beck, Taejin Kim, Hyungsoo Kim, and Kilhoon Lee. I wish to thank Chris Evans, DeeDee Bennett, and Angelika Braig, for their continuous support.

And most of all, I can't express my love and gratitude enough to my wife, Min Jung Park, for her love and support throughout all my life. My daughter, Claire, and other one still waiting for her appearance in this world have also been a great source of joy. I am especially grateful to my parents, Jun Ho An and Cho Ja Kim, and my parents-in-law, Ki Soo Park and Soon Deuk Kim for their unconditional love. Without their encouragement and support, this dissertation would not have been possible. I would also like to recognize my brother, Kyu Cheol An, who also deserves a special thank you for his support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACLR	adjacent channel leakage ratio
ACPR	adjacent channel power ratio
AM-AM	amplitude-amplitude modulation
AM-PM	amplitude-phase modulation
AMPS	advanced mobile phone service
BJT	bipolar junction transistor
CCDF	complementary cumulative density function
CDF	cumulative density function
CDMA	code division multiple access
CE	collector efficiency
CG	common gate
CMOS	complementary metal oxide semiconductor
CS	common source
DAT	distributed active transformer
DE	drain efficiency
EVM	error vector magnitude
FDD	frequency-division duplexing
FET	field effect transistor
FFT	fast Fourier Transformer
GaAs	gallium arsenide

GMSK	Gaussian minimum shift keying
GPRS	general packet radio system
GSM	global system for mobile communications
HB	harmonic balance
HBT	hetero-junction bipolar transistors
HSDPA	high-speed download packet access
IC	integrated circuit
IMD3	third-order intermodulation distortion
IP3	third-order intercept point
LNA	low noise amplifier
LSSP	large signal S-parameter
LTCC	low-temperature co-fired ceramic
NMT	nordic mobile telephone
OFDM	orthogonal frequency division multiplexing
PA	power amplifier
PAE	power-added efficiency
PAPR	peak-to-average power ratio
PAR	peak-to-average ratio
PCB	printed circuit board
PCR	power-combining ratio
PCT	parallel-combining transformer
PDF	probability density function
PGS	patterned ground shielding

P1dB	1dB compression point
RF	radio frequency
SCT	series-combining transformer
TDD	time-division duplexing
TDMA	time-division multiple access
UMTS	universal mobile telecommunications system
VCO	voltage-controlled oscillator
VSWR	voltage standing wave ratio
WCDMA	wideband code division multiple access
WiMAX	worldwide interoperability for microwave access
WLAN	local wireless area network

SUMMARY

The explosive growth of the wireless market has increased the demand for low-cost, highly-integrated CMOS wireless transceivers. However, the implementation of CMOS RF power amplifiers remains a formidable challenge. The objective of this research is to demonstrate the feasibility of CMOS RF power amplifiers by compensating for the RF performance disadvantages of CMOS technology. This dissertation proposes a parallel-combining transformer (PCT) as an impedance-matching and output-combining network. The results of a comprehensive analysis show that the PCT is a suitable solution for watt-level output power generation in cellular applications. To achieve high output power and high efficiency, the work presented here entailed the design of a class-E switching power amplifier in a 0.18-μm CMOS technology for GSM applications and, with the suggested power amplifier design technique, successfully demonstrated a fully-integrated RF front-end consisting of a power amplifier and an antenna switch. This dissertation also proposed an efficiency enhancement technique at power back-off. In an effort to save current in the power back-off while satisfying the EVM requirements, a class-AB linear power amplifier was implemented in a 0.18-μm CMOS technology for WLAN and WiMAX applications using a PCT as well as an operation class shift between class-A and class-B. Thus, the research in this dissertation provides low-cost CMOS RF power amplifier solutions for commercial products used in mobile wireless communications.

CHAPTER 1

INTRODUCTION

1.1. Background

Since the explosive growth of the wireless market took place in the late 1990s, just a decade ago, the top priority for manufacturers of mobile terminals has been to maximize their profits by lowering the cost of mobile terminals. However, to meet the demand for low-cost terminals in the ever-competitive wireless market, manufacturers must be able to facilitate their production of components to shorten the time to market. In the early era of mobile terminals, engineers spent a considerable amount of time assembling and matching components in the signal path to minimize loss. However, as the functions of wireless communication have proliferated and diversified, mobile terminal manufacturers are calling for easier solutions for modular integration, such as display, audio, and communication, in order to save time under pressure. As the wireless market trend continues to accelerate, integrated circuit (IC) manufacturers are taking on the burden of realizing modular functionality by attempting to provide a ready-made solution for cell-phone and wireless terminal designers.

Accordingly, the inevitable task and ultimate goal of the modern wireless communication industry is the full integration of digital, analog, and even radio frequency (RF) functions. To this end, the industry has devoted great effort to designing wireless terminals using a common semiconductor process that utilizes a single chip; thus,

true single-chip radio is now within the grasp of manufacturers. To date, the most successful solution for such demands has been complementary metal oxide semiconductor (CMOS) technology, thanks to its cost-effective material and great versatility.

Lower f_{max} and f_T of CMOS devices are not serious problems at the frequencies used by cellular applications. However, because of the intrinsic drawbacks of standard CMOS processes from the RF perspective, several obstacles, especially low quality factor (Q) passive structures and lossy substrate [1] and low breakdown voltage of active devices [2], are hindering the realization of a fully-integrated CMOS radio. Thus, the implementation of CMOS RF front-ends, such as an RF switch [3] or a RF power amplifier (PA) [4], remains a challenging task. Accordingly, most commercial products are based on gallium arsenide (GaAs) technologies, as shown in Figure 1.

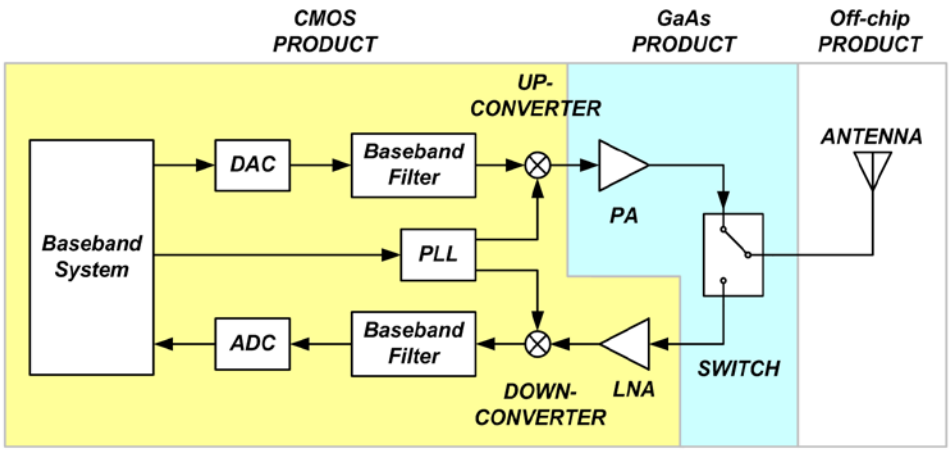


Figure 1. A block diagram of a direct-conversion transceiver and its processes

Today, the various demands from consumers have stoked the development of multiple standards. For example, global system for mobile communications (GSM), code

division multiple access (CDMA), and wideband CDMA (WCDMA) are now used for voice and relatively low data rate communications while local wireless area networks (WLANs) and worldwide interoperability for microwave access (WiMAX) mainly target high data rate communications with varying mobility. The typical output power levels of PAs for some wireless applications are listed in Table 1 and shown in Figure 2 along with their data rates. While many standards require different average output powers due to peak-to-average power ratio (PAPR), the general peak output power requirement for wireless communications is usually 30 dBm to 35 dBm as indicated in Figure 2. Due to the poor linearity performance of CMOS devices, satisfying the key specifications of commercial RF PA products (high output power, high efficiency, and high linearity) with CMOS devices poses a great challenge and even a greater one for emerging wireless communications [4] in which good linearity is a default requirement.

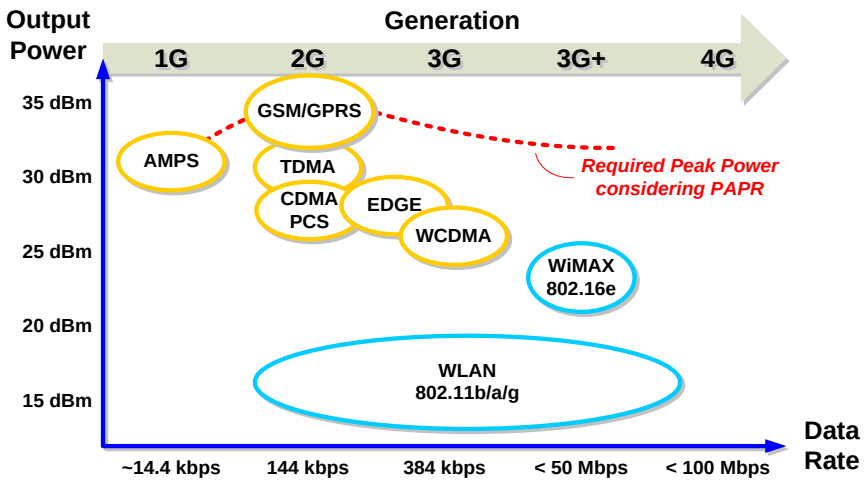


Figure 2. Output power requirements of various standards

Table 1. Typical output power of PAs for some wireless applications

Application	Standard	Frequency (MHz)	Typical Output Power (dBm)	Modulation
Cellular	GSM850	824-849	35	GMSK
	E-GSM900	880-915	35	GMSK
	DCS1800	1710-1785	33	GMSK
	PCS1900	1850-1910	33	GMSK
	CDMA (IS-95)	824-849	28	O-QPSK
	PCS (IS-98)	1750-1780 1850-1910	28	O-QPSK
	WCDMA (UMTS)	1920-1980	27	HPSK
WLAN	IEEE 802.11b	2400-2484	16-20	PSK-CCK
	IEEE 802.11a	5150-5350	14-20	OFDM
	IEEE 802.11g	2400-2484	16-20	OFDM
WiMAX	IEEE 802.16d/e	2300-2700	22-25	OFDM
	IEEE 802.16d/e	3300-3700	22-25	OFDM
	IEEE 802.16d/e	4900-5900	22-25	OFDM

1.2. Motivation

The difficulties of CMOS RF PAs have already been described in the previous section and will be revisited in Chapter III. From a purely performance-oriented standpoint, CMOS technologies are not a good solution. However, they follow an aggressive down-scaling roadmap, shown in Figure 3, that is unbeatable when compared to any other semiconductor technology; hence, the integrability and versatility of CMOS technologies will be welcomed for a long time down the road. Narrowing down the focus to the cost of PAs, CMOS technologies would be the cheapest among other candidates such as III-V HBT, III-V PHEMT, SiGe HBT, and Si-MOSFET technologies, as shown in Figure 4. Thus, it seems inevitable that both customers and manufacturers will choose

CMOS technologies over all others, as it represents a win-win strategy, depicted in Figure 5 [5].

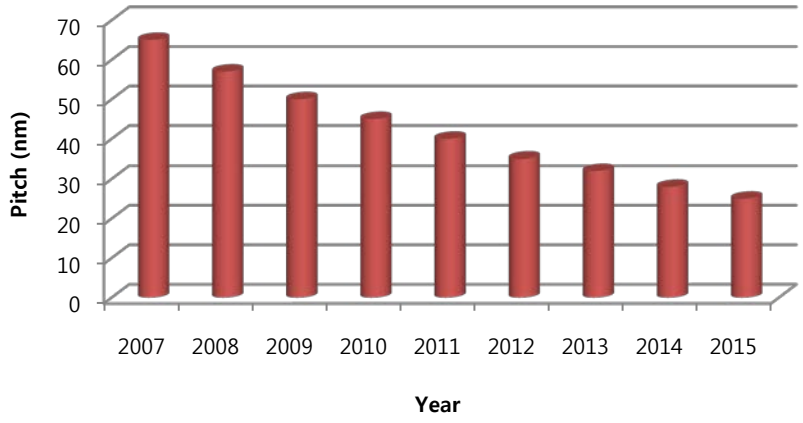


Figure 3. Scaling-down of CMOS technologies

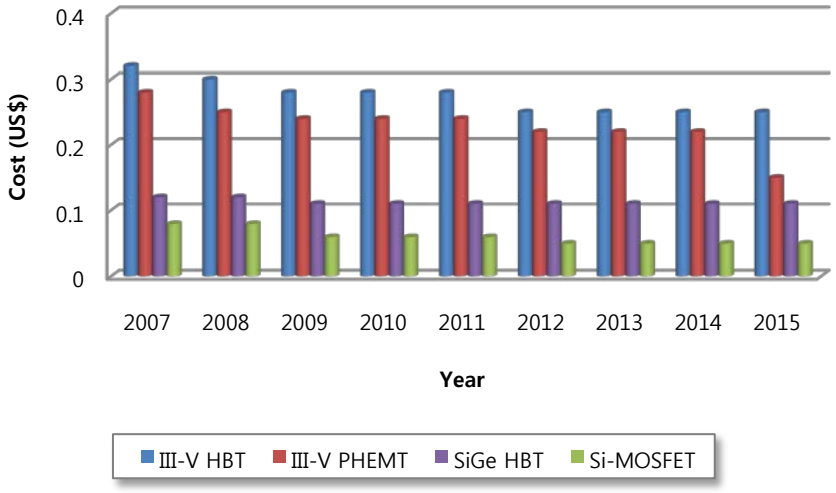


Figure 4. Cost advantages of CMOS technologies over other semiconductor technologies for PA solutions in US$/mm^2

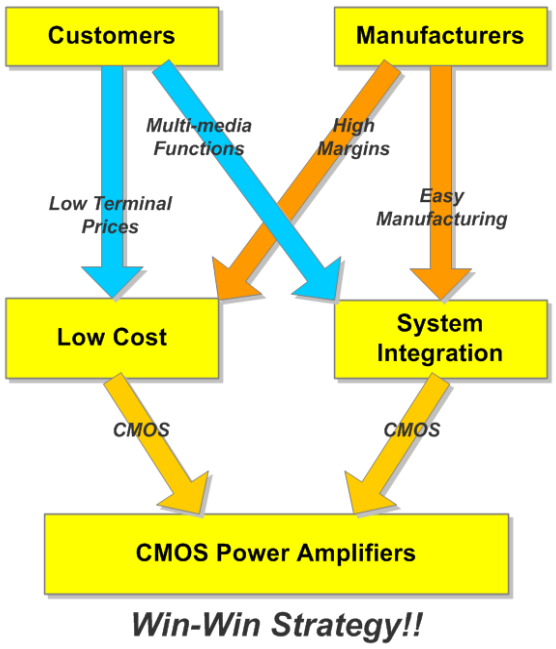


Figure 5. Benefits of CMOS PAs in cellular markets

Currently, the atmosphere is ideal for an RF CMOS PA in the wireless market. One caveat, however, is how to achieve comparable performance using CMOS in implementing a PA. Therefore, in serious consideration of the implementation of CMOS RF PAs, this research will introduce and discuss various efforts at determining good PA solutions for their commercial application in wireless communications.

1.3. Organization of the Thesis

Based on the aforementioned technological background and motivation, the purpose of this work is to exploit CMOS technologies for developing RF PAs for current and future wireless communications as illustrated in Figure 6. In this work, research on a power-combining transformer technique is proposed for a fully-integrated switching PA

that achieves both high output power and high efficiency for constant envelope communication. In addition, the analysis for power back-off efficiency is executed for the design of a fully-integrated linear PA that achieves both high output power and high efficiency in linear operation for non-constant envelope high data rate communication. The final and ultimate goal of this research is to identify the critical characteristics of an RF PA for high data rate communications: high output power, high efficiency, and high linearity.

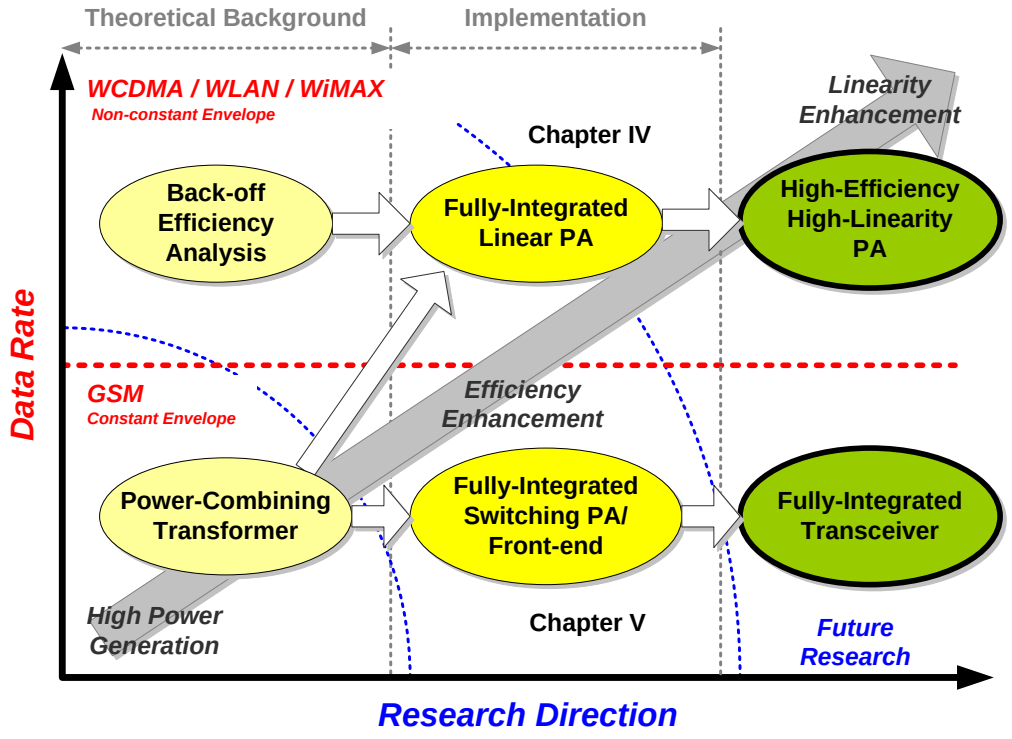


Figure 6. Outline of this research

Chapter 1 contains an introduction to the wireless market and current trends, the requirements of RF PAs, and the motivation for this work. To provide some background,

Chapter 2 presents an explanation of the basic definitions of RF PAs for wireless communications and describes the key quantities, wireless standards, and measurement methods. Chapter 3 presents CMOS technology and its shortcomings from an RF PA design standpoint, and the information presented in this chapter serves as a basis for several PA designs discussed and illustrated in the following chapters. Chapter 4 presents a new power-combining method using a monolithic transformer for PA and the design of class-E PAs and an RF front-end design, and Chapter 5 introduces a CMOS linear PA for high data rate communications with an analysis of power back-off efficiency. Finally, Chapter 6 summarizes and concludes the work in this dissertation, and posits research trends for the future.

CHAPTER 2

RF POWER AMPLIFIERS FOR WIRELESS COMMUNICATIONS

2.1. Introduction

Before advancing to the topic of CMOS RF PA, this dissertation will present an overview of PAs, which will serve as a guideline for the remainder of this research. A prerequisite for the design of PAs is a thorough understanding of the meaning and the significance of their key characteristics, such as output power, gain, efficiency, linearity, harmonic, stability, and so on. However, the complicated nature of these characteristics, particularly linearity, often hampers designers in their efforts to realize an efficient linear PA. Thus, the quantitative measures of linearity must first be understood. More importantly, for high data rate digital communications, such quantities, used to represent popular digital standards such as GSM, WCDMA, WLAN, and WiMAX must be known from the first phase of PA design.

Section 2.2 introduces the key characteristics of RF PAs such as output power, gain, efficiency, linearity, and other specifications, and Section 2.3 briefly summarizes several popular wireless standards from the PA standpoint. Section 2.4 then lists the general design procedures, and Section 2.5 lists the general measurement setups for various PA specifications.

2.2. Characteristics of RF Power Amplifiers

The characteristics of RF PAs differ according to various standards. Once a PA is given, however, common specification parameters are used to evaluate its performance. The key specification items and their meanings follow.

2.2.1. Output Power, Gain, and Efficiency

2.2.1.1. Output Power

Output power is the most important design aspect of a PA. In one sense, if the PA generates low output power, it loses its identity, making it hard to define. When a supply voltage is given as a fixed value, only the amount of current that provides a required output power can be a design parameter. Assuming a normal output load with resistance, R, the PA in Figure 7 has an output power of the following expression:

$$P_{OUT} = \frac{\left(\frac{V_{\text{peak-to-peak}}}{2}\right)^2}{2 \cdot R} \tag{2.1}$$

.

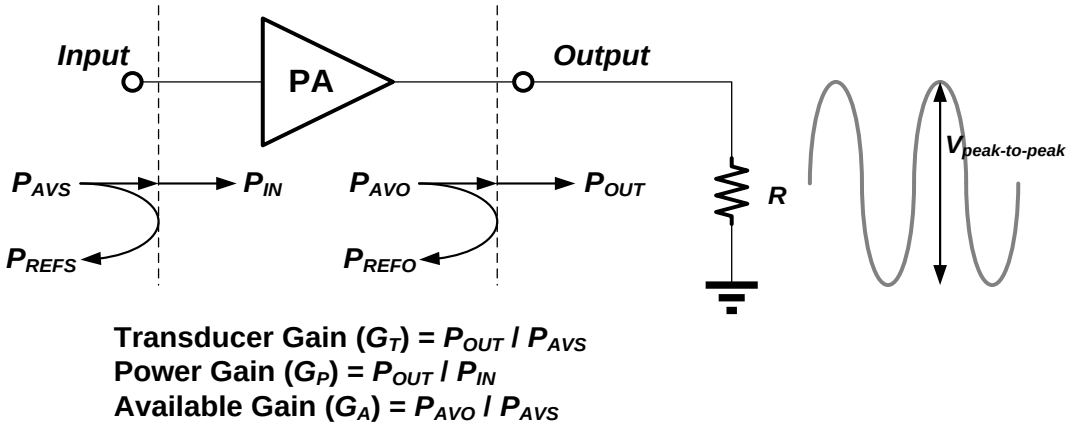


Figure 7. Definition of power and gain

In RF applications, the power level, usually defined as dBm, has a decibel value on a reference of 0.001 Watt (0 dBm). Assuming a general RF block with 50-Ohm terminations for the input and output, the voltage level can be derived. For a 30 dBm PA with 0 dBm input power (a gain of 30 dB), the peak-to-peak voltage swing for the input and output are 0.632 V and 20 V, respectively. Since an air-interface is not easily defined as fixed impedance, using a power interpretation instead of a voltage interpretation is preferable. Moreover, a link budget for a communication should be defined as a unit of power for the calculation of the dynamic range [2, 6].

2.2.1.2. Gain

The gain of the PA in Figure 7 can be defined as follows. Using the definitions given in the figure,

$$\text{Transducer Gain } (G_T) = \frac{P_{OUT}}{P_{AVS}}, \tag{2.2}$$

$$\text{Power Gain } (G_P) = \frac{P_{OUT}}{P_{IN}} = G_T \left(1 + \frac{P_{REFS}}{P_{IN}}\right), \tag{2.3}$$

$$\text{Available Gain } (G_A) = \frac{P_{AVO}}{P_{AVS}} = G_T \left(1 + \frac{P_{REFO}}{P_{OUT}}\right). \tag{2.4}$$

In the gain definitions of a PA, the transducer gain of Equation (2.2) is handily used for general measurements. The power gain of Equation (2.3) is the gain considering the input and output matching conditions. This definition is useful when the matching condition is not well optimized and the reflection at the input and load are not negligible, which is often observed in source/load-pull tests. If the reflection at the input of the PA is eliminated, the definition is the same with the transducer gain. Finally, the available gain

of Equation (2.4) is useful for the estimation of the maximum performance assuming perfect matching conditions for the input and the load of the PA. In reality, however, this gain is not feasible due to the unwanted mismatch in implementation. By ignoring the reflection at the output of the PA, the gain can be interpreted as the transducer gain, shown in Equation (2.4).

2.2.1.3. Efficiency

To generate an output power, we need to supply energy that is higher than the required output power in advance. Since running a PA requires high current consumption, any careless control of the PA may cause power dissipation in the form of heat. While this is not a critical issue for fixed terminals, for mobile terminals with limited energy supplied by battery, the savings in current consumption would be critical for longer battery life and mobility. Thus, the efficiency of PAs is crucial to wireless applications. Even for fixed applications such as baseband station PAs, if the efficiency is too low, the heat generation by low efficiency may cause a problem with reliability.

For a typical PA, as shown in Figure 8, efficiency can be defined as the ratio between the output power and the sum of all supplied energy into the black box, including the input power and the DC supplied power. The most popular and accurate definition of an efficiency, referred to as power-added efficiency (PAE) is represented in Equation (2.5).

$$PAE = \frac{P_{OUT} - P_{IN}}{P_{DC}} = \left(1 - \frac{1}{G}\right) \cdot \frac{P_{OUT}}{V_{DD} \cdot (I_{PA} + I_{DA} + I_{BIAS})} \tag{2.5}$$

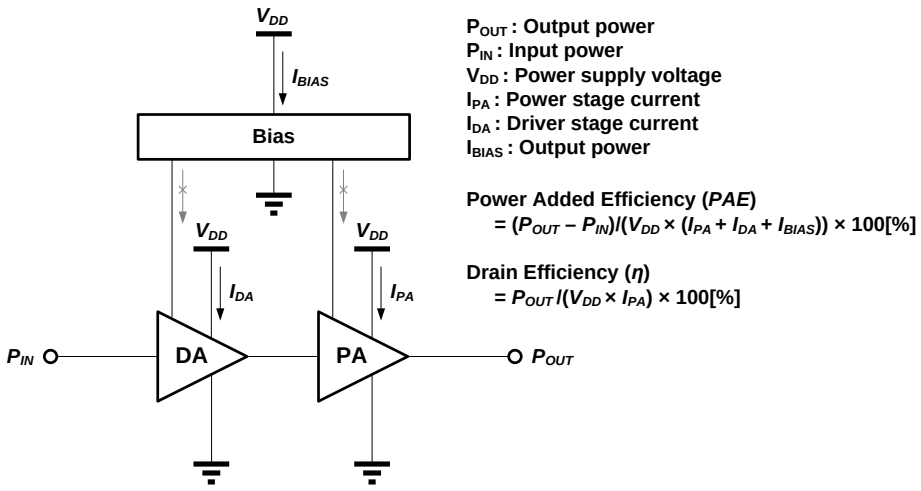


Figure 8. Efficiency calculation of a PA

Other definitions have been used for the same PA: in the case of field effect transistor (FET) circuits, it refers to drain efficiency (DE); and in the case of bipolar junction transistor (BJT) circuits, it refers to collector efficiency (CE). In this work, which assumes that all designs are FET circuits, only the term DE is used.

$$\eta(DE) = \frac{P_{OUT}}{P_{DC}} = \frac{P_{OUT}}{P_{DC} \cdot I_{PA}} \tag{2.6}$$

However, confusion may arise from this definition when dealing with a multi-stage PA in which more than one driver stage is used to drive the final stage. In some cases, DE is simply PAE without input power, but it includes driver current consumption. However, it is more intuitive to define DE as Equation (2.6), which includes only the drain bias current consumption. For a complete PA, the usage of DE can be misleading, because no information of gain is provided. If the gain is low, additional stages, which consume additional power, are needed to drive the final stage. Thus, this definition may be useful

only for defining the quality of matching conditions for the output, but not for representing power consumption used for generating output power.

2.2.2. Linearity

Linearity of a PA represents a criterion that represents how the quality of a given signal is maintained throughout the PA. However, myriad definitions and abbreviations for linearity specifications may confuse newcomers to this field who do not know which ones to use for characterization of linearity or how to interpret them. Specifically, the definition of linearity varies depending on viewpoints and modulation schemes.

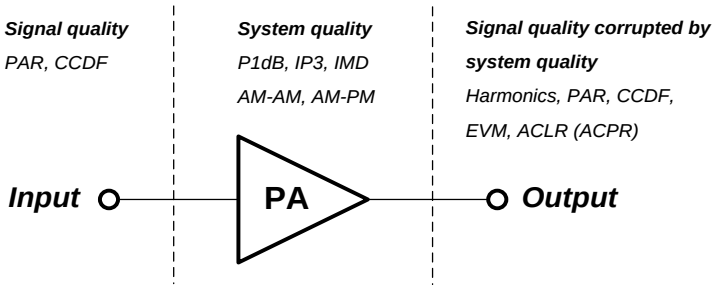


Figure 9. Linearity indicators of a PA

In Figure 9, the quality of a signal can be defined as a peak-to-average ratio (PAR) or a complimentary cumulative density function (CCDF). Moreover, the linearity of a system can be defined as a 1dB compression point (P1dB), a third-order intercept point (IP3), amplitude-amplitude modulation (AM-AM), amplitude-phase modulation (AM-PM), or third-order intermodulation distortion (IMD3). Such system qualities affect the quality of a signal, and the signal quality corrupted by the system quality can be defined by harmonics, PAR, CCDF, and the adjacent channel power ratio (ACPR), i.e., the

adjacent channel leakage ratio (ACLR), error vector magnitude (EVM), and so on. Although the system qualities are not specified by communication standards, they are specifically defined in standards such that linear PAs should keep signals within a specified limit. The purpose of maintaining linearity is two-fold: to minimize signal distortion (PAR, CCDF, P1dB, AM-AM, AM-PM, EVM) so that users maintain good connectivity and to ensure co-existence with neighboring channels (harmonics, IP3, IMD, ACLR) so that other users can also maintain good connectivity.

2.2.2.1. Harmonics, P1dB, IP3, and IMD

Equation (2.7) represents a polynomial expansion of the general PA model in Figure 10 truncated at the third-order term.

$$y(t) \approx \alpha_1 x(t) + \alpha_2 x(t)^2 + \alpha_3 x(t)^3 \tag{2.7}$$

After input $x(t) = A\cos\omega t$ is applied, then the system generates an output,

$$\begin{aligned} y(t) &= \alpha_1 A\cos\omega t + \alpha_2 (A\cos\omega t)^2 + \alpha_3 (A\cos\omega t)^3 \\ &= \frac{1}{2}\alpha_2 A^2 + \left(\alpha_1 A + \frac{3}{4}\alpha_3 A^3\right)\cos\omega t + \frac{1}{2}\alpha_2 A^2 \cos 2\omega t + \frac{1}{4}\alpha_3 A^3 \cos 3\omega t \end{aligned} \tag{2.8}$$

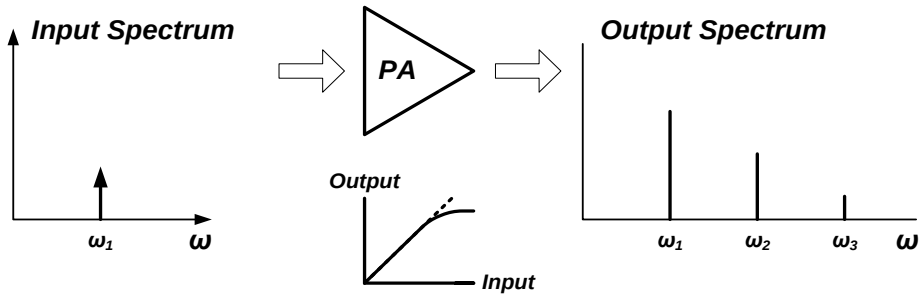


Figure 10. Distortion of a PA with one-tone input

The system generates higher-order harmonic components such as 2ω and 3ω as well as the fundamental frequency component at ω. While the cause of harmonic generation is the nonlinearity of a PA, this specification is usually dealt with independently because it can be suppressed by filtering characteristics at the output. Thus, if we simply measure the harmonic levels, which designers have more freedom to control, at the output port [6], it can differ from the actual nonlinear performance of the PA.

In the system, coefficient α_3 is assumed to be less than zero, or the system output expands with increased input to the system, violating a natural system in which new energy cannot be generated. Thus, with the increased input, the system suffers a compressive output. The point at which the original gain is compressed by 1 dB is defined as P1dB, illustrated in Figure 11, indicating a border between a reasonably linear region and a compression area.

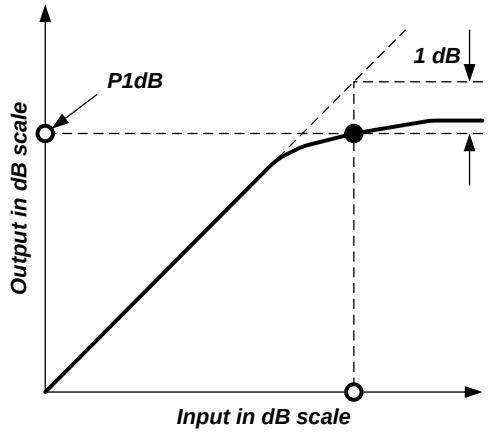


Figure 11. 1dB compression point

When two inputs with equal amplitudes are applied, as shown in Figure 12, $x(t) = A\cos\omega_1 t + A\cos\omega_2 t$, a different distortion mechanism, works through the same system,

such that not only harmonics but also inter-modulated signals appear very near the input frequency components at frequencies, $2\omega_1 - \omega_2$, $2\omega_2 - \omega_1$, $3\omega_1 - 2\omega_2$, $3\omega_2 - 2\omega_1$, and so on.

$$y(t) = \left(\alpha_1 A + \frac{9}{4}\alpha_3 A^3\right)\cos\omega_1 t + \left(\alpha_1 A + \frac{9}{4}\alpha_3 A^3\right)\cos\omega_2 t$$
$$+\frac{3}{4}\alpha_3 A^3\cos(2\omega_1 - \omega_2)\,t + \frac{3}{4}\alpha_3 A^3\cos(2\omega_2 - \omega_1)\,t + \cdots \qquad (2.9)$$

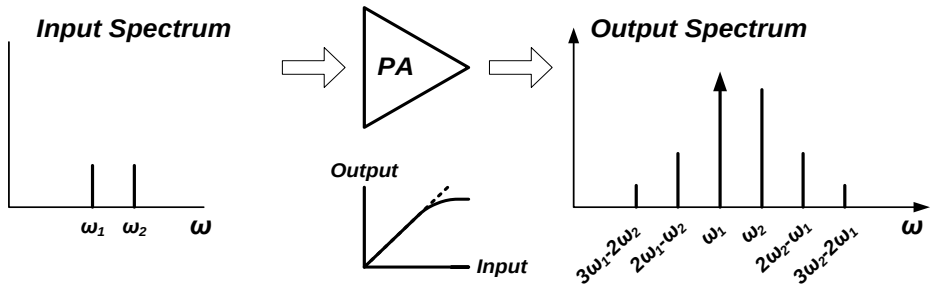


Figure 12. Distortion of a PA with two-tone input

The generated third-order nonlinearities at frequencies $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ in Equation (2.9) are called IMD3. In Figure 12, as the input level increases, the intermodulation terms increase as well but follow a steeper slope (three times in IMD3 and in general, n times of the fundamental slope for IMDn, n = 2, 3, 4, …), increases the fundamental level. At the imaginary intercept point at which the fundamental tone and the IMDn tone are equal, the n-th order intercept point (IPn, n = 2, 3, 4, …) is defined in the black dot in the figure. The input and output of the point are called input n-th order intercept point (IIPn, n = 2, 3, 4, …) and output n-th order intercept point (OIPn, n = 2, 3, 4, …), respectively. Higher-order nonlinearities such as fifth-order intermodulation (IMD5) can also be generated at frequencies $3\omega_1 - 2\omega_2$ and $3\omega_2 - 2\omega_1$, but the most

dominant intermodulation is still the third-order nonlinearities, so IMD3 and IP3 are usually used to indicate the linearity of a PA.

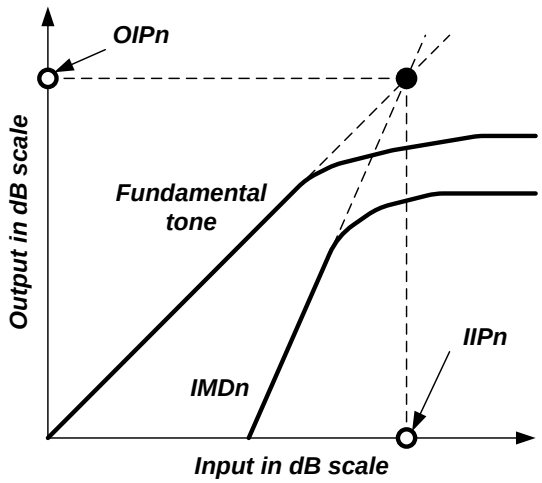


Figure 13. IP3 and IMD

2.2.2.2. AM-AM and AM-PM

AM-AM and AM-PM distortions represent the input-output relations of a PA excited by a sinusoidal input, depicted in Figure 14. AM-AM is distortion by the amplitude so that it has a strong connection with P1dB. However, this definition covers not only the compression by the P1dB but also the fluctuation of gain throughout the entire power range. In much the same way, the phase variation throughout the entire operational power range is quantified by AM-PM distortion. The quantities acquired by AM-AM and AM-PM are useful for characterizing EVM, in which distortions caused by the amplitude and the phase hinder the identification of the right constellation points. However, they are not very critical to the estimation of ACLR characteristics to which distinct frequency components generated by PA distortion are more relevant.

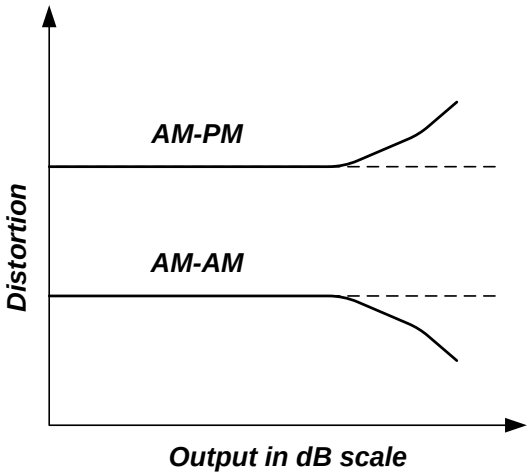


Figure 14. AM-AM and AM-PM distortion

2.2.2.3. PAR and CCDF

From the point of signal quality in a two-tone input test, the input can be rephrased as follows:

$$x(t) = A\cos\omega_1 t + A\cos\omega_2 t = 2A\cdot\cos\left(\frac{\omega_1+\omega_2}{2}t\right)\cdot\cos\left(\frac{\omega_1-\omega_2}{2}t\right), \quad (2.9)$$

$$P_{PEAK} = \left(\frac{2A}{\sqrt{2}}\right)^2 = 2A^2, \quad (2.10)$$

$$P_{AVG} = 2\left(\frac{A}{\sqrt{2}}\right)^2 = A^2, \quad (2.11)$$

$$P_{TONE} = \left(\frac{A}{\sqrt{2}}\right)^2 = \frac{A^2}{2}. \quad (2.12)$$

Thus, the ratio between the peak input signal over the average signal can be defined as PAR.

$$\text{PAR} \equiv 10\log\left(\frac{P_{PEAK}}{P_{AVG}}\right) = 3\ \text{dB} \quad (2.13)$$

For a linear operation without distortion, the input signal into the PA should be limited to a signal excursion equivalent to PAR at the output. Therefore, to guarantee that no distortion occurs, each tone of the two-tone test should be defined at the power back-off of 6 dB from the peak output power [6].

$$\text{Back-off} \equiv 10 \log\left(\frac{P_{PEAK}}{P_{TONE}}\right) = 6 \text{ dB} \tag{2.14}$$

Digital modulations usually have statistical signal distribution showing dynamic behaviors according to modulation that differs from static linearity [7]. While either the one- or two-tone signal characterization of linearity can provide an intuitive understanding of linearity simplifying the relationship between the input and output, an accurate characterization of linearity can be attained only by including the dynamics of a modulated signal. Therefore, the definition of PAR for digital modulations cannot be derived from simple derivations, and only statistical accumulation of data can provide these definitions.

CCDF indicates that the PAR of a digital communication reflects the statistical distribution of signal amplitude. As represented in Equation (2.15), the probability density function (PDF) of a signal is determined by the low power level of the signal, resulting in a cumulative density function (CDF), which is depicted in a complimentary fashion that emphasizes the peak amplitude area. As an illustration, Figure 15 shows the CCDF of a digitally-modulated signal that has the PAR of 3.2 dB.

$$CCDF \equiv 1 - \int PDF(P) \cdot dP \tag{2.15}$$

where *P* represents the instantaneous power of a PA. Depending on modulation format, CCDF curves vary for different communication standards. A signal with a high CCDF near peak output range suffers more from distortion.

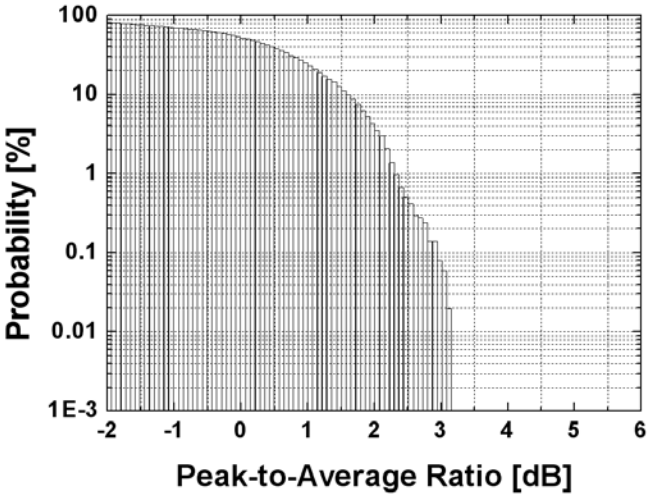


Figure 15. CCDF of a digitally modulated signal

Although PAR and CCDF determine the quality of a general signal, the signal can be distorted through a PA, which causes the PA to undergo amplitude compression; then the PAR and CCDF values may decrease. Thus, PA designers should maintain a sufficient signal excursion margin to minimize any reduction in these two linearity indicators.

2.2.2.4. EVM and ACLR

EVM is a metric of modulation or demodulation accuracy in a transmitting chain. Digital modulation requires a constellation diagram to identify each data point. Figure 16 shows that the ideal symbol location is often displaced by amplitude and phase error through a transmitting chain, and the measured symbol location is found in a different

location by an amount of an error vector. The following calculated EVM Equations (2.16) and (2.17) can be formulation on either in a dB scale or a percent scale.

$$\mathrm{EVM(dB)} \equiv 10\log\left(\frac{\text{Error vector}}{\text{Reference vector}}\right) \tag{2.16}$$

$$\mathrm{EVM(\%)} \equiv \sqrt{\frac{\text{Error vector}}{\text{Reference vector}}} \times 100(\%) \tag{2.17}$$

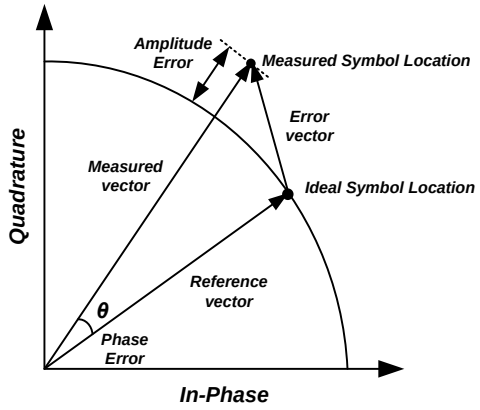


Figure 16. Error vector of symbols

Figure 17 illustrates a constellation diagram in which a digitally-modulated signal is plotted. The white circles in the figure represent a data point with errors, and the lines between the white circles represent the movement of the amplitude information. As the group of white circles expands, the percentage vector displacement represented by EVM also degrades, increasing the likelihood that it will be incorrectly interpreted as a different data point.

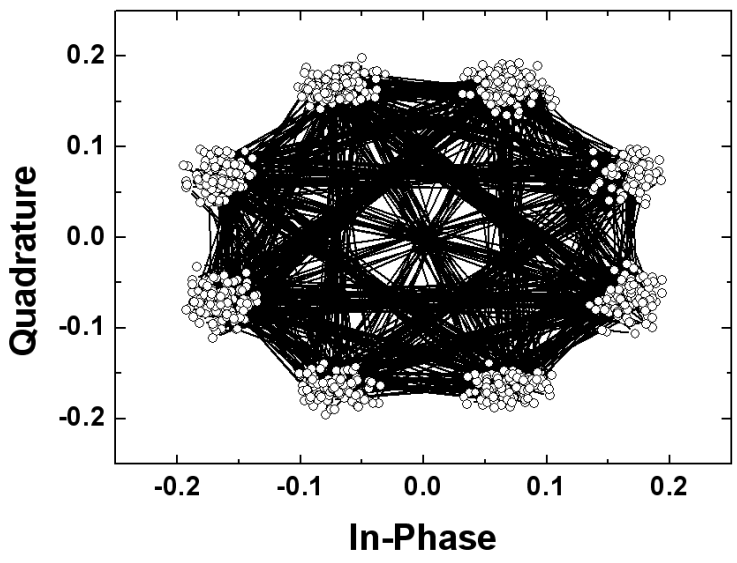


Figure 17. EVM of a digitally-modulated signal

Another metric for linearity in digital communication is ACLR, which follows:

$$ACLR\ (\text{dBc}) \equiv 10 \log\left(\frac{\text{Power in Adjacent or Alternative Channel in Watt}}{\text{Power in Main Channel in Watt}}\right) \tag{2.18}$$

Intermodulation by the transmitter odd-order nonlinearities widens the signal spectrum, shown in Figure 18. A general term for this phenomenon is "spectral regrowth." The power of spectral regrowth in the adjacent channel acts as interference for other users in the cell using this channel.

Channel power adjacent to the main channel power is referred to as the "adjacent channel leakage ratio" while the channel power neighboring the adjacent channel is referred to as the "alternative channel leakage ratio" for the same abbreviation—ACLR.

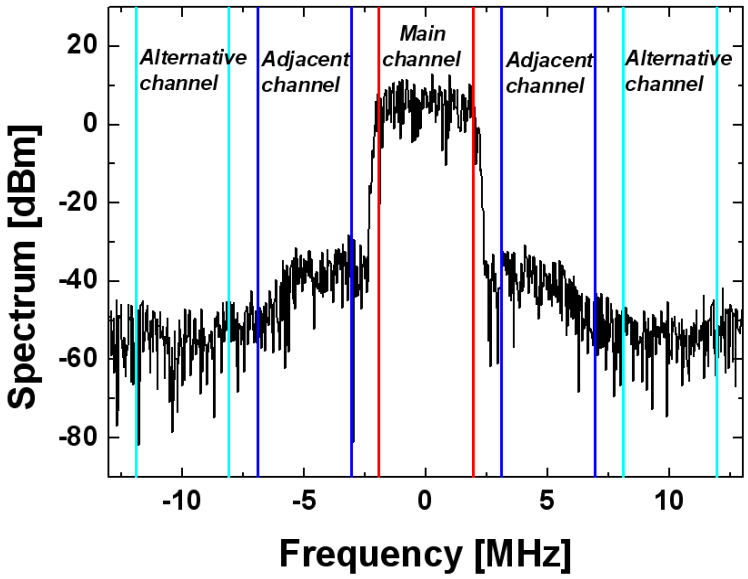


Figure 18. ACLR of a digitally-modulated signal

While the definitions of EVM and ACLR are generally described in this section, the exact values are defined only by each standard characteristic. Since the two metrics are based on different nonlinearity mechanisms, the exact relationship between them cannot easily be identified.

2.2.3. Other Characteristics

2.2.3.1. Stability and Ruggedness

The usage environment of a general mobile terminal is very unpredictable due to the electromagnetic absorption or reflection of human bodies and other structures near the antenna of the mobile terminals. Thus, the usable range in such an environment, or mismatch, should be defined. VSWR is a measure of mismatch from the point of voltage reflection. In general cases, for low VSWR (e.g., less than 6:1), it is defined as the

guarantee of stability such that the PA is within a normal operating range. For high VSWR (e.g., about 10:1), it is defined as ruggedness that guarantees no damage to the PA, even in a high reflection environment.

In Figure 19, the general reflection coefficient and the voltage standing wave ratio (VSWR) of a load can be defined as Equations (2.19) and (2.20), respectively.

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{2.19}$$

$$VSWR = \frac{1 + |\Gamma|}{1 - |\Gamma|} \tag{2.20}$$

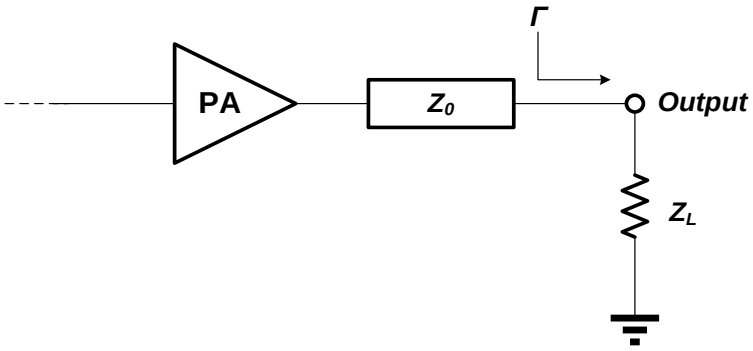


Figure 19. Reflection at output node.

By combining Equations (2.21) and (2.22), the load with a constant reflection coefficient can be summarized as Equation (2.23):

$$|\Gamma| = \frac{VSWR - 1}{VSWR + 1}, \tag{2.21}$$

$$\Gamma = \left|\frac{VSWR - 1}{VSWR + 1}\right| e^{j\theta}, \tag{2.22}$$

$$Z_L = \left[\frac{1 + \frac{VSWR - 1}{VSWR + 1} e^{j\varphi}}{1 - \frac{VSWR - 1}{VSWR + 1} e^{j\varphi}}\right] \cdot Z_0, \tag{2.23}$$

where θ represents the phase of Γ, and φ is the angle of VSWR. With φ in a range of 0 to 2π radians, constant VSWR circles for some VSWR values can be plotted, depicted in Figure 20. As can be seen in the figure, the perfect matching point can be found at VSWR = 1:1, and the VSWR = 3:1 circle indicates the area in which half of the incident power is reflected at the load. The VSWR circle close to 10:1 is already close to the outer border of the Smith chart, indicating that most of the power will be reflected at the load.

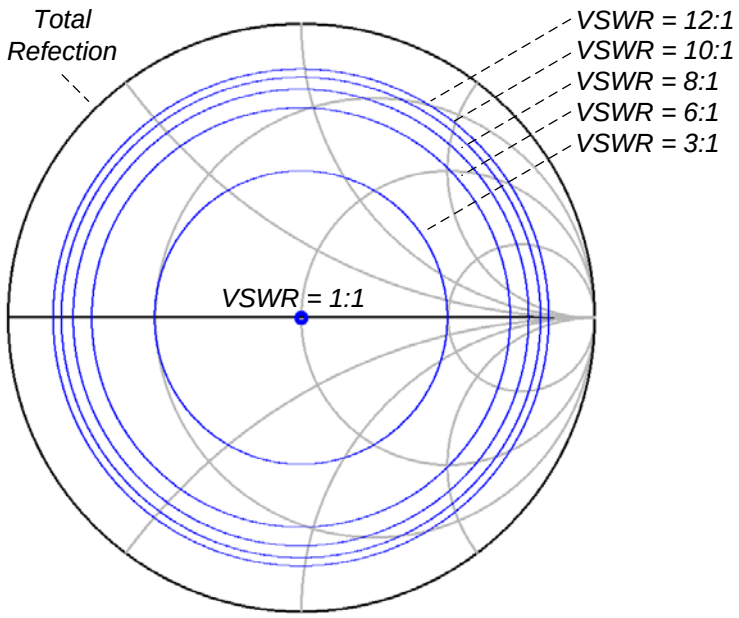


Figure 20. Constant VSWR circles in the Smith chart

For a varying load, PAs should be stable; that is, they should generate neither spurious signals nor oscillation throughout the whole frequency band of interest. Stability is measured by the Rollet stability factor, can be defined in the following Equation (2.24) [8], [9] for any two-terminal device:

$$K \equiv \frac{1 + |\Delta|^2 - |S_{11}|^2 - |S_{22}|^2}{2|S_{21}||S_{12}|} \tag{2.24}$$

where $\Delta = S_{11}S_{22} - S_{12}S_{21}$, and $K > 1$ and $\Delta > 0$ should be satisfied for unconditional stability.

2.2.3.2. Reverse Intermodulation Product

When a transmitting signal and an interfering signal from a co-located transmitter, which conducts back from an air-interface, are intermodulated at a PA output, a new distortion, shown in Figure 21, is generated. This IMD product, unlike the IMD product caused by the PA distortion, is called the "reverse intermodulation product." By adding an isolator after the PA, this IMD distortion can be suppressed. The measurement method for this phenomenon is described in detail in Section 2.5 of this chapter.

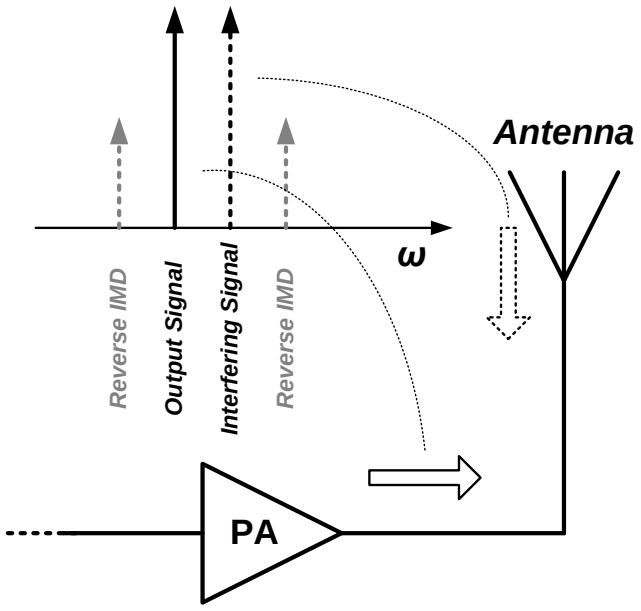


Figure 21. Reverse IMD generation

2.2.3.3. Idle Current

The quantity of an idle current is defined for linear applications. In case of a time-division duplexing (TDD) system, the transmitter of the system goes into an off state consuming no current. However, in a frequency-division duplexing (FDD) system, in which the transmitter should be ready to send a signal at all times, the PA in the system draws an idle current. Often this quantity is also called a "quiescent current." Since a quiescent current involves power dissipation that does not entail sending information, such unnecessary power dissipation should be suppressed as much as possible. If the idle state is statistically very long compared to the communication time, most of the battery power is simply wasted while the transmitter is waiting for usage.

$$I_{AVG} = T_{ON} \cdot I_{ON} + (1 - T_{ON}) \cdot I_{IDLE} \tag{2.25}$$

$$P_{AVG} = V_{DD} \cdot I_{AVG} \tag{2.26}$$

where

I_{AVG} = Average current drain.

T_{ON} = Fraction of time the PA is on.

I_{ON} = Current drain from the battery when the PA is on.

I_{IDLE} = Current drain from the battery when the PA is on standby.

P_{AVG} = Average power consumption.

2.3. Wireless Standards

Among the many wireless standards, the most popular ones in the market that describe the specifications related to the design of PAs are introduced in this section. For 2G communications, the global system for mobile communications, used mostly for

voice communication is also described; and for 3G mobile communications, wideband code division multiple access, and specifically for dedicated data communications in 3G communications, WLAN and WiMAX standards are briefly introduced.

2.3.1. Global System for Mobile Communications

Before GMS appeared in the market in the early 1990s, analog cellular communications such as advanced mobile phone service (AMPS) or Nordic mobile telephone (NMT) service, both introduced around 1980, were available. Evolving throughout the decades, GSM highlighted the need for and popularity of digital communication. Despite the advancement of new standards, GSM is currently the most popular cellular communication standard worldwide. The GSM standard is based on Gaussian minimum shift keying (GMSK) modulation in which time-division multiple access (TDMA) is used for user capacity in which one frame must be divided into eight slots. For a higher data rate, multiple slots can be assigned to one user, referred to as the "general packet radio system" (GPRS). By distinguishing the transmitting and receiving frequency bands, it is categorized as FDD, but instead of a duplexer at the antenna port, an RF switch can be used, owing to the nature of TDMA.

The design of a GSM PA does not demand a linearity requirement, so only output power and PAE are key design specifications to meet. High efficiency is the most important design target for this standard. The key specifications of GSM PAs are summarized in Table 2. Since the PA is turned on and off for each time slot, the PA should meet not only a spectral mask but also a time domain mask. The control of GSM PA is by a control port, not by input power requiring a good analog bias circuitry.

Table 2. Commercial GSM PA specifications

Specification	Value
Operating frequency	824-849 MHz 880-915 MHz 1710-1780 MHz 1850-1910 MHz
Transmission rate	270.833 kbps
Maximum output power	+33 dBm-+35 dBm
PAE @ Maximum output power	~50%

2.3.2. Wideband Code Division Multiple Access

Wideband CDMA is a standard for "3GPP" (Third Generation Partnership project), the so-called "UMTS" (Universal Mobile Telecommunications System), targeting high-speed mobile data communications over simple voice communications. A typical data rate is 3.84 Mbps, but by reducing a spreading factor in a high-speed download packet access (HSDPA), a higher data rate up to 14 Mbps is also possible. In the case of WCDMA PAs, the required specifications for class 3 (24 dBm output power at antenna port) are listed in Table 3. The usual products in the market deliver an output power of 28 dBm, a PAE of 40% and an ACLR of -40 dBc at 5 MHz offset [10], [11] while the standard specification is -33 dBc at this offset. ACLR at 10 MHz offset should be less than -43 dBc.

Using the minimum transmit power of -50dBm required for all power classes, it can say that a class 1 mobile station has a transmit power dynamic range of -50 dBm to +33 dBm, producing a total dynamic range of 83dB [12].

Table 3. Commercial WCDMA PA specifications

Specification	Value
Operating frequency	1.92-1.98 GHz
Bandwidth	5 MHz
Chip rate	3.84 Mcps
Maximum output power (class 3)	+27 dBm - +28.5 dBm
Dynamic range (class 3)	78 dB (-50 dBm - +28 dBm)
PAE @ Maximum linear output power	~ 40%
ACLR (3.84 MHz integration)	< -40 dBc @ 5 MHz offset
	< -50 dBc @ 10 MHz offset
EVM	< 2.5%

2.3.3. Wireless Local Area Network and Worldwide Interoperability for Microwave Access

Ever-increasing demands for a high data rate have not been successfully satisfied by mobile standards such as GSM, GPRS, EDGE, CDMA, and WCDMA. While the mobility of the standards is lower than those listed as mobile standards, an increased data rate could successfully be realized by WLAN and WiMAX due to their orthogonal frequency division multiplexing (OFDM) based on IEEE standards 802.11 and 802.16, respectively. In this multiplexing scheme, a set of carriers located individually and independently in close proximity allows frequency diversity.

Table 4 lists some commercial specifications of IEEE 802.11g.

So far, an IEEE 802.11 WLAN is considered the most suitable application for the CMOS PA. Since this application is operated at a comparably low power level with a low voltage swing, the burden of reliability and ruggedness is significantly reduced. In addition, the time division duplexing (TDD) mode of the 802.11 WLAN helps integration

with the transceiver since a TDD-based system does not concurrently operate both the receiver and transmitter parts. This confines the substrate coupling problem to the transmitter only [13].

Table 4. 802.11g WLAN transmitter specifications

Specifications	Value
Frequency band	2.4-2.4835 GHz
Number of carriers	52 (48 data and 4 pilots)
Channel bandwidth	16.25 MHz
Data rate	6 to 54 Mbps
Carrier type	OFDM
Modulation	BPSK, QPSK, 16QAM or 64QAM
Max. instantaneous output power	1 W (in USA)
EVM	2-3% or -25dB for 54 Mbps
Spectrum mask	-20 dBc @ 11MHz offset -28 dBc @ 20MHz offset -40 dBc @ 30MHz offset

2.4. Design of RF Power Amplifiers

2.4.1. Design Procedure

In the design of a PA with given specifications, a systematic approach can be helpful, so this research proposes the following design procedure. The first step of the procedure is to evaluate the transistor in light of the available semiconductor solutions and then to check the transistor characteristics from the perspective of the loadline, the determination of size, and finally the characteristics of a given size. The next step is to design a power stage in which both load-pull and source-pull analyses are repeated until the design

optimization for output power, efficiency, and sometimes even linearity are complete. The final step is to evaluate the spectrum performance of the power stage. The same procedure can be applied to the driver stage. By securing the designed power stage and the driver stage, a complete PA that focuses on the output matching, interstage matching, and input matching can be designed. Since the performance should be optimized, even though each block is already characterized well, all design parameters should be readdressed from the perspective of the full stage design. Since all specifications are in trade-off relations, designers are required to spend considerable time optimizing and repeating the same procedure until all the specifications have sufficient margins.

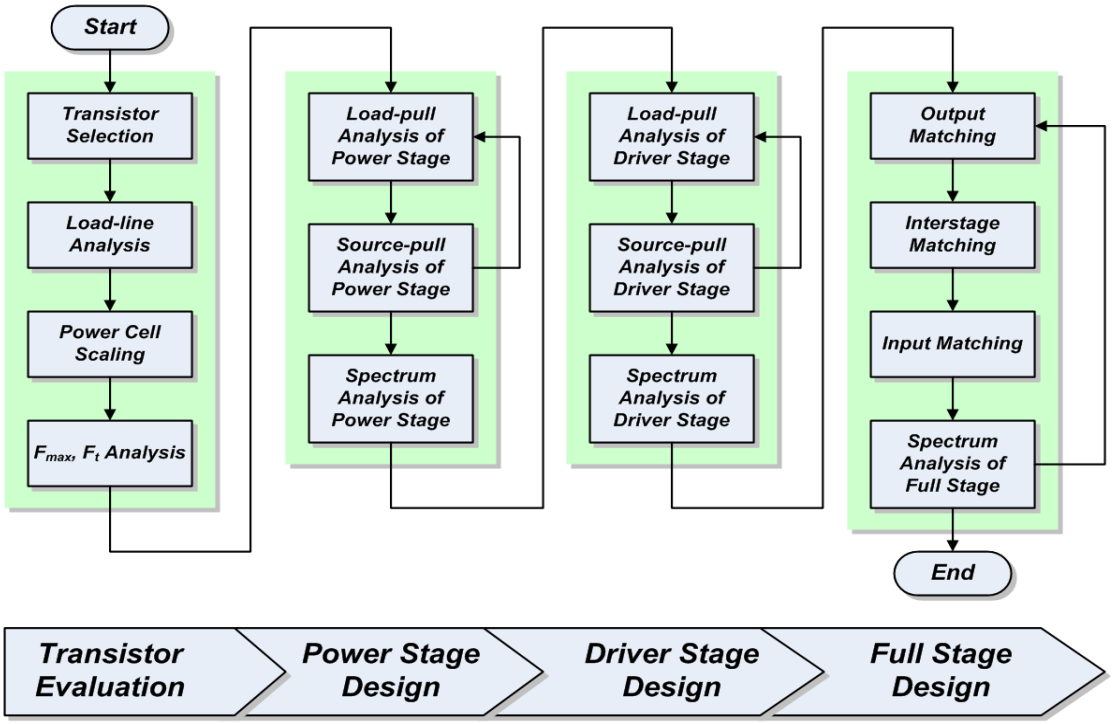


Figure 22. General design procedure of a PA

In the abstract, this suggested design procedure is simple and clear. However, the actual design procedure often encounters more obstacles than expected, so designers should be prepared for long, tedious optimization procedure. Often, designers fall in a quagmire of going back-and-forth among all the trade-off relations that have deviated from the measured steps. Furthermore, the design procedure can vary according to different class operations. For example, if a class-E PA is designed, the required output impedance can be intentionally manipulated without regard to the optimal load-pull result. Such general guidelines can be helpful but not always the best approach for designers to take.

2.4.2. Simulation Techniques

The characterization of RF PAs for digital communications requires computer-assisted tools for calculating complicated signals. While transient analysis works in the time domain, harmonic balance, or Volterra series analysis, works in the frequency domain. To include the advantages of both domains, envelope simulation runs in the time domain and at each time point, harmonic simulation in the frequency domain works [14].

2.4.2.1. Transient Analysis

Transient analysis is not easily applicable to digital domain signals, particularly for data-spreading spectrum techniques in which the PN code division of signals requires an impractically short time step. More importantly, the signal in the time domain should be converted to the frequency domain at a cost of the fast Fourier transformer (FFT) consuming computational time. Thus, this technique is rarely used in the design of a PA.

When the PA includes time-step input, the transient response of the PA can be checked for stability and settling time.

2.4.2.2. Harmonic Balance Analysis

While harmonic balance (HB) is considered a technique in the frequency domain, it is a hybrid between the time and frequency domains. While the latter deals with the linear part of a signal, the former deals with the nonlinear part. This technique is applicable to strongly nonlinear systems with time-invariant coefficients for a Fourier series, representing only periodic and quasi-periodic responses. By simplifying a digitally-modulated signal with a few tones, it is widely used for RF simulation, including very nonlinear voltage-controlled oscillators (VCOs) and PAs. Some simulation techniques such as large signal S-parameter (LSSP) simulation or envelope simulation are also based on the HB. Since HB also consumes considerable computation time and memory, the smart matrix manipulation technique such as the Krylov solver has been widely adopted as an option. The phase information of each tone in HB should match to that of the envelope for an accurate modeling of digital signals.

2.4.2.3. Volterra Series Analysis

Like HB, Volterra analysis also uses a multi-tone signal in a system expanded by the Volterra series, in which the memory effect of PAs is also included. It is good for a weakly nonlinear system below P1dB. Since most linear PAs are operated in the back-off region below P1dB, this technique is useful for a linear PA design.

2.4.2.4. Envelope Analysis

Standard HB cannot handle a circuit with digitally-modulated signals or a multiple time-scale signal. Instead, in envelope analysis, different time rates are used for the sampling of the circuit variables [15]. By using time-variant phasors, this analytical technique can efficiently simulate complex regimes at low computational cost. Digitally-modulated signals can be modeled by one carrier with a time-varying complex envelope. Following the time step, the circuit is analyzed using single-tone HB, which calculates the instantaneous envelope at the point. Then another HB is executed during the next time step with the interval decided by the envelope bandwidth [14].

2.5. Measurement of RF Power Amplifiers

The measurement setup of a PA requires a vector signal generator, a vector signal analyzer or a spectrum analyzer, a power meter, and several power suppliers. Attenuators for the input and output of the PA can help reduce reflection by mismatch. Since the allowable input power for the vector signal analyzer is usually less than 1 Watt, use of coupled output, shown in Figure 23, is recommended. While the loss is represented in dB (the power has a unit of dBm), it is often overlooked that the same 0.1 dB difference is actually a huge difference for peak output power and back-off output power. Therefore, the measurement setup must be calibrated with a particular focus on the output side, on which a slight calibration error could cause a significant percentage of efficiency miscalculation.

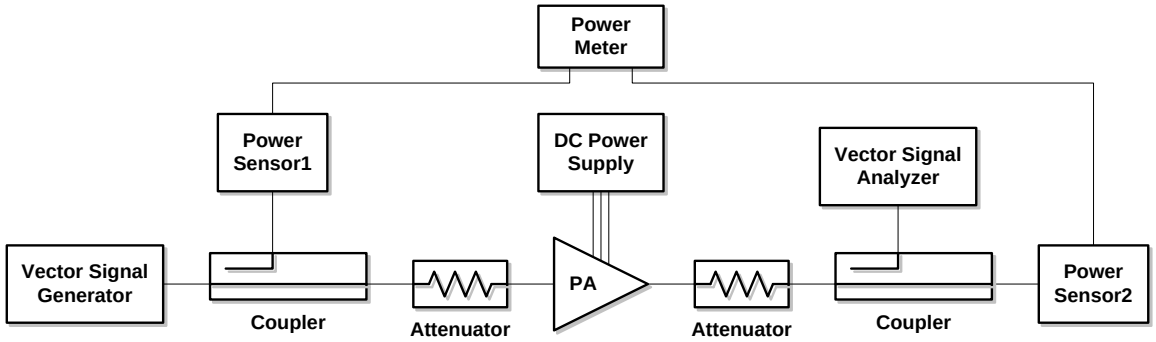


Figure 23. Measurement setup of a PA

To minimize the error of the output power measurement, we must assure a power calibration condition. The calibration process, in which a power meter must have at least two independent sensors and readers or two separate power meters, consists of several steps. The first is to calculate the input offset for the first power meter by reading the power difference for the input part of the measurement setup. The next is to operate the measurement setup without a PA using a through component and then to record the difference as the offset of the second power meter. Figure 24 shows the setups for the calculation of offsets at the input and output, respectively, based on Equations (2.28) and (2.29).

$$Offset_{PM1} = Power_{PM2} - Power_{PM1} \tag{2.28}$$

$$Offset_{PM2} = Power_{PM1} - Power_{PM2} \tag{2.29}$$

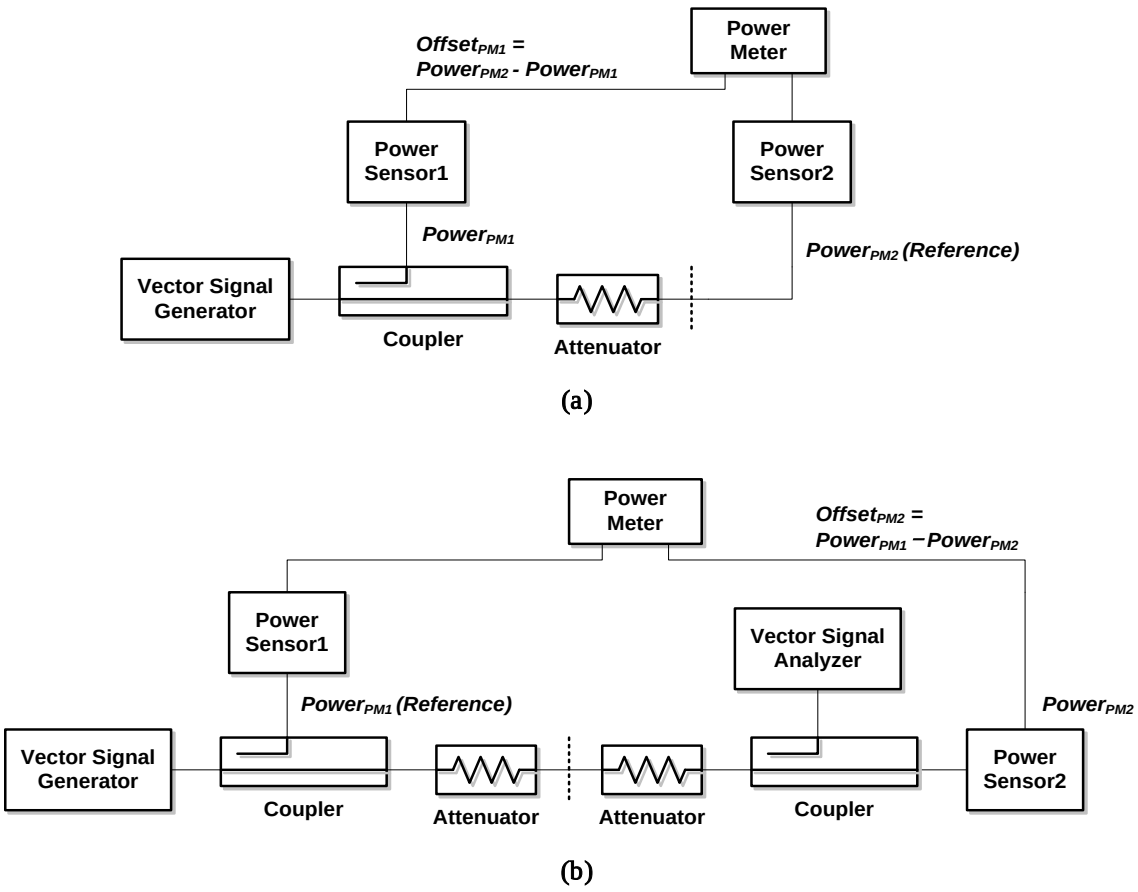


Figure 24. Calibration procedure for a PA measurement setup:
(a) offset calculation of input and (b) offset calculation of output

Figure 25 shows a typical source/load-pull setup for VSWR measurement. Under mismatched conditions, the operation of the PA is performed by controlling two tuners. A mismatch affects the load conditions of the matching network, causing the performance of the PA to deteriorate due to abnormal stress. Thus, in the evaluation process, such extreme cases should be verified as described in the previous section on VSWR specifications.

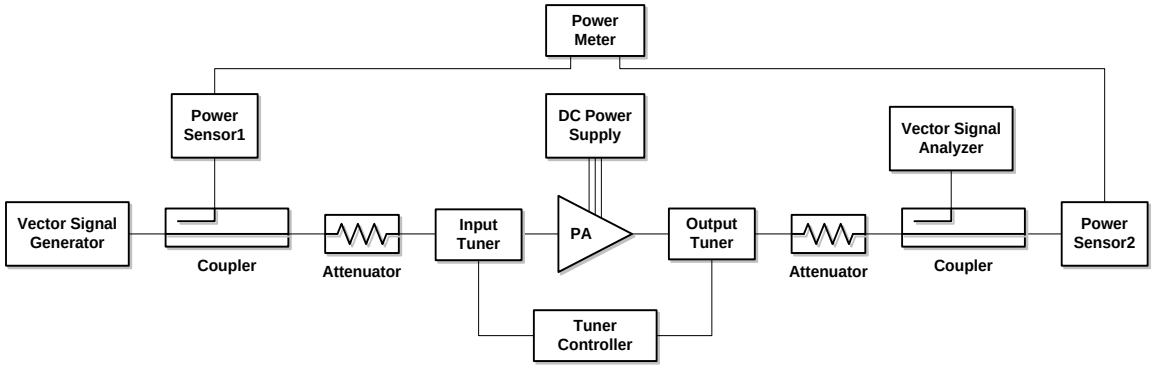


Figure 25. Source/load-pull setup

Figure 26 shows the setup for reverse IMD measurement. The figure shows that the two generators are used both at the input and output of the PA and that the IMD is measured at the output. The underlying mechanism of reverse IMD is briefly discussed in the previous section.

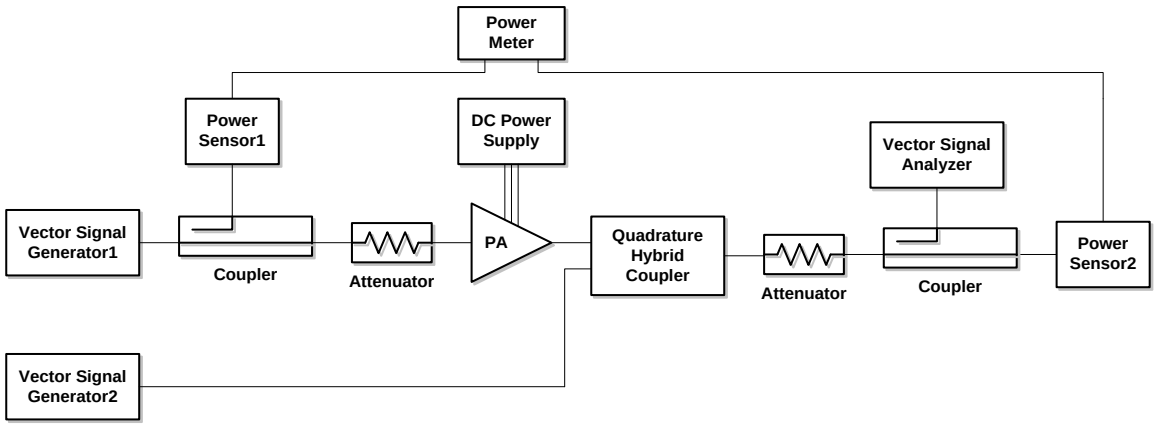


Figure 26. Reverse IMD test set up

2.6. Conclusion

This chapter provided general information that explains the design of RF PAs. The key quantities used to characterize a PA are output power, gain, efficiency, linearity, and so on. Each parameter has several different definitions according to a perspective within which the PA has been understood. In particular, the indicators of linearity can vary according to different standards and viewpoints, so the definitions of linearity such as P1dB, IP3, IMD, ACLR, and EVM can be selectively used for various digital standards. Once a standard is chosen as a target for PA design, proper design procedures and simulation techniques are used. For PA design, most design efforts go into the output matching of the last stage for maximum output and efficiency generation with high linearity. Estimating the correct behavior of PAs also entails the appropriate selection of a simulation tool from among the many time and frequency domain techniques. Accordingly, measurement setups for various PA specifications should also be properly understood and prepared to retrieve the accurate characteristics of a measured PA.

CHAPTER 3

CHALLENGES AND TECHNIQUES OF CMOS RF POWER AMPLIFIERS

3.1 Introduction

Designing PAs is often considered a very special field distinctive from other RF block designs because of the lack of well characterized models for large signal operations and high voltage and current stress. Thus, designers often rely more on their experience than simulation results when they characterize PAs. Moreover, guaranteeing a reliable operation is as important as achieving good performances such as high output power, high efficiency, and high linearity. In addition to these design difficulties of general PAs, a CMOS PAs are even harder to design. While CMOS technology is welcomed because of its cost-effective material and great versatility, as mentioned in Chapter I, the commercialization of CMOS PAs has not been easily achieved due to the intrinsic drawbacks of standard CMOS processes from RF perspectives: a low-quality factor (Q), the lossy substrate of passive structures, low breakdown voltage, and low transconductance of active devices. Thus, many efforts, including the development of high power generation techniques, linearity enhancement techniques, and efficiency enhancement techniques, have focused on overcoming the drawbacks of CMOS technology for PA designs. Thus, Section 3.2, introduces the structure and the difficulties of designing PAs. Section 3.3 follows with a discussion of the challenges of CMOS

technology in the design of PAs. In response to the challenges outlined in the previous sections, Section 3.4 introduces several state-of-art performance enhancement techniques for CMOS PAs.

3.2 General Issues in Designing RF Power Amplifiers

As shown in Figure 27, a PA consists of four important blocks: an active block that consists of power cells and three passive blocks that comprise bias feeding for the DC current, power combiners, and impedance transformers. Since the passive blocks usually contain inductive characteristics, they can be combined functionally.

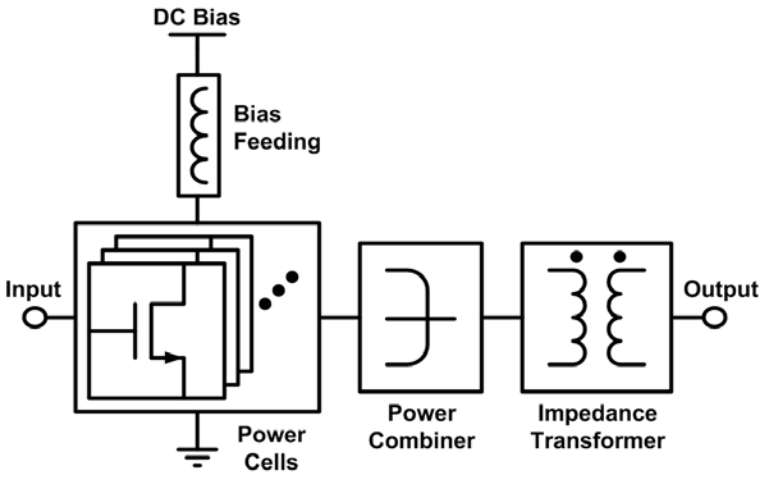


Figure 27. Block diagram of the PA output network

In the bias feeding for the DC current, minimizing the voltage drop is a key factor to keeping the drain voltage as high as possible. PAs typically consume high current, around a few amperes, so even a small series resistance of the bias feeding line can severely deteriorate efficiency. Fully-integrated PAs, the bias feeding line can be implemented in

two ways. One is to use bond wires, well known for their high Q [2] of around 40 to 50 for cellular bands; however, they are subject to inductance variations. The other is to use an integrated inductor that consumes a large area and normally has low Q of less than 30, a best guess by the author.

An impedance transformer is also a key block to determining the amount of power to be transferred from the power device to the load. Even though it draws a large current from the power cells for high output power, the transformed load impedance at the input of the output network turns out to be as small as a few ohms [16]. Since the parasitic resistance of the output network, caused by resistive loss, the skin effect, and the proximity effect [17], occurs together with the transformed load impedance, a portion of power is consumed during this parasitic resistance. Hence, a high Q impedance transformer is preferred if power loss is to be minimized.

Last, when the required output power is higher than the power that a single PA can provide, a power combiner that guarantees output power specifications becomes essential. Since typical power combiners lead to an increase in power loss as the structure becomes more and more complex, the misuse of power combining consumes a large die area and even degrades performance. Therefore, this method of meeting power requirements should be a last resort.

In the design of PAs, a lossy output network is detrimental to achieving high output power, efficiency, and linearity, all crucial characteristics of PAs for wireless applications; thus, output networks require special attention. Output networks have been successfully implemented using conventional techniques such as high-Q off-chip components [18, 19] and integrated patterns onto low-loss substrates (e.g., low-temperature co-fired ceramics)

(LTCC) [20]. Nonetheless, such a modular design requires additional cost. Furthermore, because PAs are extremely sensitive to output networks, the use of semiconductor processes is preferred as they have better tolerance and repeatability [6].

3.3 Challenges of CMOS Technology for RF Power Amplifiers

The challenges mentioned in this section originate from the devices and process characteristics of CMOS field effect transistors (FETs). While such devices are used for other purposes such as digital and small signal RF circuits, if the aim is to design PAs, these characteristics pose serious challenges to PA designers. Before developing performance enhanced PAs, designers must possess a thorough understanding of the issues that will lead to better designs.

3.3.1 Lossy Substrate and Thin Top Metal of Bulk CMOS

As mentioned in Section 3.2, power loss caused by passive structures in a PA layout must be minimized. Figure 28 shows the loss mechanism of inductive structures in bulk silicon (Si) CMOS technologies. On top of a low resistance substrate, dielectric layers are stacked vertically and metal traces are formed within the dielectric layers. When current conduction flows through the metal traces that have series resistance and inductance with parasitic capacitance with neighboring traces, ohmic loss by the resistance affects the quality factor of the metal traces. At the same time, the induced magnetic field penetrates the dielectric and substrate layers. Since most materials, except certain magnetic materials rarely used in semiconductor processes, have the permeability constant value of unity, the magnetic field naturally makes a closed loop not bothered by any layers in the

figure. The field can cause eddy current flow for a low resistive substrate. The substrate conduction current can also flow through the parasitic capacitance in-between the metal traces and the substrate, leading to additional loss. Thus, maximizing the quality factor by reducing eddy currents [21] and minimizing parasitic capacitance [8] are the main design goals of an inductive passive component in CMOS.

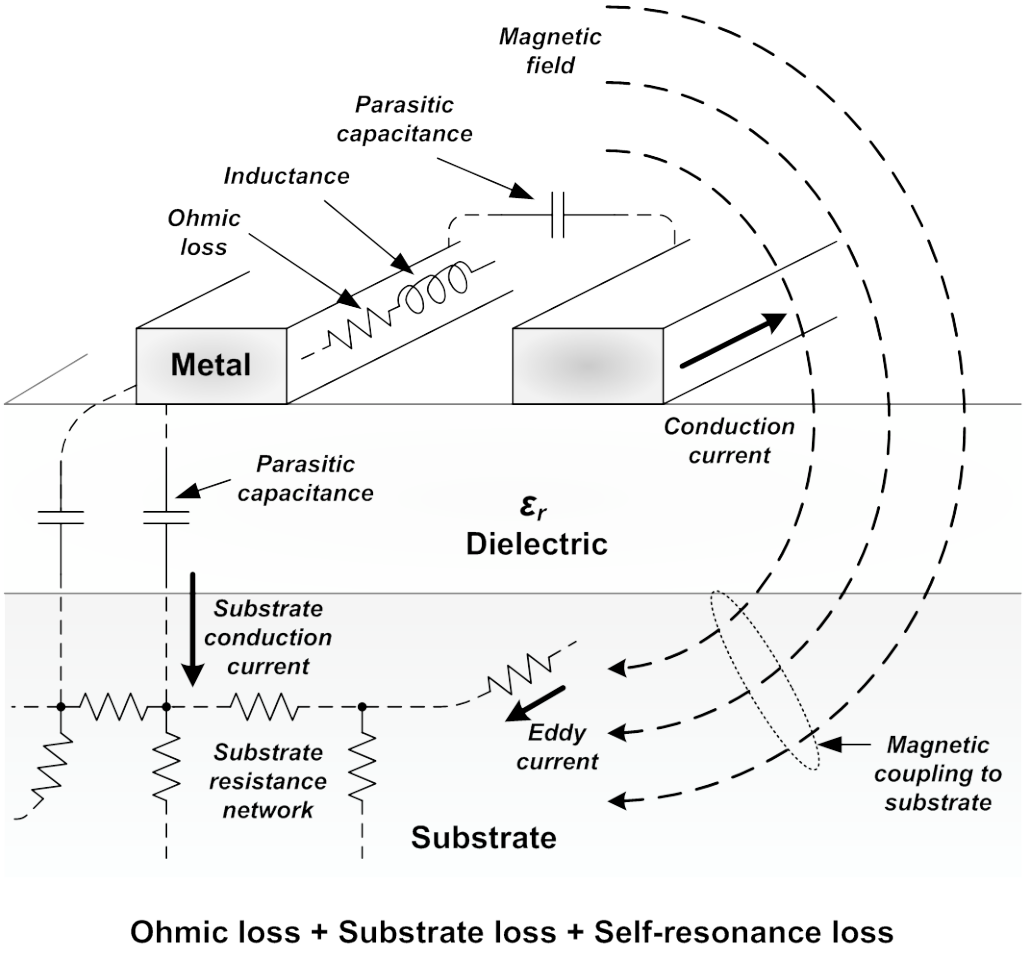


Figure 28. Loss mechanism of inductive structures in CMOS technologies.

The quality factor of an inductive structure is defined as [22]

$$Q = 2\pi \cdot \frac{\text{energy stored}}{\text{energy loss in one oscillation cycle}} \tag{3.1}$$

$$= \frac{\text{peak magnetic energy} - \text{peak electric energy}}{\text{energy loss in one ocsillation cycle}} \tag{3.2}$$

As represented in Equations (3.1) and (3.2), the definition can be further categorized according to three primary mechanisms: the ratio of magnetic energy storage capacity to energy loss (the ohmic loss in the series resistance), substrate loss dissipated in the silicon substrate, and the self-resonance factor by the parasitic capacitance [6, 22] in Figure 28. In a more analytical description, a single-ended spiral inductor can be modeled as illustrated in Figure 29(a), in which R_s and L_s represent the series resistance and inductance, respectively. C_s is the interwinding parasitic capacitance, and C_{ox} is the oxide capacitance on top of the substrate resistance and capacitance, R_{Si} and C_{Si}, respectively.

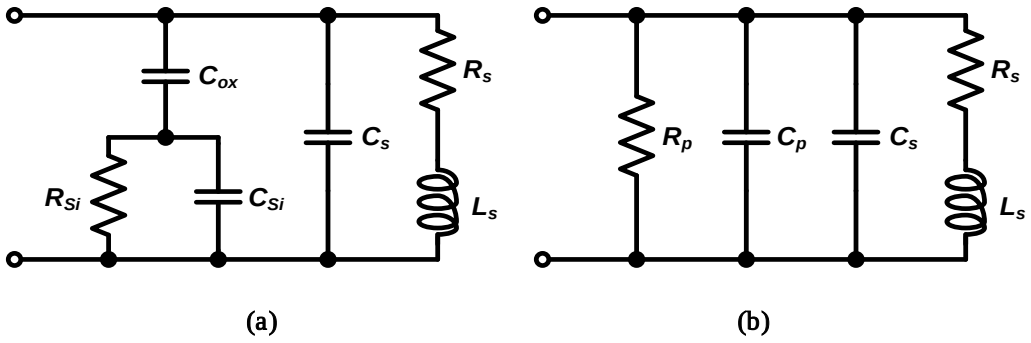


Figure 29. Lumped model of a spiral inductor on silicon: (a) physical model and (b) simplified equivalent model

In the simplified equivalent model in Figure 29(b), R_p and C_p can be derived as

$$R_p = \frac{1}{\omega^2 C_{ox}^2 R_{Si}} + \frac{R_{Si}(C_{ox} + C_{Si})^2}{C_{ox}^2} \tag{3.3}$$

$$C_p = C_{ox} \cdot \frac{1 + \omega^2 (C_{ox} + C_{Si}) C_{Si} R_{Si}^2}{1 + \omega^2 (C_{ox} + C_{Si})^2 R_{Si}^2} \tag{3.4}$$

The quality factor of the inductor can be derived again as follows [22]:

$$Q = \frac{\omega L_s}{R_s} \cdot \frac{R_p}{R_p + \left[\left(\frac{\omega L_s}{R_s}\right)^2 + 1\right] R_s} \cdot \left[1 - \frac{R_s^2 \left(C_s + C_p\right)}{L_s} - \omega^2 L_s \left(C_s + C_p\right)\right] \tag{3.5}$$

$$= \text{ohmic loss factor} \cdot \text{substrate loss factor} \cdot \text{self-resonance factor} \tag{3.6}$$

Ohmic loss stems from the quality of the metal, reflected in its conductance, width, and thickness. The design parameters that control the quality factor of an inductor are the width and the thickness for R_s, and the distance from the substrate to the metal layer. The reason for the series resistance is twofold: the DC metal resistance for low frequency and the skin effect for high frequency [8]. The physical cause of the skin effect is the eddy current induced by a magnetic field. While an AC current flows through the inductor trace, the eddy current redistributes the total current, and most charges are spelled from the center of the metal traces, increasing the series resistance of the inductor. This is called “skin depth” with the following relation,

$$\delta = \sqrt{\frac{2\rho}{\omega\mu}} \tag{3.7}$$

$$R_{skin} = \sqrt{\frac{\omega\rho\mu}{2}} \tag{3.8}$$

where δ (m) is the skin depth, ρ (C/m^3) is charge density, and μ (H/m) is the permeability. To reduce the ohmic loss by the skin effect, the thickness, the width, and the conductivity of the metal traces should be controlled so that it has low series resistance. In a spiral structure, the currents of neighboring metal traces in an equivalent phase expel each other,

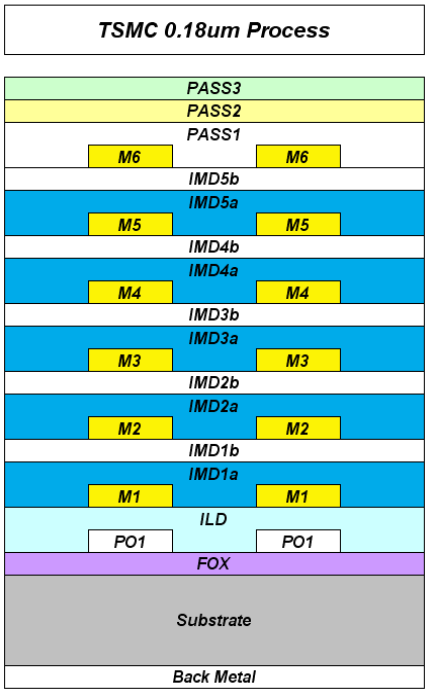
rendering the current distribution more complex. This phenomenon is referred to as the "proximity effect," which is also caused by magnetic flux through the metal traces.

The substrate of CMOS technologies also has finite resistance causing dissipation by the substrate resistance. The parasitic capacitance causes the current through the metal trace of the inductor to leak out along the electric field. To suppress parasitic capacitance, the distance from the substrate to the metal traces should be sufficiently far, but in CMOS technologies, the distance is not controllable. Although techniques such as patterned ground shielding (PGS) can be used to reduce substrate loss, they come at the cost of low self-resonance frequency. Often, PGS is believed to reduce the magnetic coupling to the substrate resistance, but it has little effect [8, 22]. If the substrate resistance were either infinite or zero, no substrate loss would occur and a high quality factor could be expected [8].

While a purely magnetic component can store only magnetic energy, parasitic capacitances store electric energy, and when two energies are equal, self-resonance occurs. This is the point in which the magnetic energy of the inductor is converted to purely electric energy, so the behavior of the inductor resembles a capacitor. To avoid such an effect, the substrate under an inductor or transformer layout is blocked from doping or limited to light doping.

In dedicated RF processes such as the GaAs HBT process, substrate resistivity is higher than 10,000 Ohm·cm (=0.01 S/m); that is, it is virtually an insulator, and the loss in the substrate is low because substrate leakage and the eddy current are suppressed [21]. However, standard CMOS processes are built on a bulk silicon substrate, which has a low substrate resistivity, or high substrate conductivity, of less than 10 Ohm·cm (=10 S/m).

Figure 30 provides an example of standard CMOS layer information. In this example, 1 poly 6 metal layers are shown, and the substrate has a conductivity of 12.5 S/m (=8 Ohm·cm) and the top metal thickness is 2.34 μm. The quality factor has a direct relationship to the metal thickness. A thin top metal that is less than 3 μm provides only a low quality factor far less than 10 for passive designs. If an on-chip inductor is used for matching, this low quality inductor can cause power loss.



TSMC 0.18um Layer

Dielectric Layer	Dielectric Thickness	Er	Conductivity	tan
	um		S/m	
PASS3	0.6	7.9		0.03
PASS2	0.15	4.2		0.03
PASS1	2.5	4.2		0.03
IMD5b	0.35	4.2		0.03
IMD5a	1.18	3.7		0.03
IMD4b	0.2	4.2		0.03
IMD4a	1.18	3.7		0.03
IMD3b	0.2	4.2		0.03
IMD3a	1.18	3.7		0.03
IMD2b	0.2	4.2		0.03
IMD2a	1.18	3.7		0.03
IMD1b	0.2	4.2		0.03
IMD1a	1.18	3.7		0.03
ILD	0.75	4		0.03
FOX	0.35	3.9		0.03
Substrate	250 (10mil)	11.9	12.5	0.03

Metal Layer	Metal Thickness	Conductivity
	um	S/m
M6	2.34	2.40E+07
M5	0.53	2.40E+07
M4	0.53	2.40E+07
M3	0.53	2.40E+07
M2	0.53	2.40E+07
M1	0.53	2.40E+07
PO1	0.2	
BM	10	PEC

Figure 30. Layer information of a standard CMOS process

To contain the magnetic field above the substrate, various methods have been developed. The most popular is to add a PGS on the *M1* layer or on the *PO1* layer in Figure 30 with slots perpendicular to the direction of the eddy current, but with increased

parasitic coupling to the substrate [21]. However, the misuse of this technique often leads to an inferior quality factor.

The loss mechanism of inductive structures in CMOS technologies is summarized in Table 5, which also lists some simple approaches overcoming the loss mechanisms.

Table 5. Loss mechanism of inductive structures in CMOS technologies

Loss factor	Cause	Techniques
Ohmic loss	▪ DC resistance ▪ Skin effect ▪ Proximity effect	▪ High conductivity metal traces ▪ Thick and wide metal traces
Substrate loss	▪ Capacitive coupling ▪ Magnetic coupling	▪ High substrate resistance ▪ Long distance from substrate to metal traces
Self-resonance loss	▪ Parasitic capacitance	▪ Low parasitic layout ▪ Long distance from substrate to metal traces

If a PA is integrated into a transceiver, the large signal coupling to the substrate can corrupt the operation of neighboring transistors, illustrated in Figure 31. The large signal swing at the drain of the center transistor is coupled through the junction capacitance of the drain-body diode to the substrate, and the resistor-diode (or capacitance) network is undesirably modulated. Substrate coupling affects sensitive blocks, especially the low-noise amplifier (LNA), which is very susceptible to noise. Toward the ultimate goal of designing a fully-integrated transceiver, isolation techniques should be properly applied between the PA and other blocks. From a circuit perspective, differential topologies can consume an AC current swing within the circuit itself and minimize the injection of stray electrons and holes into the substrate. The other way is to make guard-rings a more

intuitive and direct method of cutting off the leakage itself, but at the cost of a large signal routing area.

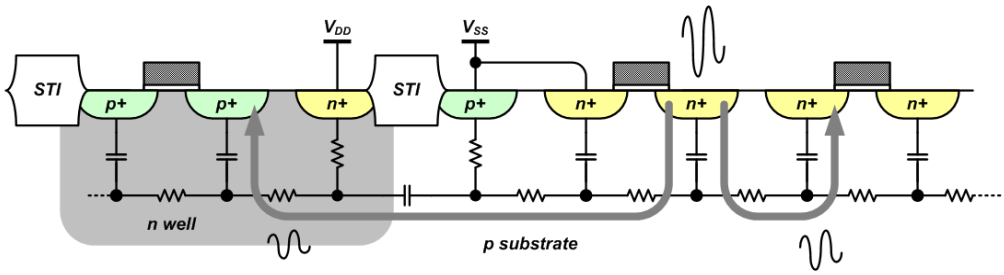


Figure 31. Substrate coupling of a CMOS process

3.3.2 Reliability of Bulk CMOS

Since PAs deal with a large output power, the reliable operation range is often violated in the design of CMOS PAs. CMOS devices in PAs are under high stress, causing a breakdown of devices in extreme cases. The breakdown mechanism of field effect transistors (FETs) can be categorized as shown in Table 6 [23]. Among them, hot carrier effects and oxide breakdown are destructive while junction breakdown and punch-through effects are recoverable when stress conditions are removed. The breakdown mechanisms caused by a large signal swing between the drain and the source or between the drain and the gate are junction breakdown and oxide breakdown, respectively. Although junction breakdown is recoverable, oxide breakdown permanently damages transistors. Thus, preventive measures must be taken in PA designs. Whereas GaAs hetero-junction bipolar transistors (HBTs), which constitutes a dedicated RF process, have high breakdown voltage sustaining up to 20 V, CMOS transistors endure only up to twice the supply voltage [2], which barely meets the supply voltage of cellular terminals.

Table 6. Breakdown mechanism of FETs

Breakdown	Cause	Effect	Damage
Junction (Avalanche) Breakdown	▪ Increase in the drain-source voltage causes excess voltage for the drain-substrate pn-junction ▪ The drain to body diode is operated in the reverse breakdown area ▪ Impact ionization	Abrupt increase in the drain current	Not inherently destructive
Punch-through	▪ Punch-through of the drain depletion region to the source depletion region	Gradual increase in the drain current	Not inherently destructive
Hot Carriers	▪ Injection of electrons and holes into the oxide region with sufficient horizontal or vertical electric fields ▪ Increase in the gate current and carrier trapping	Turn-on of transistors with turn-off conditions	Destructive
Oxide Breakdown	▪ Strong vertical electric field due to excessive gate voltages	Gate leakage current	Destructive

Several efforts have been made to overcome the low breakdown voltage of the CMOS process. One way is to make a thicker gate oxide that bolsters the high gate bias voltage. However, this approach requires process revision and is cost prohibitive. Thus, simple circuit techniques are preferable in a standard CMOS process. For instance, the vertical stacking of devices can reduce the burden on each transistor as long as the number of stacks is large enough to distribute the voltage stress [18].

In addition to the voltage stress of transistors, PAs also suffer from high current driving. A large current swing through passive structures can cause a gradual increase in metal resistance in a long-term operation and the Joule heating effect in a short-term operation ending in the melting of metal lines. These problems call for some guidelines

that will guarantee a reliable operation. While the rule of thumb value, 1 mA/μm, is often suggested, a more rigorous criterion that affords such a guarantee should be used for each process.

3.3.3 Low Transconductance of Bulk CMOS

The current for forward bias and the transconductance of a CMOS device can be expressed [23] as

$$I = \frac{\mu C_{ox}}{2}\frac{W}{L}(V_{GS} - V_t)^2, \tag{3.7}$$

$$g_m = \mu C_{ox}\frac{W}{L}(V_{GS} - V_t), \tag{3.8}$$

respectively, in which all symbols follow the conventional MOSFET parameters of [23]. Then, the ratio of the transconductance to the current is expressed as

$$\frac{g_m}{I} = \frac{2}{V_{GS} - V_t}, \tag{3.9}$$

while the transconductance of a bipolar junction transistor (BJT) shows a relation of

$$\frac{g_m}{I} = \frac{q}{kT}. \tag{3.10}$$

For example, by simply comparing the ratios of both the MOSFET and the BJT at room temperature, the BJTs have a higher value of $q/\,kT$ ~ 1/26 mV, while a typical MOSFET shows $2/(V_{GS} - V_t)$ ~ 1/50 mV with V_{GS} = 600 mV and V_t = 500 mV. In other words, BJTs usually have a higher transconductance ratio per given current through the output by one order of magnitude.

Thus, obtaining both a high gain and a high output current in the CMOS process requires large power cells and strong driving power from previous stages. However,

neither way is desirable in a PA design, and PAs that use BJTs are generally considered more efficient at output power generation.

3.3.4 Nonlinearity of Bulk CMOS

In the design of a linear PA using a standard CMOS process, the first step is to understand the limitations of the given process. Figure 32 shows the equivalent circuit model of a standard RF NMOS transistor. In addition to the transconductance provided by the MOS transistor (based on the BSIM3 model with turned-off junction diodes), other resistances, capacitances, and junction diodes can represent the RF behavior of the transistor. Such ever-present parasitic capacitances and junction diodes distort the AC signal applied to terminals and cause distortion, imposing serious constraints on their anticipated performance [2]. If such distortions are not efficiently suppressed at the device or circuit level, then the PA designed using such a device can cause nonlinear circuit operation.

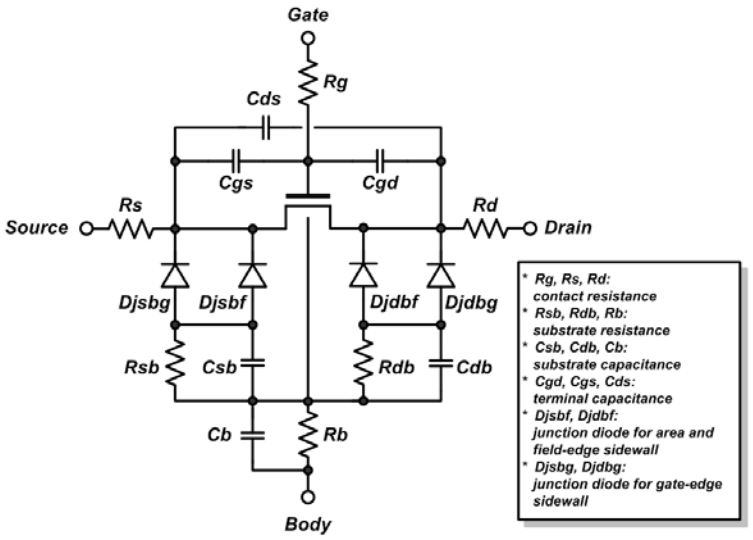


Figure 32. Equivalent circuit model for RF NMOS transistor

Researchers have devoted their efforts analyzing the effects of the components on linearity and the factors responsible for nonlinearity [24]. It is claimed that the factors responsible for the nonlinearity of a device are gate-source capacitance, gate-source transconductance, gate-drain transconductance, and drain-body junction diodes, listed in order of most to least dominant.[25, 26]. However, analyses, which are limited to a small-signal range and ignore the large-signal swing in PA operation, have led to disagreement, requiring further investigation that leads to the generalization of the ideas.

In summary, CMOS processes have inherent problems such as poor reliability and linearity, and poor passive and active performances in RF CMOS PA designs. Thus, before undertaking the next step of developing proper measures for handling the large-signal operation of PAs, we must first understand the challenges that such a task entails.

3.4 Techniques for CMOS RF Power Amplifiers

As introduced in Section 3.3, CMOS devices have low power capability compared to compound semiconductor devices. Due to the low Q of on-chip inductors in CMOS processes, off-chip matching networks have been the dominant methods for completing the PA functions. One breakthrough in the effort to achieve the integration of an entire PA came in the form of the design of an efficient power-combining method using transformers in series [16, 27]. Although this design represented proof of the feasibility of achieving high output power from CMOS processes, because of its bulky size and instability, further effort is required to gain the same output power capability using alternative methods [28]. Since the initial efforts focused only on power capability, the linearity performance was poor, marring the progress toward linear applications. Only

through the broadening of the understanding of CMOS nonlinearities can any approach that eliminates nonlinear factors in CMOS devices be established. Upon until now, however, only a few noticeable findings have been reported [24, 26, 29]. Despite such efforts, system concepts are also being applied to implement a PA in which linearity or efficiency enhancement is realized using more complex signal manipulation [30-33].

3.4.1 Output Power-Combining Techniques

In wireless applications, output power usually reaches watt levels (shown in Figure 2 of Chapter 1). For example, most PA products for GSM applications generate more than two watts. Hence, CMOS processes, known for low power-driving capabilities, necessitate an additional function, i.e. a power combiner, that combines several unit power cells that generate the required output power. Figure 33 depicts several popular power-combining techniques. The *LC* matching and combining network is popular but not suitable for IC integration due to its bulky size and narrowband characteristics. The Wilkinson combiner is also a popular method, especially for high frequency or off-chip implementation, but its bulky quarter-wave transmission lines thwart implementation efforts in cellular applications. Compared to the above techniques, the transformer provides differential combining in a compact form factor with a broadband advantage.

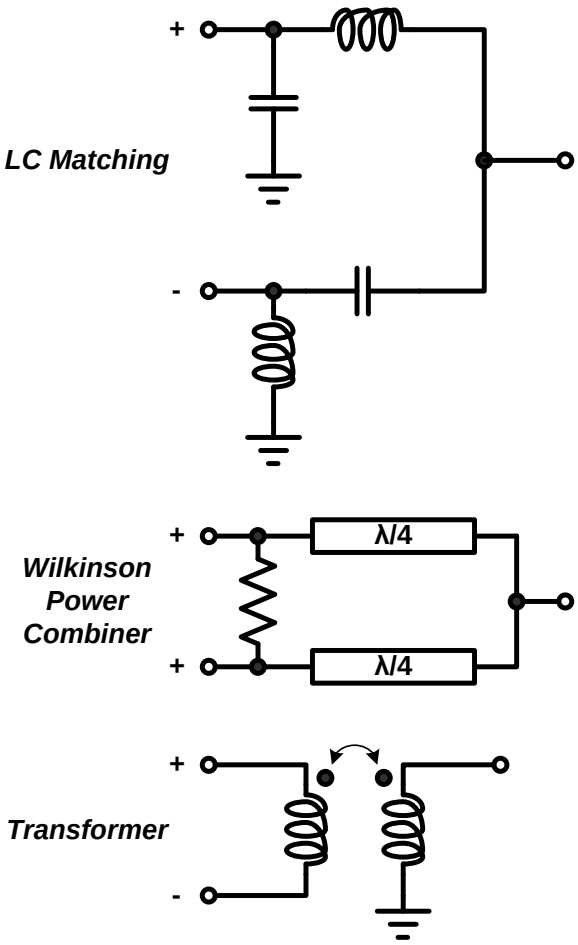


Figure 33. Conventional power-combining techniques

Thus, with the use of transformer-type output networks, several successful efforts have been reported for both power combining and impedance transformation. Some reported techniques, such as distributed active transformers (DATs) [16, 27], shown in Figure 34, or their variations [28, 34, 35], and figure-eight structures [36] are good examples of series-combining transformer (SCT); and recently introduced voltage-boosting parallel-primary transformers [37, 38] are examples of parallel-combining transformer (PCT). Detailed analyses of both techniques are presented in Chapter 4.

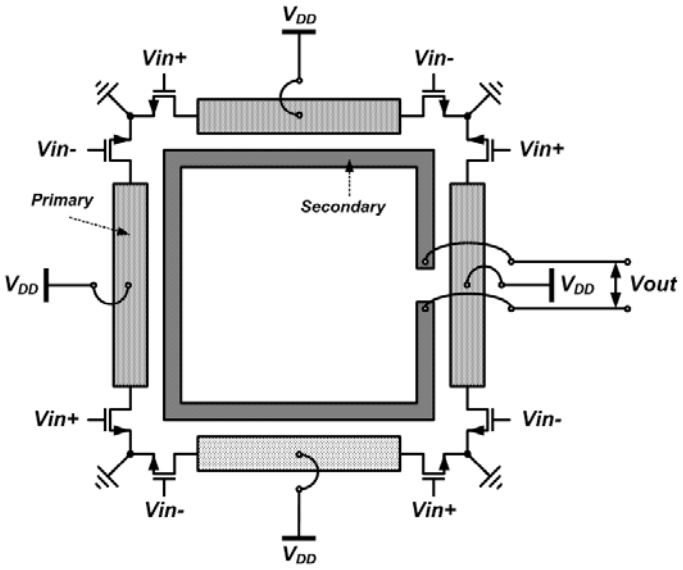


Figure 34. Diagram of a DAT PA

3.4.2 Linearity Enhancement Techniques

To address the linearity of a PA, one must consider both the fundamental cause of nonlinearity and the nonlinear behavior of the PA topics of interest. Pertaining to the causes of nonlinearity, transconductance and gate-source capacitance are considered the key factors that determine linearity among the many conductance and capacitive components [24, 29]. For capacitive components, the absolute amount of capacitance itself is not a problem, but the amount of transition resulting from a large-signal swing is the actual cause of nonlinearity. The neutralization of a capacitance transition can be attained by adding a transistor with an opposite characteristic in parallel with the main transistor. In Figure 35, the main transistor generating output power is an NMOS transistor M_N, and PMOS M_P is in parallel with it, described in Equation (3.11). Then, the input signal swing does not suffer from the transition of input capacitance, and the linearity of the PA can be improved [29]. In this figure, the termination of a second

harmonic is also shown to suppress the return of a second harmonic to the gate of the transistor, thus averting the third-order inter-modulation [24].

When the behavior of a PA is known in advance and the nonlinearity behavior of the PA can be mimicked using other components, the distortion can be suppressed. The pre-distortion of a PA in which the compression of gain and phase are adequately matched with the pre-distorter generating a very linear output signals is shown in Figure 36.

For meeting the specifications of high data-rate communications, linearity enhancement is one of the most important design aspects. Therefore, more efforts are required to guarantee good linearity with increases in PAR.

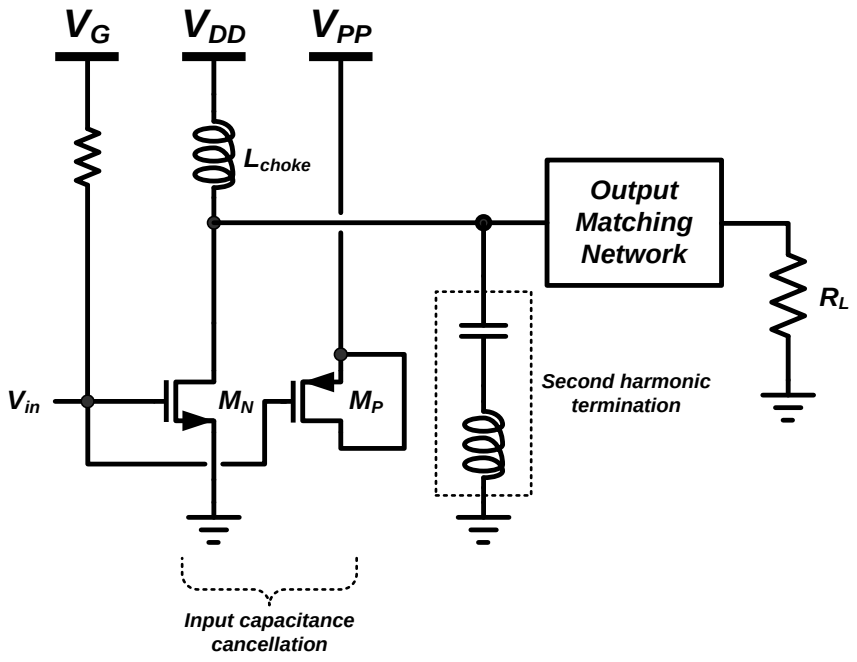


Figure 35. Input capacitance cancellation technique

$$C_{in}(V_{in}) = C_{g,M_N} + C_{g,M_P} \tag{3.11}$$

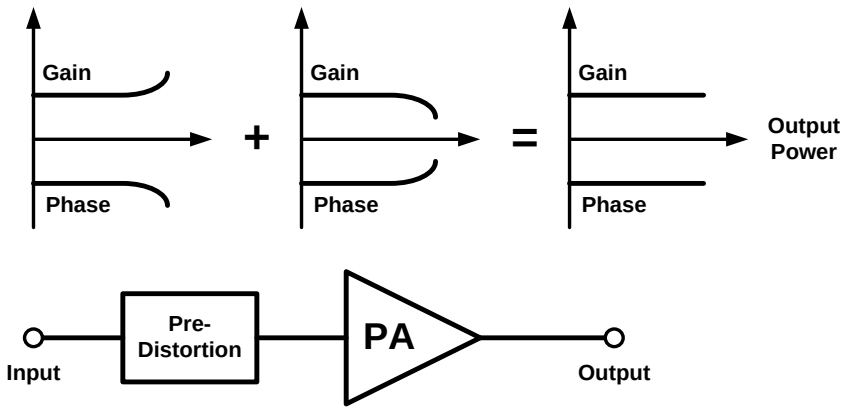


Figure 36. Pre-distortion of a PA

3.4.3 Efficiency Enhancement Techniques

Since the early era of radio signaling a century ago, concepts for system PAs such as the Doherty PA [31], the Kahn PA [32], and the Chireix PA [33] have been developed. Such techniques have been well established and adopted successfully in base stations. However, even for GaAs HBT PA modules, realizing such techniques in a small form factor poses a formidable challenge, so only a few relatively unsuccessful examples have been reported. However, these techniques require the application of efficiency enhancement techniques to CMOS PAs, posing even greater challenges due to their poor passive performance. Therefore, their performance is far from that demanded in commercial production [39-41].

Among such techniques, the polar transmitter, illustration in Figure 37, became very popular recently with the aid of digital signal processing. From the baseband system, an amplitude modulated signal, $A(t)$, and quadrature phase signals, $P_i(t)$ and $P_q(t)$, are generated. The PA has the input of phase modulation but with a constant envelope, but the amplitude modulation through the operation amplifier and PMOS keeps the operation

point of the PA in the saturation mode with a constant envelope that maximizes efficiency. The output of the PA can be represented as Equation (3.12).

$$v(t) = A(t) \cdot \cos\left(\omega_c t + \sqrt{P_i(t)^2 + P_q(t)^2}\right) \tag{3.12}$$

Unfortunately, the polar transmitter system suffers from a mismatch of amplitude and phase loss by the PMOS, and the complicated operation requirements. Therefore, the trade-off between efficiency enhancement and circuit complexity should be carefully considered in advance.

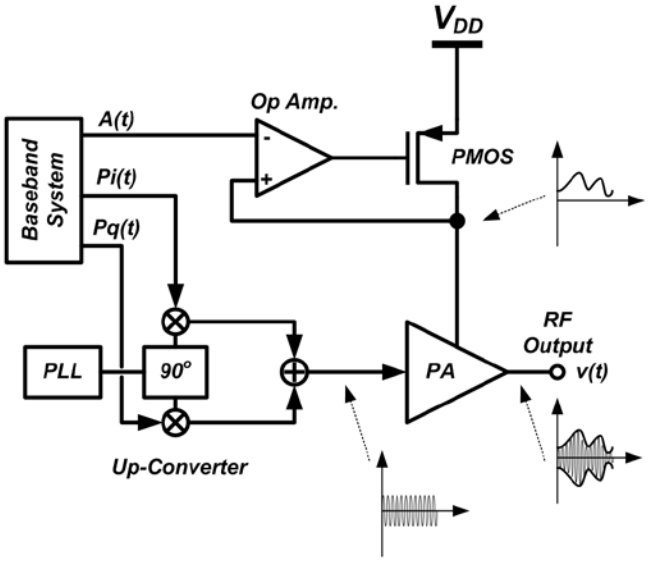


Figure 37. Polar transmitter system

3.5 Conclusion

This research has discussed the challenges for designing CMOS PAs. PA design itself is a challenging task due to large signal operation; however, in CMOS technology, the burden doubles because of the deficiencies of CMOS technology. The design of CMOS

PAs necessitates the mitigation of the CMOS problems: low transconductance, low breakdown voltage, and a low quality factor of passives. To overcome such problems and to generate high output power with high efficiency and high linearity, techniques such as power combining, efficiency enhancement, and linearity enhancement techniques should be used.

CHAPTER 4
POWER-COMBINING TRANSFORMERS AND CLASS-E POWER AMPLIFIER DESIGN

4.1 Introduction

The first obstacle in designing a PA is to attain high output power. Due to the acknowledged inferiority of CMOS technology, simple *LC* networks frequently used for PA products based on compound semiconductors are not suitable for CMOS PAs. Thus, as a viable option for applications that require more than one watt, a transformer technique can be adopted, as stated in Chapter 3.

This chapter presents fully-integrated CMOS PAs with parallel power-combining transformers. For high-power CMOS PA design, Section 4.2 analyzes and compares two types of transformers, a series-combining transformer (SCT) and a parallel-combining transformer (PCT), revealing parasitic resistance and a turn ratio as the limiting factors of power combining. Based on the analysis, Section 4.3 describes two kinds of PCTs, a two-primary with a 1:2 turn ratio and a three-primary with a 1:2 turn ratio, which are incorporated into the design of fully-integrated CMOS PAs in a standard 0.18-μm CMOS process.

4.2 Power-Combining Transformer

As an output network, a general transformer has useful characteristics. First, by controlling the turn ratio, it can obtain impedance matching. Second, it can convert a differential signal to a single-ended signal. Third, as it has physically detached primary and secondary windings, it can provide DC isolation [42] and ESD protection. Therefore, a good solution for integrated RF circuits is a multi-functional transformer. However, a typical transformer cannot be easily matched to the output load for watt-level power generation due to an unrealistically high turn ratio, low Q, and low magnetic coupling [16]. Instead, the distribution of a high turn ratio using multiple transformers with a reduced turn ratio is more practical for integrated designs [16, 27, 28, 34, 35, 43].

Figure 38 shows a conceptual power-combining network interpreted as an impedance network [Z], in which M input ports are connected to the differential unit power cells aligned on the left side, and the single output is defined with the load on the right side. All voltages and currents applied to the ports can interact with each other, which can be modeled with the following relation:

$$\begin{bmatrix} V_{11} \\ \vdots \\ V_{1M} \\ V_2 \end{bmatrix} = \begin{bmatrix} z_{11} & \cdots & z_{M1} & z_{(M+1)1} \\ \vdots & \ddots & \vdots & \vdots \\ z_{1M} & \cdots & z_{MM} & z_{(M+1)M} \\ z_{1(M+1)} & \cdots & z_{M(M+1)} & z_{(M+1)(M+1)} \end{bmatrix} \cdot \begin{bmatrix} I_{11} \\ \vdots \\ I_{1M} \\ I_2 \end{bmatrix} \tag{4.1}$$

Unnecessary complexity in the relationship between the input and output can be averted if it is assumed that the interactions between the primary windings do not exist and that the voltage and current are equally distributed to the primary windings. The input signals can be summarized as follows:

$$V_1 \equiv V_{11} = V_{12} = \cdots = V_{1M} \tag{4.2}$$

$$I_1 \equiv I_{11} = I_{12} = \cdots = I_{1M} \tag{4.3}$$

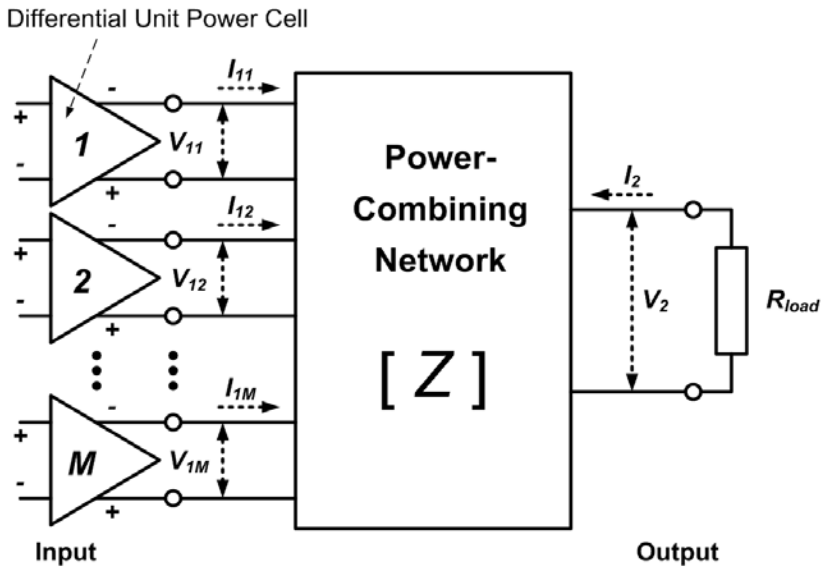


Figure 38. Conceptual power-combining network

Based on these conditions, two types of power-combining networks, series and parallel combining, are analyzed and compared in detail. Figure 39(a) and (b) show the conceptual diagrams of the SCT and PCT, respectively. The system of transformers can be defined as $M \times N_1 : N_2$ [*M : the number of primary windings,* N_1 *: the number of primary turns,* N_2*: the number of secondary turns*]. This analysis focuses on topology-dependent parameters such as N_1, N_2, and M, which are set prior to the layout. Among the layout-dependent parameters such as parasitic series resistance, mutual inductance, and the coupling coefficient, only parasitic series resistance is included in the analysis as the dominant cause of loss in CMOS processes. The mutual inductance and coupling coefficients are assumed to be ideal.

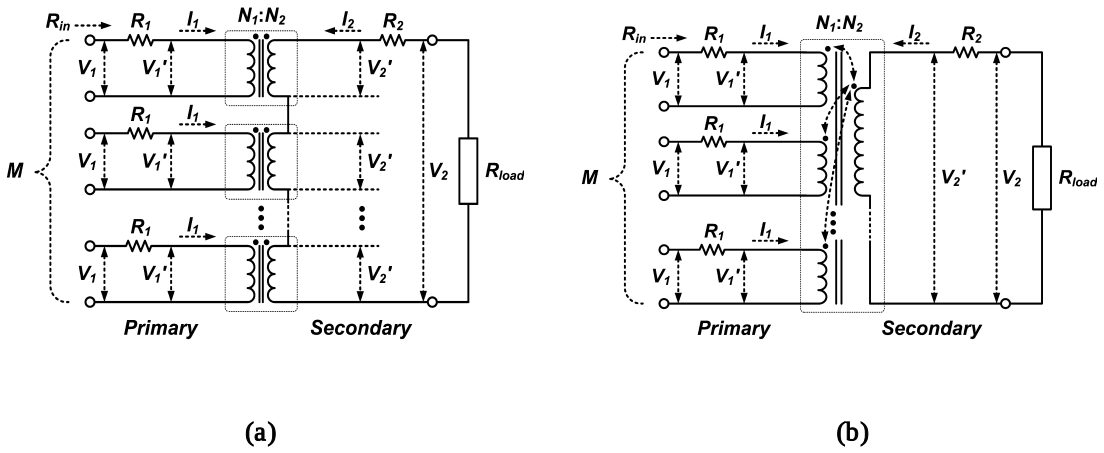


Figure 39. Power-combining transformers (a) SCT and (b) PCT

In Figure 39(a), the secondary windings of the ideal transformers are connected in series sharing the same current, I_2. Accordingly, the current and voltage relations can be given as follows:

$$\frac{I_2}{I_1} = \frac{N_1}{N_2} \tag{4.4}$$

$$V_1^{'} = \frac{N_1}{N_2} V_2^{'} = \frac{1}{M}\frac{N_1}{N_2}(R_2 + R_{load})I_2 \tag{4.5}$$

The input impedance, R_{in}, looking into the transformer, is as follows:

$$R_{in} = \frac{V_1}{I_1} = R_1 + \frac{1}{M}\left(\frac{N_1}{N_2}\right)^2 (R_2 + R_{load}) \tag{4.6}$$

Equation (4.6) consists of two parts: the parasitic resistance term, R_1, and the remaining transformed load impedance term. R_{in}, which decreases as N_2/N_1 and M increase, is limited by R_1. The power-combining ratio (PCR) of the SCT, PCR_{SCT}, can be defined as the actual power-combining capability when M is given.

$$PCR_{SCT} \equiv \frac{P_2}{P_1} = \frac{V_2 \cdot I_2}{V_1 \cdot I_1} = \frac{\left(\frac{N_1}{N_2}\right)^2 R_{load}}{R_1 + \frac{1}{M}\left(\frac{N_1}{N_2}\right)^2 (R_2 + R_{load})} \tag{4.7}$$

where P_1 is the power of the unit power cell and P_2 is the output power. Representing the performance of the transformers, transformer efficiency is defined as the ratio of output power to total input power. For the SCT,

$$\eta_{SCT} \equiv \frac{P_2}{MP_1} = \frac{PCR_{SCT}}{M} = \frac{1}{1 + \left[M\left(\frac{N_1}{N_2}\right)^2 R_1 + R_2\right]/R_{load}}. \tag{4.8}$$

In the ideal case where R_1 and R_2 are zero, PCR_{SCT} and η_{SCT} become M and unity, respectively.

A similar approach is used for the analysis of the PCT in Figure 39(b). In the PCT, it is assumed that the current at the secondary winding is equally attributed from the primary windings. The current and voltage relationships follow:

$$\frac{I_2}{I_1} = M\frac{N_1}{N_2} \tag{4.9}$$

$$V_1^{'} = \frac{N_1}{N_2}V_2^{'} = \frac{N_1}{N_2}(R_2 + R_{load})I_2 \tag{4.10}$$

The input impedance can be derived as

$$R_{in} = \frac{V_1}{I_1} = R_1 + M\left(\frac{N_1}{N_2}\right)^2 (R_2 + R_{load}) \tag{4.11}$$

Equation (4.11) also has two parts: the parasitic resistance, R_1 and the transformed load impedance term. An increase of N_2/N_1 can cause R_{in} to decrease, but unlike Equation (4.6), M can compensate for the effect of N_2/N_1. Thus, as M increases, the input impedance of the PCT becomes less sensitive to R_1 compared to that of the SCT. As for the power-combining capability, PCR_{PCT} is represented as

$$PCR_{PCT} \equiv \frac{P_2}{P_1} = \frac{V_2 \cdot I_2}{V_1 \cdot I_1} = \frac{M^2 \left(\frac{N_1}{N_2}\right)^2 R_{load}}{R_1 + M\left(\frac{N_1}{N_2}\right)^2 (R_2 + R_{load})}. \quad (4.12)$$

For a high M value, the effect of R_1 is overwhelmed by the transformed load impedance; thus, PCR_{PCT} approaches M. The transformer efficiency of the PCT is

$$\eta_{PCT} \equiv \frac{P_2}{MP_1} = \frac{PCR_{PCT}}{M} = \frac{1}{1 + \left[\frac{1}{M}\left(\frac{N_1}{N_2}\right)^2 R_1 + R_2\right]/R_{load}}. \quad (4.13)$$

An increase of M and a decrease of N_2/N_1 would make η_{PCT} approach unity, the ideal transformer efficiency.

Such derived equations can be plotted according to M and N_2/N_1, assuming that R_1 = 3 Ω and R_2 = 6 Ω, which are closely chosen from the example in Table 7 in Section 4.3. Although these values can vary, they still produce the same tendencies. The output impedance is set at R_{load} = 50 Ω for real cases. For the power calculation, the unit power cells in Figure 38 are assumed to be in a differential class-A operation with a 3.5 V supply bias; that is, V_1 = 7 V. Then, the power of the unit power cell, P_1, and the output power, P_2, can be calculated as follows:

$$P_1 = \frac{V_1^2}{2R_{in}} \quad (4.14)$$

$$P_2 = \frac{V_2^2}{2R_{load}} = PCR \cdot P_1. \quad (4.15)$$

As shown in Figure 40(a), the input impedance, R_{in}, of the SCT decreases with an increase of N_2/N_1 and M. A smaller input impedance indicates that higher input power can be delivered from the unit power cells into the transformer by generating more current. On the other hand, in Figure 40(b) for the PCT, the input impedance increases

according to M. Thus, it is expected that the input power decreases with a higher M in the PCT. The effects can be observed in Figure 41.

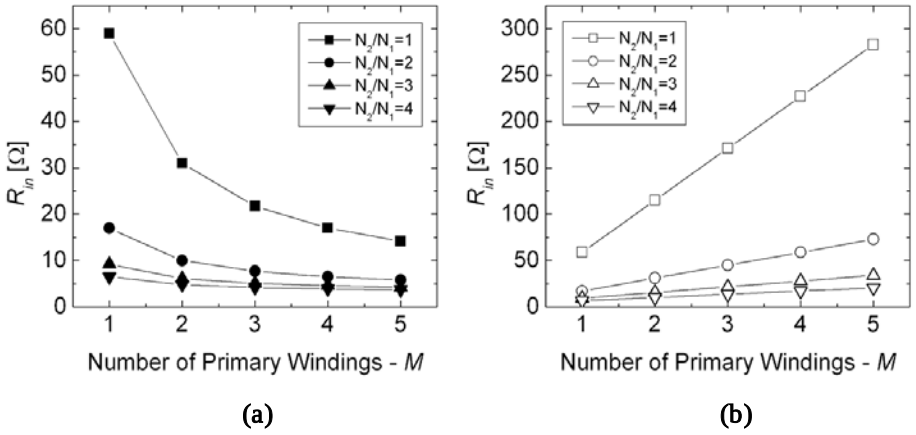


Figure 40. Input impedance (R_1 = 3 Ω, R_2 = 6 Ω, R_{load} = 50 Ω) (a) SCT and (b) PCT

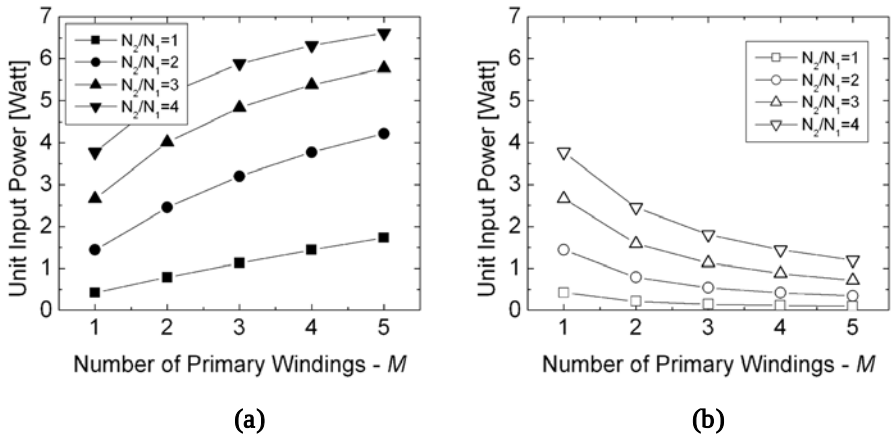


Figure 41. Power of unit power cell (R_1 = 3 Ω, R_2 = 6 Ω, R_{load} = 50 Ω, differential class-A operation, $V_1/2$ = 3.5 V) (a) SCT and (b) PCT

Figure 41 shows the calculated unit input power, P_1, to each primary winding using Equation (4.14). In Figure 41(a) of the SCT case, the unit input power rapidly increases with N_2/N_1, requiring a large unit power device that can produce more than one watt,

possibly leading to instability. Hence, use of unity N_2/N_1 for the SCT is recommended. However, Figure 41(b) shows that the input power of the PCT case gradually increases as N_2/N_1 increases.

Figure 42 shows the comparison of PCRs using Equations (4.7) and (4.12). Figure 42(a) shows that the SCT loses its capability to combine unit input powers as N_2/N_1 increases, limiting the effective range of N_2/N_1. In Figure 42(b), the PCT shows a slight degradation of the PCR with an increase of N_2/N_1 and a linear relationship with M.

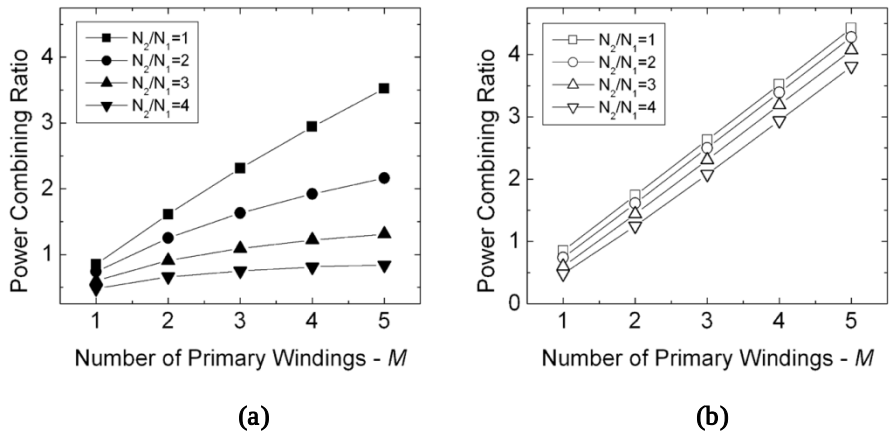


Figure 42. Power combining ratio (R_1 = 3 Ω, R_2 = 6 Ω, R_{load} = 50 Ω) (a) SCT and (b) PCT

Transformer efficiencies between the SCT and PCT are also compared in Figure 43. The efficiency drop in the SCT in Figure 43(a) with higher turn ratios is attributed to the effect of R_1. While PAs transform the load impedance to a very low input impedance through an output network, this parasitic resistance, R_1, becomes of comparable value or even the dominant value with higher turn ratios, and in turn consumes input power at the primary windings. However, in the case of the PCT in Figure 43(b), the increased M

compensates for this effect, as expected in Equations (4.8) and (4.13), and the transformer efficiency for the PCT can even be improved.

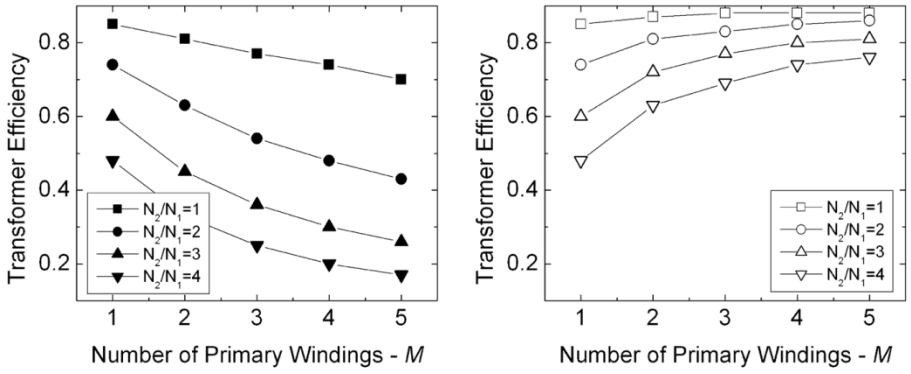


Figure 43. Transformer efficiency (R_1 = 3 Ω, R_2 = 6 Ω, R_{load} = 50 Ω) (a) SCT and (b) PCT

Figure 44 shows the calculated output power. The comparison shows that the SCT has a higher power-driving capability, generating output power as high as ten watts, while the PCT has low output power below four watts under the given parasitic assumption. The output power is the product of the unit input power and the PCR. In the case of the SCT, even though the PCR is not good enough, a high input power capability can compensate for this low PCR. However, high output power is achievable only when the large unit power device can generate high enough power, as much as the input power handling capability of the transformer, without any stability issue. By contrast, the PCT has a good passive characteristic with high PCR and transformer efficiency, but because of low input power driving, the total output power is relatively low.

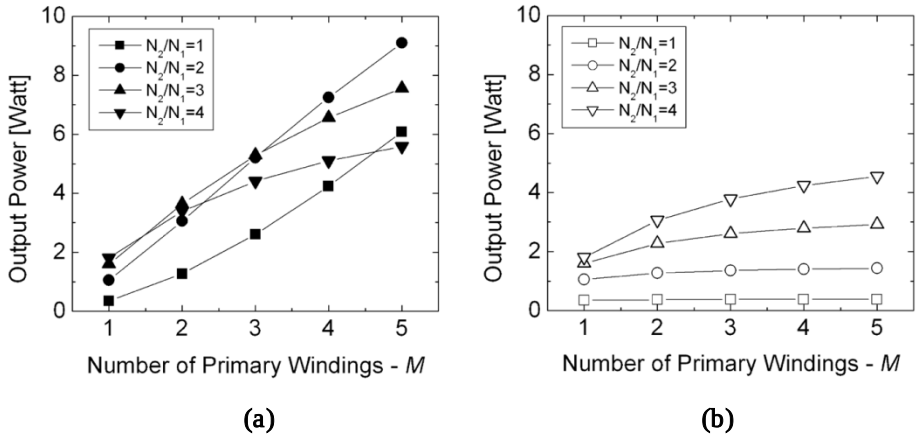


Figure 44. Output power (R_1 = 3 Ω, R_2 = 6 Ω, R_{load} = 50 Ω, differential class-A operation, $V_1/2$ = 3.5 V) (a) SCT and (b) PCT

To define the usable ranges of the two types of transformers, the input impedances, R_{in}, looking into the transformers in Equations (4.6) and (4.11) can be represented as Figure 45. The transformed load impedances in Equations (4.7) and (4.8) are desired to be less than the load impedance to generate lower impedance for power cells, which is the key role of matching. At the same time, however, the transformed load impedances should be larger than the parasitic series resistances; otherwise, parasitic power loss by this resistance may be dominant over the transferred power to the load. Therefore, the suggested usable range of impedance step-down is limited, as shown in Figure 46. For SCT,

$$R_1 < \frac{1}{M}\left(\frac{N_1}{N_2}\right)^2 (R_2 + R_{load}) < R_{load} \tag{4.16}$$

and it can be rearranged as

$$\sqrt{\frac{1}{M}} < \frac{N_2}{N_1} < \sqrt{\frac{1}{M} \cdot \frac{R_2 + R_{load}}{R_2}} \tag{4.17}$$

For PCT,

$$R_1 < M\left(\frac{N_1}{N_2}\right)^2 (R_2 + R_{load}) < R_{load} \tag{4.18}$$

and it can be rearranged as

$$\sqrt{M} < \frac{N_2}{N_1} < \sqrt{M \cdot \frac{R_2 + R_{load}}{R_2}} \tag{4.19}$$

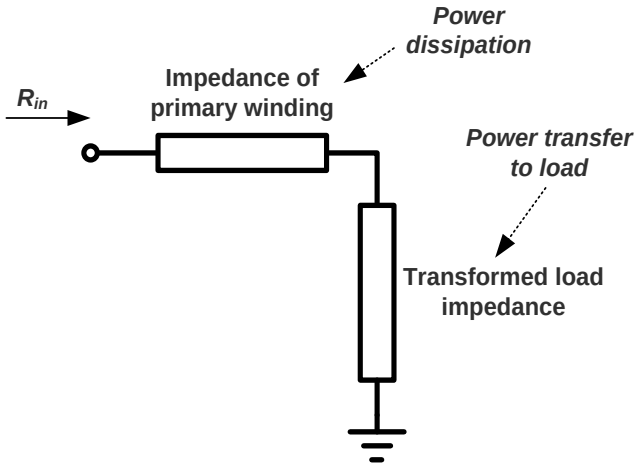


Figure 45. Input impedance looking into a transformer

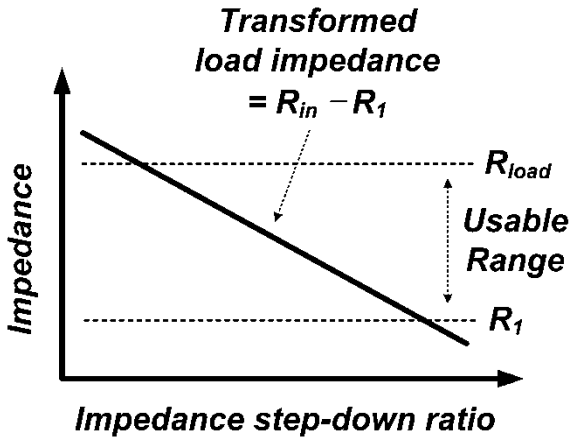


Figure 46. Usable range of impedance transformation

The design range can be plotted as shown in Figure 47.

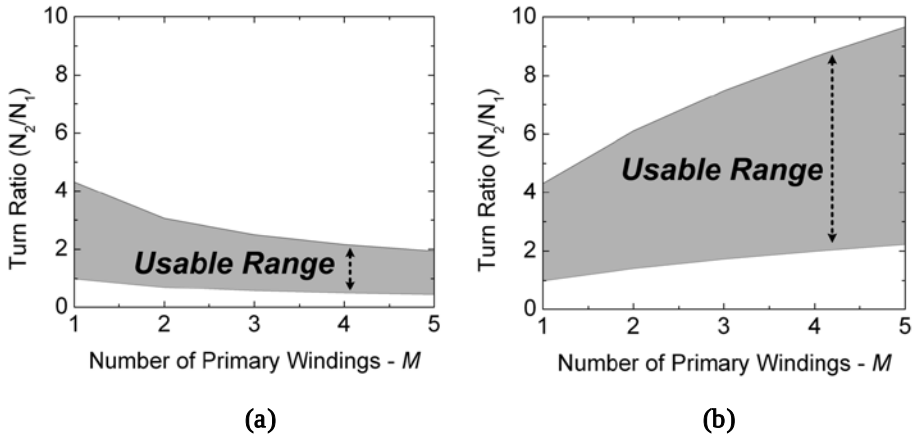


Figure 47. Usable range of turn ratio (R_1 = 3 Ω, R_2 = 6 Ω, R_{load} = 50 Ω, differential class-A operation, $V_1/2$ = 3.5 V) (a) SCT and (b) PCT

In summary, the SCT is advantageous for its high output power generation based on good input power driving but shows comparably poor transformer efficiency. The reasonable design criterion is to have a low turn ratio, N_2/N_1, close to 1 and an appropriate M for a required output power based on the applications. PAs with DATs [16, 27, 34, 35] or figure-eight structures [36], most of which have a turn ratio of N_1: N_2=1:1 and four primary windings, are good design examples within this criterion. On the other hand, the PCT is efficient for watt-level PAs using a high quality passive characteristic and low input power driving. If possible, the turn ratio, N_2/N_1, and M are chosen within a proper range of low values to avoid unnecessary layout complexity. If the size of the unit power cells is limited because of stability and if only a lossy substrate such as a CMOS process is available, the PCT becomes a good candidate for the output network by providing low-input driving capability and insusceptibility to the parasitic resistance of the primary windings.

Once a required output power is given, N_1, N_2, and M are chosen accordingly. As shown in Figure 44(b) and Figure 47(b), a PCT with N_2/N_1 = 2~3, M = 2~3 is deemed to be a candidate for a one- to two-watt PA design.

4.3 Switching Power Amplifier Design Using Parallel Power-Combining Transformer

4.3.1 Transformer Design

In the physical implementation of the PCT, non-ideal behaviors such as the effects of the skin effect and the proximity effect, loss resistance, substrate coupling, and parasitic capacitance between windings degrade performance, causing a power loss. To minimize the loss, a process with low metal resistance, low substrate conductivity, and larger distance between the metal and substrate is preferred [1, 42]. From this point of view, a general standard CMOS process is not considered a good candidate for on-chip transformers in terms of material quality. To meet such a stringent requirement of PA output networks using a qualitatively inferior passive structure, metal traces with very wide widths such as “slab inductors” for DATs [16, 27] or “power inductors” [44] have been used to implement high-Q passives.

In this design, a standard 0.18-μm CMOS process with six aluminum metal layers is used. This process provides a top metal with a thickness of 2.34-μm and a dielectric thickness of 8 μm. To draw the layout of the proposed transformers, we must follow several design guidelines. First, adjacent metal traces should belong to different windings; that is, the primary winding should not neighbor the same or different primary windings;

however, the secondary winding should not neighbor the secondary winding itself. Both situations decrease self inductance and increase mutual inductance [1]. Second, the length of each primary winding is adjusted until all the primary windings have equal phase delays from the inputs of the primary windings to the ports of the secondary winding, which will prevent destructive current coupling. Hence, primary metal traces need to be interweaved so that they have equal electrical length. Third, the distance between opposite edges should be as far as possible to suppress negative magnetic coupling; that is, the shape should be similar to a regular polygon. Fourth, the ports of primary windings should be aligned at one edge of the layout for easy connection to unit power cells, while the secondary port needs to be positioned on the opposite side to maximize the distance between the input and output.

In addition, physical specifications such as the outer dimension, the width of the metal traces, and the spacing between them are key design parameters optimizing the performance. The outer dimension is closely related to the inductance for mutual coupling, the width and thickness determine the Q of metal traces, and the spacing determines the dominant mechanism between capacitive coupling and magnetic coupling [17].

Based on these design guidelines and parameters, we have designed two transformers, 2×1:2 and 3×1:2, illustrated in Figure 48. Transformers with a similar structure but different layouts have been studied in [1, 45]. The 2×1:2 transformer in Figure 48(a) has two input ports (*P1, P2*) and one output port (*P3*), while the 3×1:2 transformer in Figure 48(b) has three input ports (*P1-P3*) and one output port (*P4*). The width and the spacing of the metal traces are 30 μm and 5 μm, respectively. The outer dimensions are adjusted

according to the size of the die and the frequency band of interest, resulting in 0.85 mm × 0.73 mm for the 2×1:2 transformer and 0.90 mm × 0.68 mm for the 3×1:2 transformer. The specific characteristics of the primary and secondary windings are summarized in Table 7.

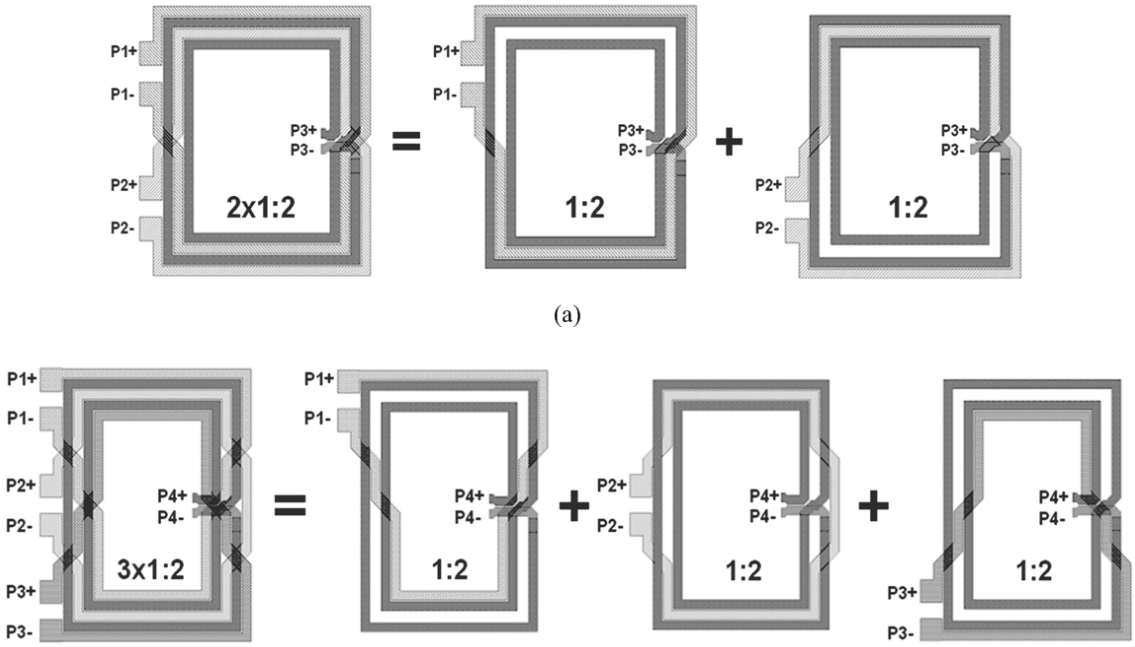


Figure 48. Proposed transformers (a) 2×1:2 PCT and (b) 3×1:2 PCT

The actual turn ratios are calculated from the primary and secondary self inductances [16]:

$$\frac{N_2}{N_1} = \sqrt{\frac{L_2}{L_1}} \tag{4.20}$$

and the mutual inductances and the coupling coefficients are calculated based on the following equations, respectively.

$$L_T = k \cdot \sqrt{L_1 \cdot L_2} \tag{4.21}$$

$$k = \sqrt{\frac{\mathrm{Im}(Z_{12}) \cdot \mathrm{Im}(Z_{21})}{\mathrm{Im}(Z_{11}) \cdot \mathrm{Im}(Z_{22})}} \tag{4.22}$$

Table 7. Simulated characteristics of the primary and secondary windings

Transformer	**2×1:2**		**3×1:2**	
	Primary	Secondary	Primary	Secondary
Self Inductance (L_1, L_2)	1.89 nH	5.07 nH	1.68 nH	5.25 nH
Resistance (R_1, R_2)	2.93 Ohm	5.74 Ohm	3.16 Ohm	7.29 Ohm
Quality Factor (Q)	7.31	9.99	6.02	8.14
Turn Ratio (N_2/N_1)	1.64		1.77	
Mutual Inductance (L_M)	2.23 nH		1.78 nH	
Coupling Coefficient (k)	0.72		0.58	
Total Mutual Inductance (L_{TM})[1]	2.03 nH		1.85 nH	
Total Coupling Coefficient (k_T)[2]	0.78		0.82	

Differential mode, Frequency = 1.8 GHz

1,2 All inputs of the primary windings are excited in-phase as a single input.

The frequency responses of the designed transformers when driven differentially are shown in Figure 49. As the frequency increases, the primary windings of the 2×1:2 transformer differ only slightly in their magnitudes of the transmission coefficients (*S31, S32*) and phase differences, $|\angle(S13) - \angle(S23)|$, as shown in Figure 49(a). However, in the case of the 3×1:2 transformer, the primary winding, *P2* of Figure 48(b), suffers from an increasing difference in the magnitude of the transmission coefficients (*S41, S42, S43*) and the phase response, $|\angle(S14) - \angle(S24)|$, as illustrated in Figure 49(b), which is partially caused by an non-equivalent layout of the primary windings, specifically the center primary winding, which has a different layout, and partially because of the

increasing effect of the interweaving capacitance as the layout becomes complicated. The fringing capacitance between the adjacent metal traces can be minimized when the traces have equal phases. The amplitude and phase errors can cause a destructive overlap of currents at the secondary winding, resulting in failure to achieve optimum performance. The best performance is expected only when all the primary windings are excited in synchronization. Therefore, the phase error in the 3×1:2 transformer may lead to an increase in the parasitic capacitance of high-frequency applications [22]. Since such a distributed parasitic effect is not easily detuned, the initial layout requires meticulous attention to detail so that this imbalanced can be minimized [1].

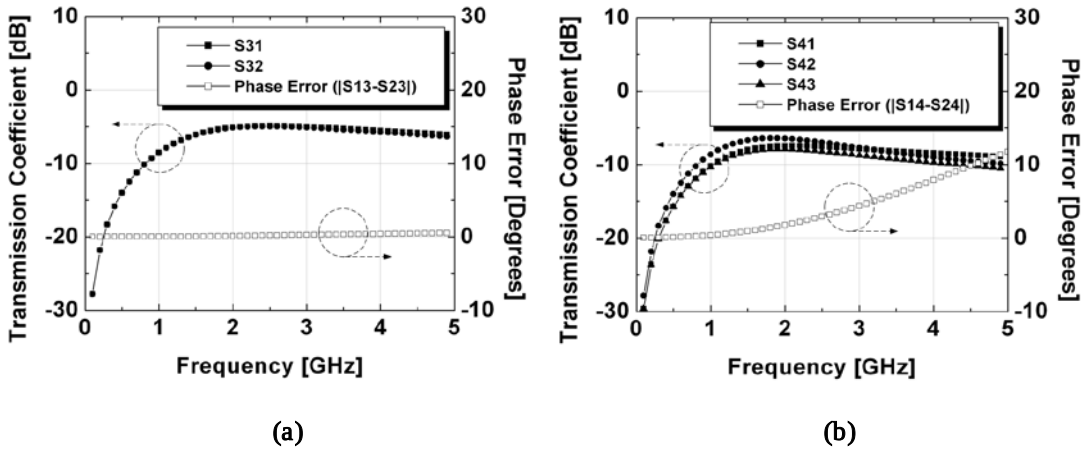


Figure 49. Simulated transmission coefficients and phase error (a) 2×1:2 PCT and (b) 3×1:2 PCT

Figure 50 shows the simulated efficiencies of the transformers, 70% in a single-ended load and 78% in a differential mode based on the calculation of maximum available gain [17]. In a differential mode, a higher quality factor and broad bandwidth are expected [1, 8]. The calculation based on Table 7 with Equations (4.8) and (4.13) provides 84% and 80% transformer efficiencies for the 2×1:2 and 3×1:2 transformers, respectively. The

difference comes from an ideal assumption of the analysis, in which mutual inductance and coupling coefficients are not considered.

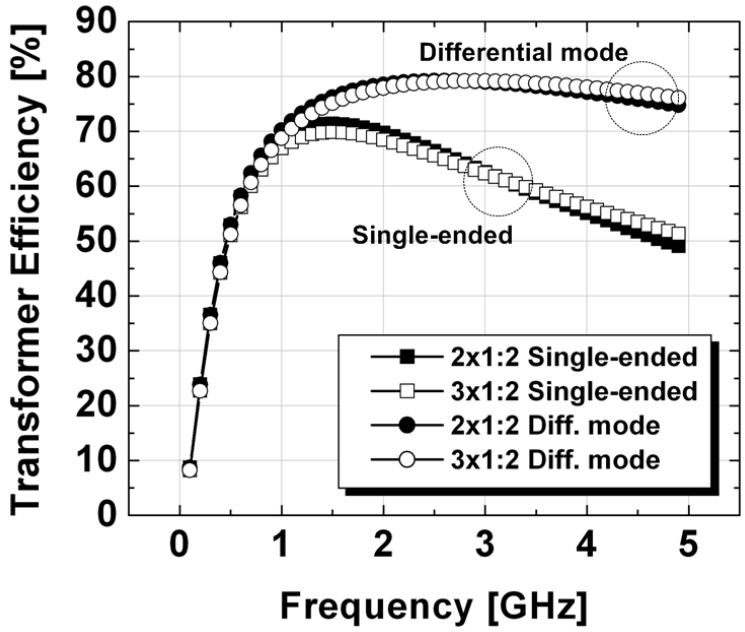


Figure 50. Simulated transformer efficiencies

4.3.2 Switching Power Amplifier Design

A schematic diagram of the designed CMOS PA is presented in Figure 51. The second driver stage and the power stage can be extended by adding the same blocks in parallel. In this figure, three parallel stages are shown. To obtain enough gain and power driving, we adopted a three-stage amplifier cascade chain. For the single-ended input and output, two balun functions are required at the input and output. Thus, the five key functional blocks are an input balun (*T1*), first and second driver stages, a power stage, and an output network with a balun function (*T2*).

For the driver amplifiers, inverters that include an NMOS-PMOS stack like *M1* and *M3* in Figure 51 are used. These driver amplifiers provide signals to the power stage with a maximum swing from the ground to the supply voltage without using inductor bias feeding, resulting in design simplification. In the case of a switching PA, only the phase information of signals is transferred through the PA so that the saturated driver stages can drive the power stage without a loss of information. Between the second driver stage and the power stage, each unit power cell has its own interstage matching inductor (*L1-L3*) to resonate out the gate capacitance. Since a unit power cell is a parallel connection of transistors, an unwanted large capacitance hinders driving of the power stages. The DC common mode of the driver output is the gate bias of the power stage. For 5 dBm input power, 0.85 V of the gate bias is supplied to the power stage.

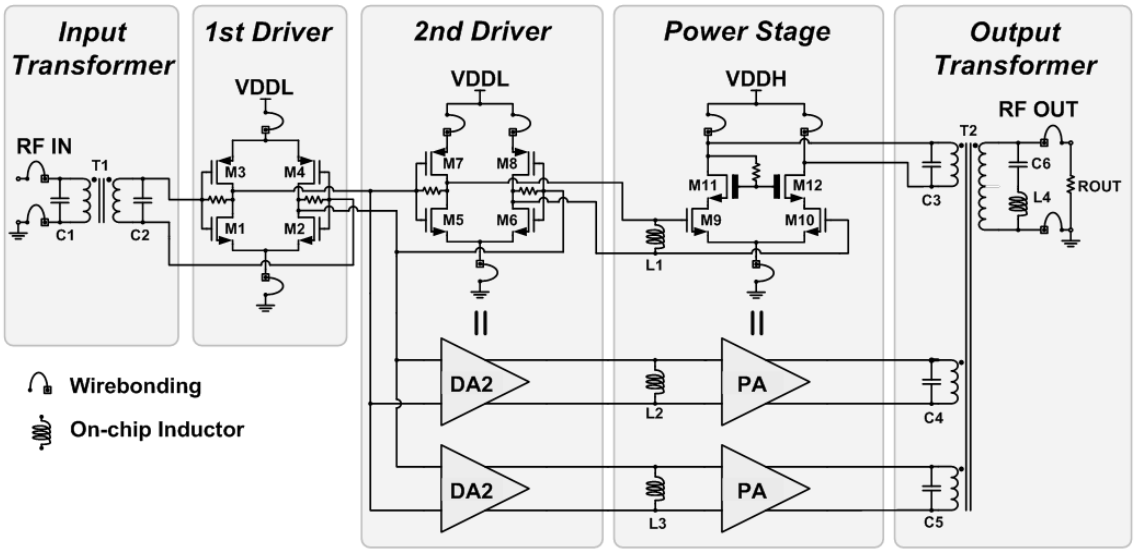


Figure 51. Schematic diagram of a class-E PA

For high efficiency, a topology of the class-E switching PA family [46] is used in the power stage as stripped to its essentials in Figure 52. A DC supply, switch, switch-off bypass element, phase corrector, resonance, and load are connected in series to constitute a class-E PA. When the switch is on, it transfers the supplied DC current to the output load. When the switch is off, the switch-off bypass element takes over the role of load modulation. The resonance block guarantees a pure sinusoidal signal at the output load. For maximum efficiency, the generated voltage and current should not be overlapped following the class-E PA conditions [47, 48]. The phase corrector performs the role of fine tuning by adding a reactive phase amount until the current and voltage at the switch node do not overlap. Two single-ended class-E PAs are combined as a differential pair, the passive elements are simplified, and the final load is replaced with a transformed load to further reduce the number of matching elements. The final topology in Figure 52 is the reference topology used in the schematic diagram of the designed PA in Figure 51.

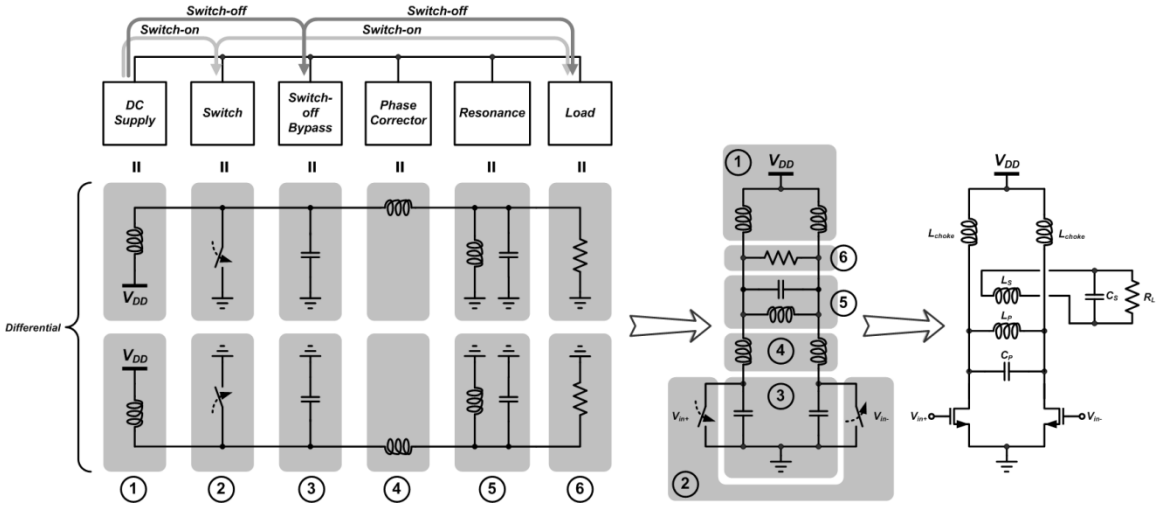


Figure 52. Design of the class-E PA

Because of the high voltage stress of class-E PAs that occurs from the drain to the gate, and from the drain to the source, a protection scheme is required. Generally, a CMOS device can sustain only double a given supply voltage [2]. Therefore, to prevent the voltage stress of CMOS devices, we use a cascode topology, particularly for the power stage, as *M9* and *M11*. In the simulation, the stacking of the thick-oxide 0.35-μm common gate (CG) and thin-oxide 0.18-μm common source (CS) can endure up to around 10 V, distributing the voltage stress to the CG and CS stacks as 6.2 V and 3.6 V, respectively as shown in Figure 53 [18].

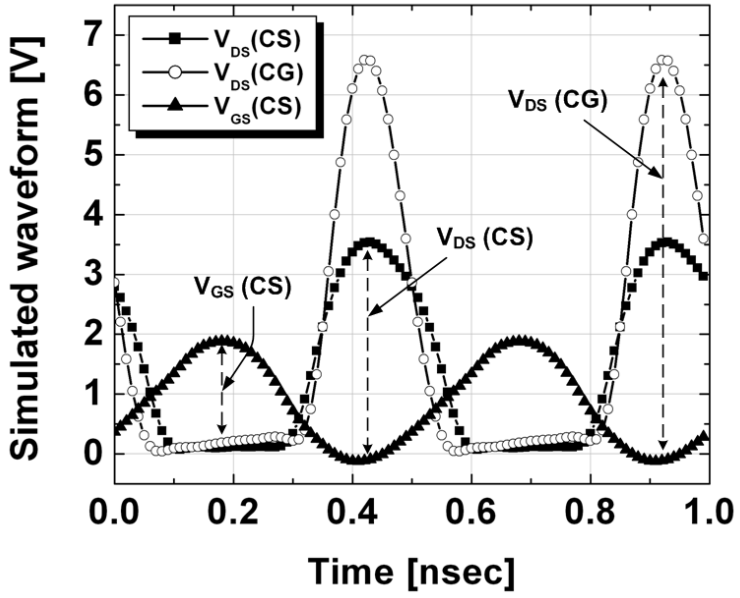


Figure 53. Simulated waveform of the power stage

At the output port, a notch filter (*C6, L4*) is attached in parallel to suppress the third order harmonic component. The chain blocks, except the input and output ports, are all differential operations that will double the voltage swing, achieving high output power at

the same supply voltage. The output transformer (*T2*) has input matching capacitors (*C3-C5*) and an output matching capacitor (*C6*) for maximizing the transfer of output power.

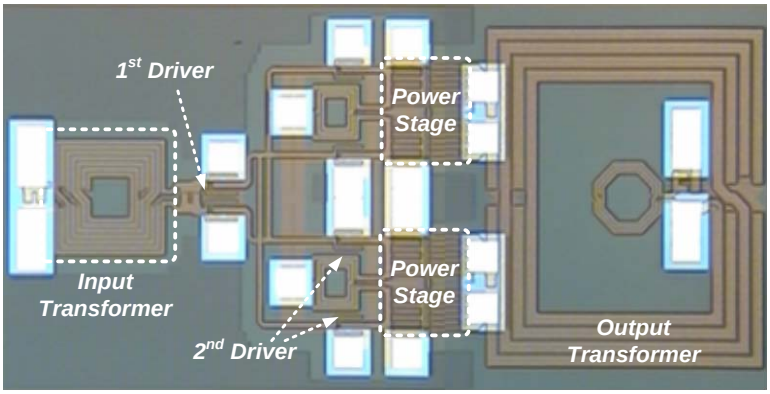
4.3.3 Measurement Results

To verify the feasibility of the proposed power-combining transformers, we have fabricated two fully-integrated switching PAs with the 2×1:2 and 3×1:2 PCTs, respectively, in a standard 0.18-μm RF CMOS process. The 2×1:2 transformer PA in Figure 54(a) is 0.96 mm × 2.04 mm, and the 3×1:2 transformer PA chip with bonding wires in Figure 54(b) is 1.18 mm × 1.9 mm. Some critical bonding wires over the transformers are bonded at a right angle for minimizing unwanted magnetic coupling as illustrated in Figure 55. With the exception of the number of power stages, the two designs are similar in composition.

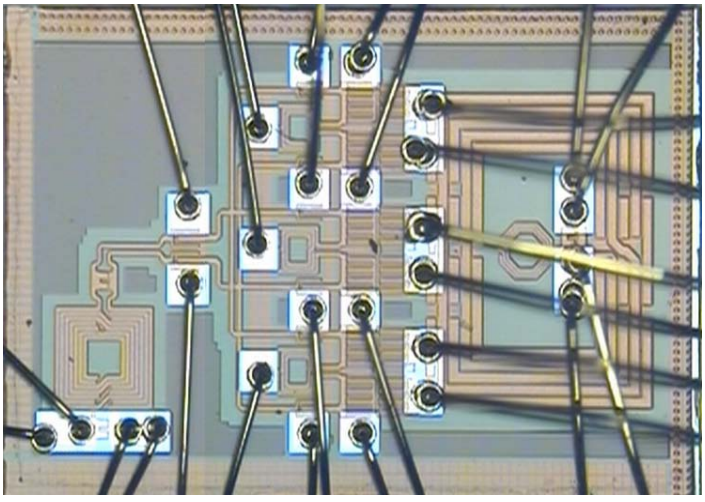
Figure 56 shows the assembled measurement board. The bare chip dies are directly attached to the ground plane of the printed circuit board (PCB) of FR-4 using an electrically conductive silver epoxy. All signal and bias feeding pads are wire bonded to the PCB with gold bond wires. The board loss and the voltage drop of the bias cables are compensated for in the calculation of output power and efficiency while the loss caused by the bond wires is included in the experimental results. No external components are used for the operation.

The measurement test bench in Figure 57 (based on Figure 23) is used for the characterization of the PAs. The bias voltages are set as VDDH = 3.3 V for the power stage and VDDL = 1.8 V for the driver stages while the input and output ports are all matched to 50 Ω. In Figure 58, a frequency sweep of the PAs shows a flat output power

and efficiency throughout the band of interest, which can cover the GSM frequency band. The measurement results in Figure 58(a) exhibit output powers of 31.2 dBm and 32 dBm for the 2×1:2 PCT and 3×1:2 PCT PAs, respectively. Thus, the increase in the output power according to the number of the primary windings is experimentally shown. The overall PAE decreases from 41% to 30%, shown in Figure 58(b), but this decrease results from the mismatch of the 3×1:2 PCT.



(a)



(b)

Figure 54. Microphotographs of the class-E PAs (a) using 2×1:2 PCT and (b) using 3×1:2 PCT

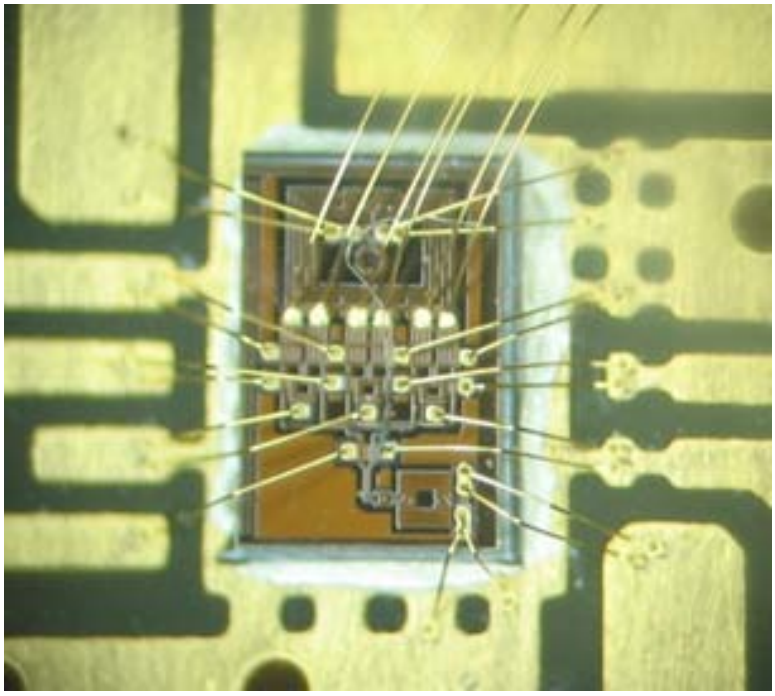


Figure 55. Microphotograph of the board assembly

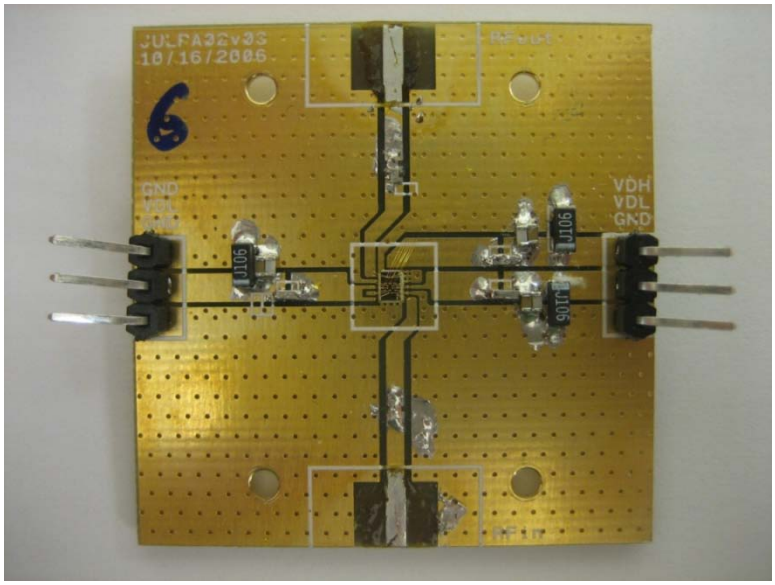


Figure 56. Assembled measurement board (4cm × 4cm)

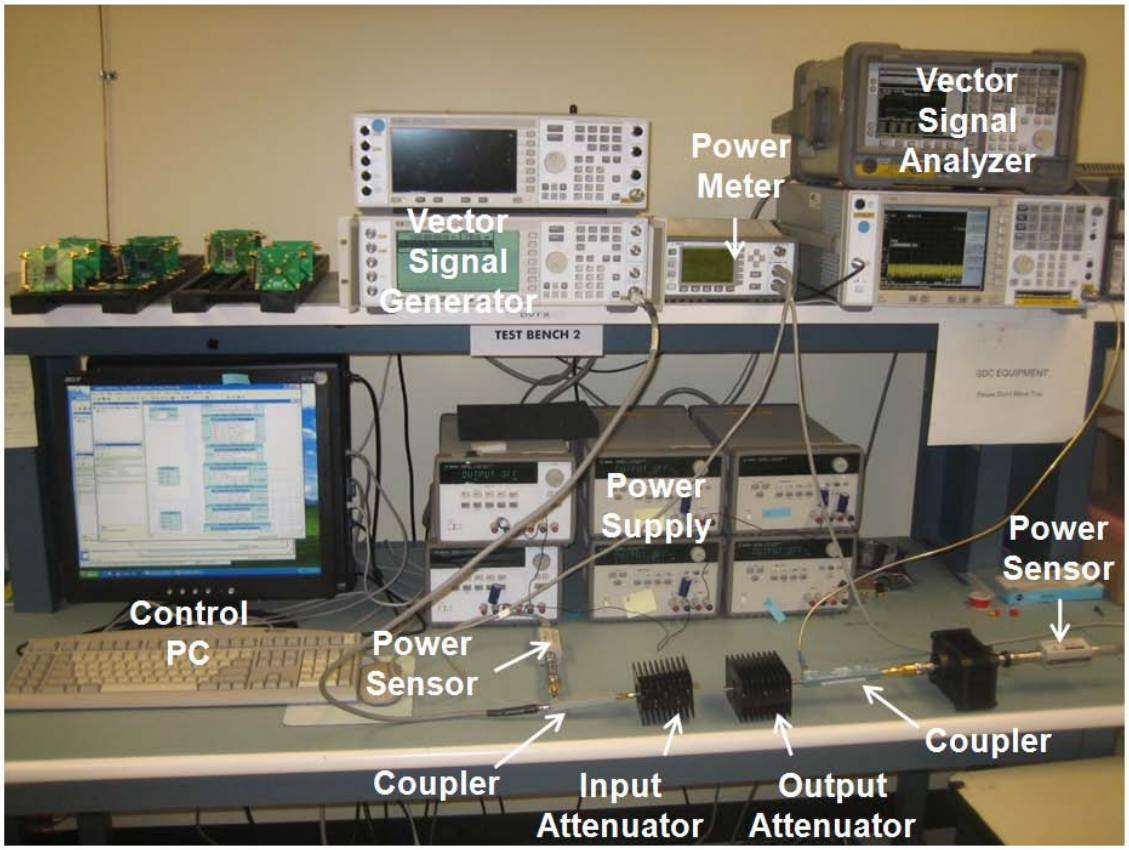


Figure 57. Measurement test bench

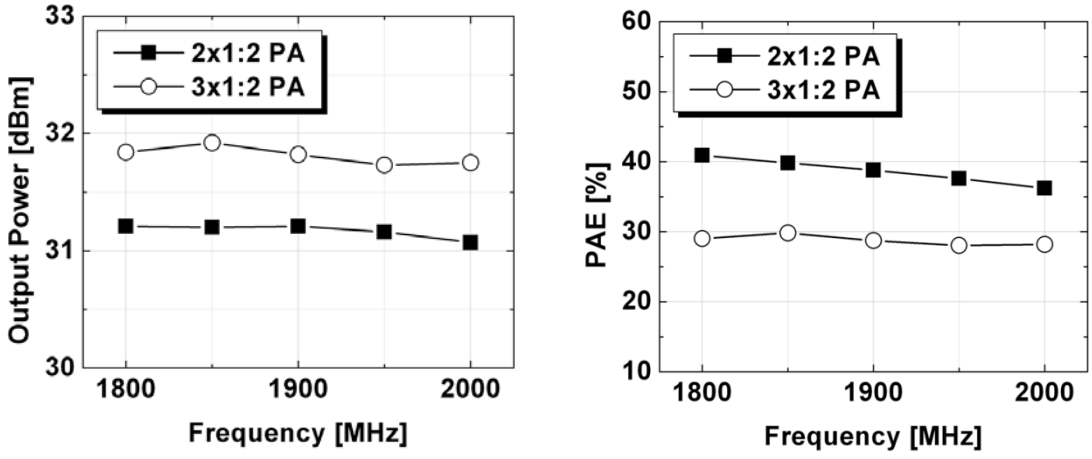


Figure 58. Measured results versus frequency (Input power of 5 dBm) (a) output power and (b) PAE

The measurements and simulations of the PAs are summarized in Table 8. The PAs show rather poor and even harmonic suppression due to an imperfect balance caused by the switching-mode operation.

Table 8. Summary of measured and simulated PA performance

PA	IDDH[1] (A)	IDDL[1] (A)	PAE (%)	Pout (dBm)	2Ho (dBc)	3Ho (dBc)
With 2×1:2 transformer (Measurement)	0.923	0.113	41	31.20	-20.8	-32.7
With 2×1:2 transformer (Simulation)	0.895	0.097	43	31.3	-19.9	-34.6
With 3×1:2 transformer (Measurement)	1.509	0.153	30	31.94	-19.1	-31.0
With 3×1:2 transformer (Simulation)	1.466	0.125	36	32.7	-16.1	-32.7

VDDH = 3.3 V, VDDL = 1.8 V, Input power = 5 dBm, Frequency = 1.8 GHz

[1]IDDH is the current supplied to power stage by VDDH

[2]IDDL is the current supplied to the 1st and 2nd driver stages by VDDL

Finally, the performance of the proposed power combining technique can be compared with state-of-the-art CMOS PAs using the figure of merit (FoM) in [49, 50], properly adopted according to the size of the power-combining structures as follows:

$$FoM_{PA} = P_{OUT} \cdot PAE \cdot f_0^2 \cdot \frac{(\lambda_0/c)^2}{A_{PC}} \tag{4.23}$$

where the key specifications of PAs such as the output power (P_{OUT}), PAE, the square of the operating frequency (f_0^2), and the area of the power-combining structure normalized to the area by the operating wavelength ($A_{PC} \cdot (c/\lambda_0)^2$), are described in the unit of [$W \cdot GHz^2/mm^2$]. The reason for the normalization of the area by the operating wavelength comes from the frequency dependency of the RF and microwave structures. The performance of the linearity of PAs strongly depends on a given standard, so

determining the quantity of the characteristics is difficult. Thus, in this figure of merit, linearity is not considered.

Table 9 shows the results of a comparison of various CMOS PAs with output-combining and matching networks. While various output powers, efficiency, operating frequencies, and areas of matching structures are considered, the proposed structures in this research have the highest figure of merit, proving that the proposed method is the most compact and efficient power-combining method ever reported. The ever-increasing demand for the low-cost PA solution for cellular application requires a very compact output matching network while guaranteeing high output power and efficiency. Thus, this PCT technique can be the most optimum candidate for the market demand.

Table 9. Comparison of studies that dealt with output combining networks for fully-integrated CMOS PAs based on the figure of merit

	Technique	P_{OUT}	PAE	f_0	A_{PC}	FoM_{PA}
This work	PCT (2×1:2)	31.2 dBm	41 %	1.8 GHz	0.70 mm^2	0.772
	PCT (3×1:2)	31.9 dBm	30 %	1.8 GHz	0.70 mm^2	0.664
[16]	DAT (SCT)	33.4 dBm	31 %	2.44 GHz	1.69 mm^2	0.404
[34]	TLT	32 dBm	40 %	1.9 GHz	1 mm^2	0.634
[43]	Figure-8 (SCT)	24.3 dBm	27 %	5.8 GHz	0.2 mm^2	0.363
[50]	LC balun	30.5 dBm	48 %	1.6 GHz	2 mm^2	0.269
[51]	LC balun	23 dBm	29 %	2.45 GHz	2 mm^2	0.029

FoM_{PA} [W·GHz^2/mm^2]

4.4 Fully-Integrated RF Front-End

The ultimate goal of the CMOS RF PA is the full integration of the transceiver. As an initial step toward this goal, a CMOS front-end is introduced in this section. The front-end block includes a PA, a switch, and an LNA, highlighted in Figure 59. Except for the antenna and the front-end, all other blocks are already commercialized based on CMOS technologies.

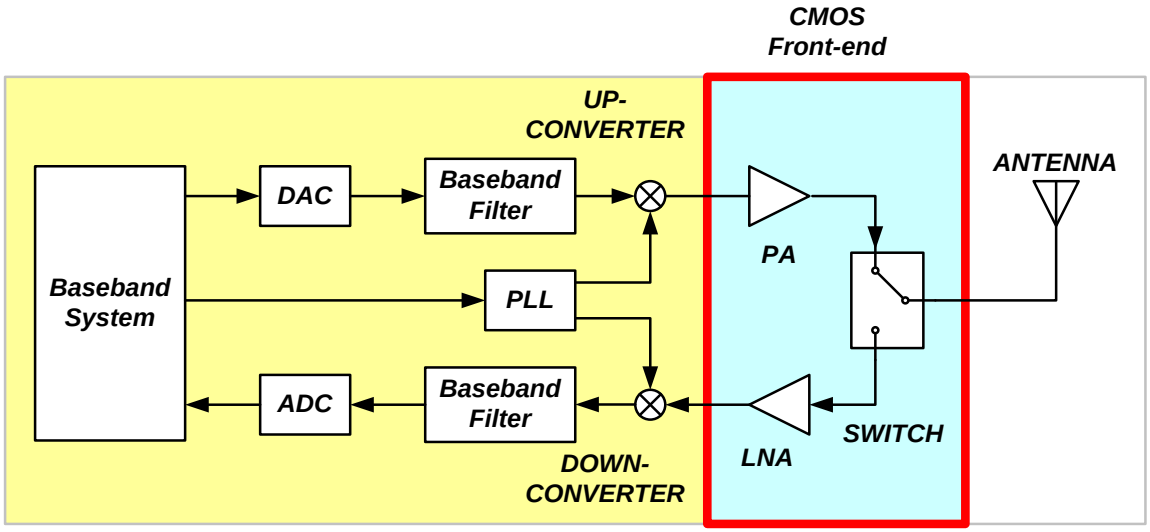


Figure 59. Block diagram of RF front-end in a direct-conversion transceiver

4.4.1 CMOS RF Power Amplifier Design

The design of a CMOS PA follows the procedures described in Section 4.3 of this chapter. Only a slight modification is required in the schematic and layout of the designed PA. Figure 60 shows the schematic diagram of the front-end. The designed PA uses a 2×1:2 PCT for output power combining. The output port of the PA is connected to the transmitting arm of the switch. For the purpose of tuning, a parallel capacitor, used

between the PA and the switch, works as an *LC* notch filter with the bond wire. The expected performance of the PA is plotted in Figure 61 with a 3.3-V bias supply at a 1.8 GHz operation.

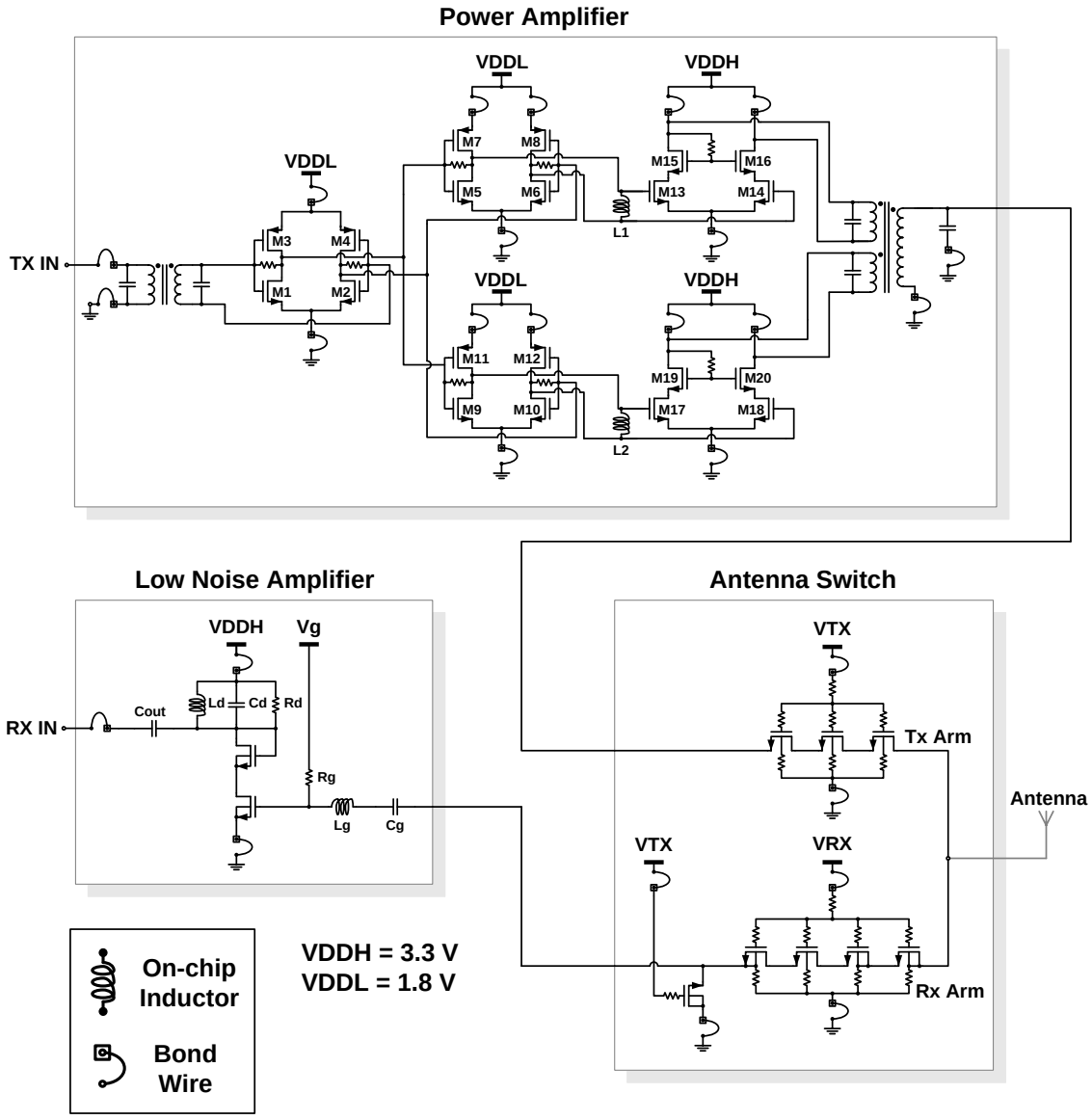


Figure 60. Schematic diagram of the fully-integrated RF front-end

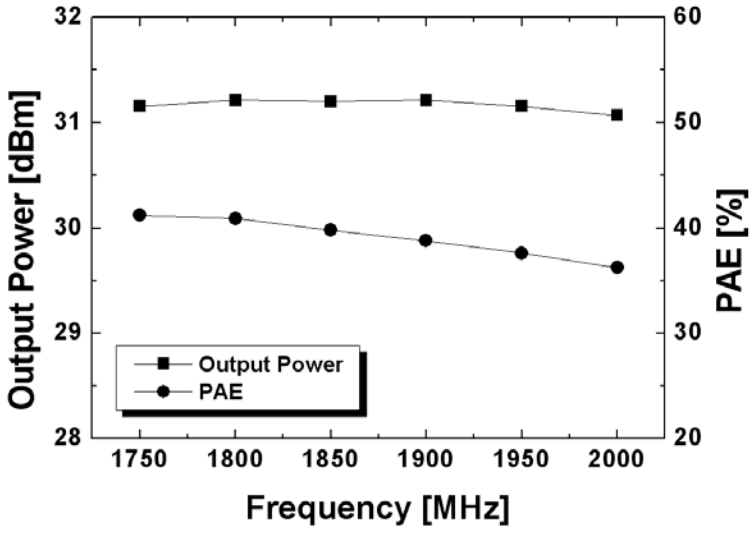


Figure 61. Simulated PA performance in the front-end

4.4.2 CMOS Antenna Switch Design

Since the antenna switch is located right after a high power generating PA, a high power handling capability is a necessity in antenna switch design. The issue of high power handling is even more important when using CMOS technology. Similar to that of the CMOS RF PA, a CMOS switch suffers from the existence of a junction diode, a low breakdown voltage, and a substrate with low resistivity [52]. In this antenna switch design, two techniques are mainly used for high power handling: a resistive body-floating technique that prevents junction diodes from turning on after a high voltage swing [53], and an adaptive voltage distribution method that shares the voltage stress burden among the four stacked transistors in the receiving arm [54]. The measured power handling capability of the switch is 33 dBm input power P0.3dB (input power for 0.3 dB output compression point). The insertion losses of the transmitting and receiving switches are

1.4 dB and 1.8 dB at 1.9 GHz, respectively. The second and third harmonic components are suppressed as much as -50 dBc and -60 dBc, respectively.

4.4.3 Measurement Results

The microphotograph of the fully-integrated RF front-end is shown in Figure 62. The front-end consists of a PA, a switch, and an antenna implemented by 0.18-μm standard CMOS process. The target application of this structure is a GSM application in which the FDD and TDD operation of transmitting and receiving slots can alleviate the burden of isolation. However, the isolation between them is still very important in general. By placing deep trench isolation between the blocks can solve the P-substrate coupling, but it is not provided in this bulk CMOS technology. Instead, a P-substrate guard around each block with a width of 50 μm is used for the basic isolation. Quantification of isolation should be further studied to generalize the isolation techniques of the RF front-end.

Figure 63 shows the measured output power and PAE. The saturated output power is 30 dBm, and the peak PAE is 35% at the antenna port. Figure 64 explains the amount of degradation from the PA to the antenna port. An observed 1-2 dB output drop and around 10% PAE degradation are well expected from the loss of the antenna switch. In addition, the harmonic suppression of the antenna switch can help minimize the leakage of the PA harmonic into the air interface by -50 dBc and -60 dBc harmonic suppressions for the second and third harmonics, respectively. The measured harmonic levels at 2 GHz are shown in Figure 65.

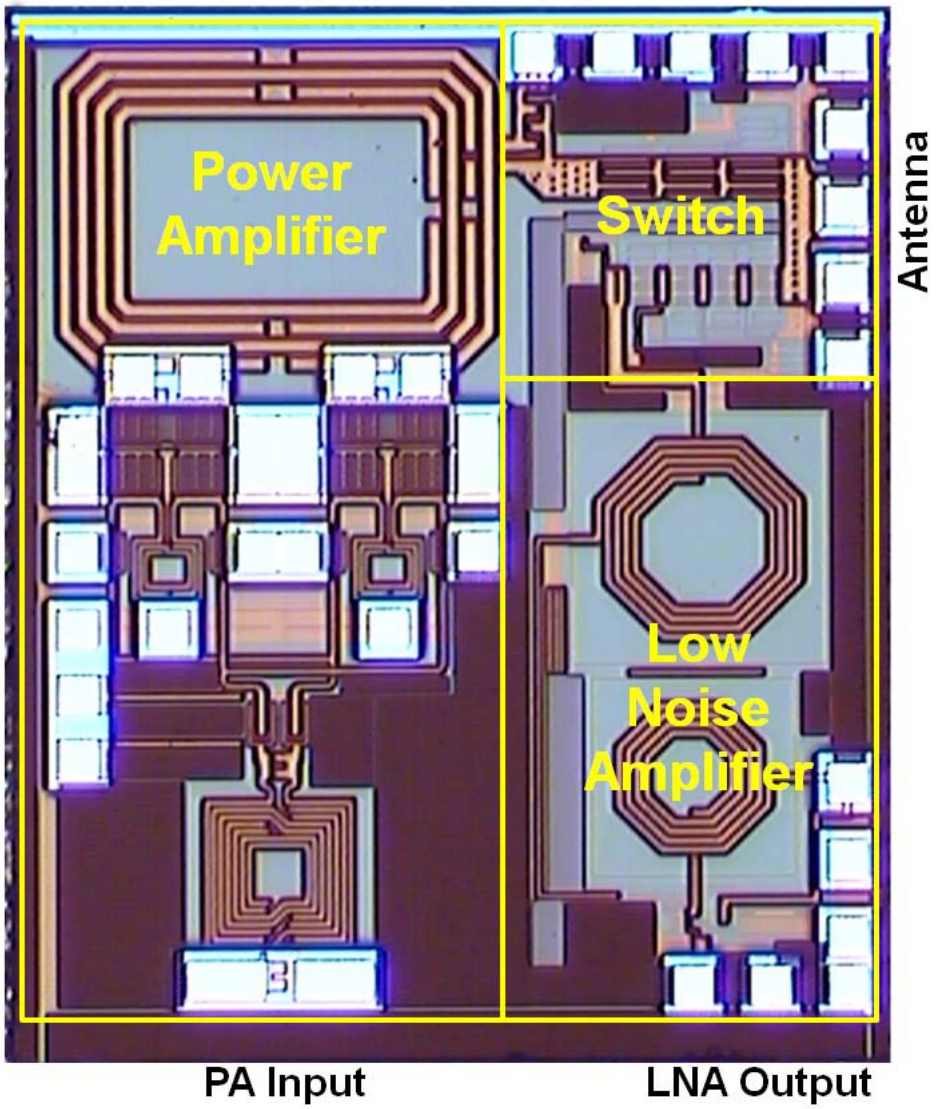


Figure 62. Microphotographs of the fully-integrated RF front-end

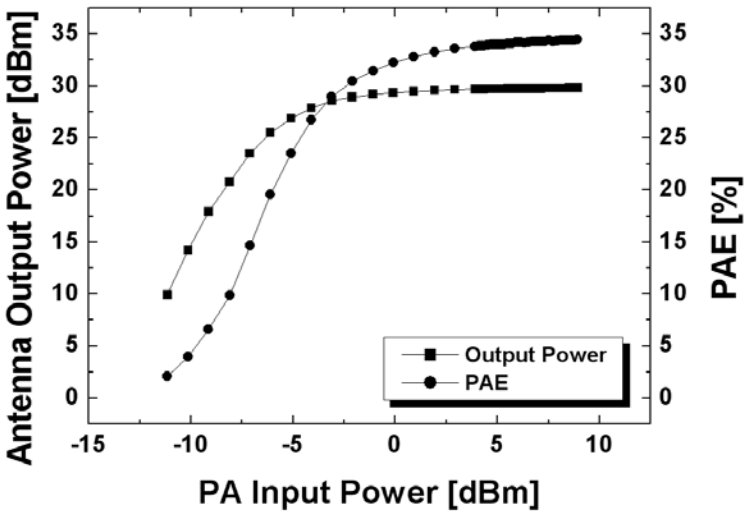


Figure 63. Measured front-end performance vs. PA input power at 2 GHz

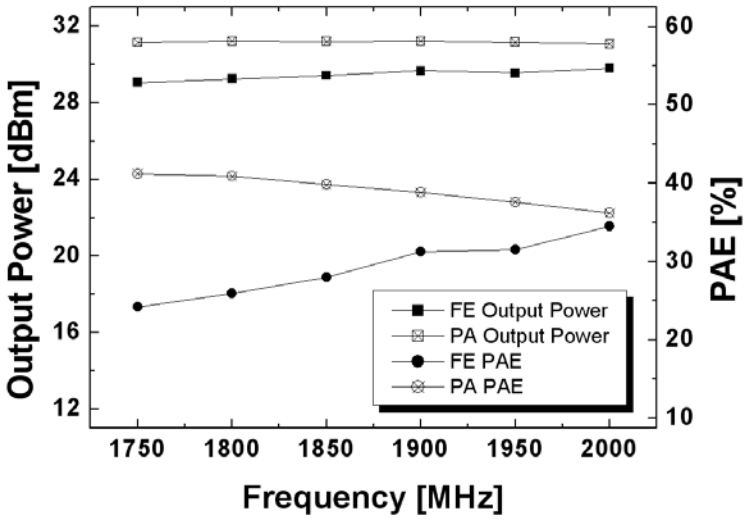


Figure 64. Performance comparison between the simulated PA and measured front-end

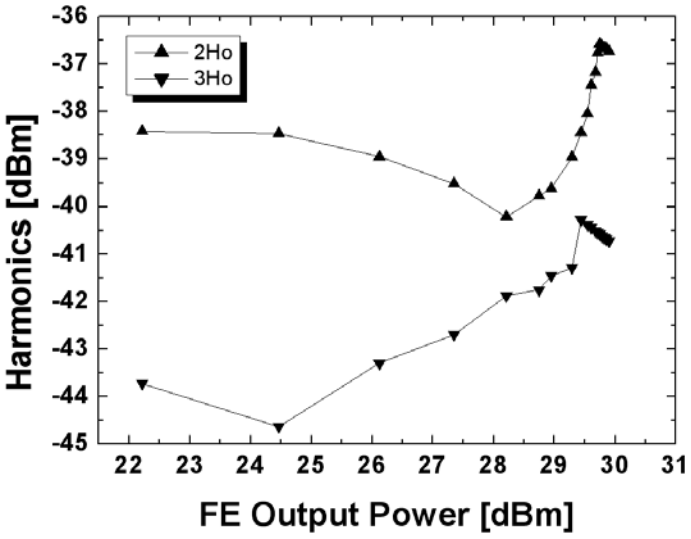


Figure 65. Measured harmonic performance of the front-end at 2 GHz

4.5 Conclusion

This section of the thesis thoroughly analyzes and compares two types of transformer output networks for impedance matching and power combining in CMOS PAs, SCT and PCT, and then introduces the effect of the parasitic resistance of the primary windings and the turn ratios as factors that limit their power-combining capability. The analysis shows that SCTs have a good power-driving capability, but they suffer from susceptibility to parasitic resistance; and PCTs enable efficient power combining with good passive characteristics and reduced output driving. Two kinds of PCTs, the 2×1:2 and 3×1:2 transformers, are adopted for the design of watt-level class-E fully-integrated PAs using a standard 0.18-μm CMOS process. While the PA with the 2×1:2 PCT achieves 31.2-dBm output power with 41% PAE, that with the 3×1:2 PCT delivers 31.9-dBm output power with 30% PAE at 1.8 GHz for a 3.3-V power supply. This analysis then integrates one of the designed PAs with an LNA and an antenna switch to

implement a watt-level RF front-end. The RF front-end can generate 30-dBm output power and provide high harmonic suppression performance of more than 50 dBc for a GSM application at 2 GHz, realizing the first watt-level RF front-end ever reported.

CHAPTER 5
LINEAR POWER AMPLIFIER DESIGN FOR HIGH DATA-RATE APPLICATIONS

5.1 Introduction

The advances in high data-rate modulations with a large PAR have forced PAs to operate at a power back-off below the 1dB compression point. Since the efficiency of general PAs decreases as output power decreases, enhancing the efficiency at the power back-off to minimize the waste of supplied DC power and lengthen battery life in mobile devices can be beneficial.

The design of a linear PA in the back-off area considering the efficiency of the PA necessitates a good understanding of the efficiency characteristic in the back-off area. Section 5.2 of this chapter describes the analytical aspects of back-off efficiency, including the effect of mismatch for linear PAs such as class-A, class-AB, and class-B PAs. In Section 5.3, a fully-integrated linear CMOS PA using a 0.18-μm CMOS process is presented. For power combining and discrete power control, the PCT mentioned in Chapter IV is used [55]. The PA achieves a peak output power of 31 dBm and a peak drain efficiency of 33% at 2.4 GHz. By the discrete power control of the PA, the reduction in current consumption is successfully demonstrated within the EVM requirements of WLAN and WiMAX applications. Finally, conclusive remarks follow in Section 5.4.

5.2 Efficiency at Power Back-off

5.2.1 Efficiencies of Linear Power Amplifiers

For an analysis of back-off efficiency, the typical linear PA such as that illustrated in Figure 66 is provided. The transistor is biased by the RF choke, and the generated voltage signal is a pure sinusoidal by the tank block at the fundamental frequency. By assuming an infinite quality factor of the tank, no harmonic components can be generated, even with a finite $R_{loadline}$. The matching network has no loss such that the efficiency in the back-off area is derived only by operation of the transistor.

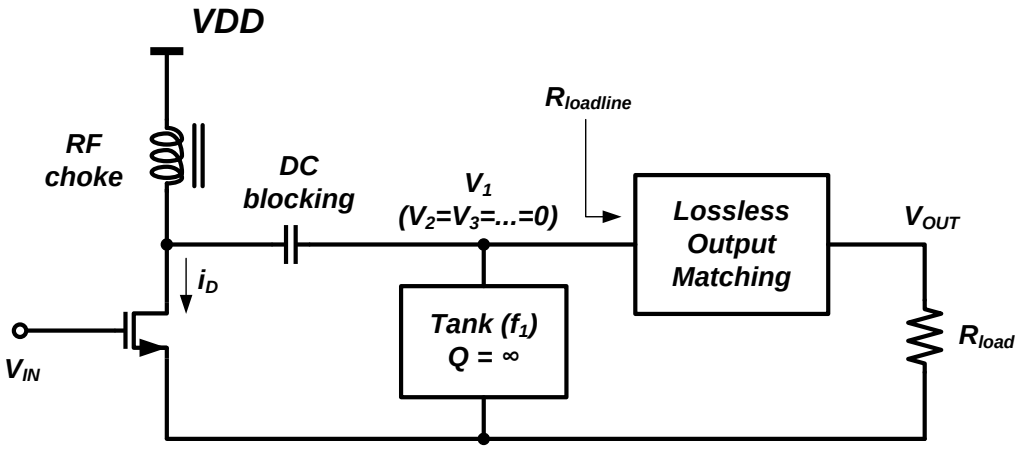


Figure 66. A simplified linear PA topology

Figure 67 shows the loadline analysis of the PA in Figure 66. This PA operates at a reduced conduction angle working as a class-AB PA. In the following analysis, it is also assumed that the knee voltage, V_k, is negligible for the simplicity of the derivation. According to the assumption, the fundamental voltage amplitude, V_1, and the bias point, V_Q, are equal in value. While the voltage swing has only fundamental components, the

current swing is clipped by the limited conduction angle, α, for a chosen loadline so that some harmonic components are generated. For the calculation of the fundamental output power, a spectral analysis of the current is required.

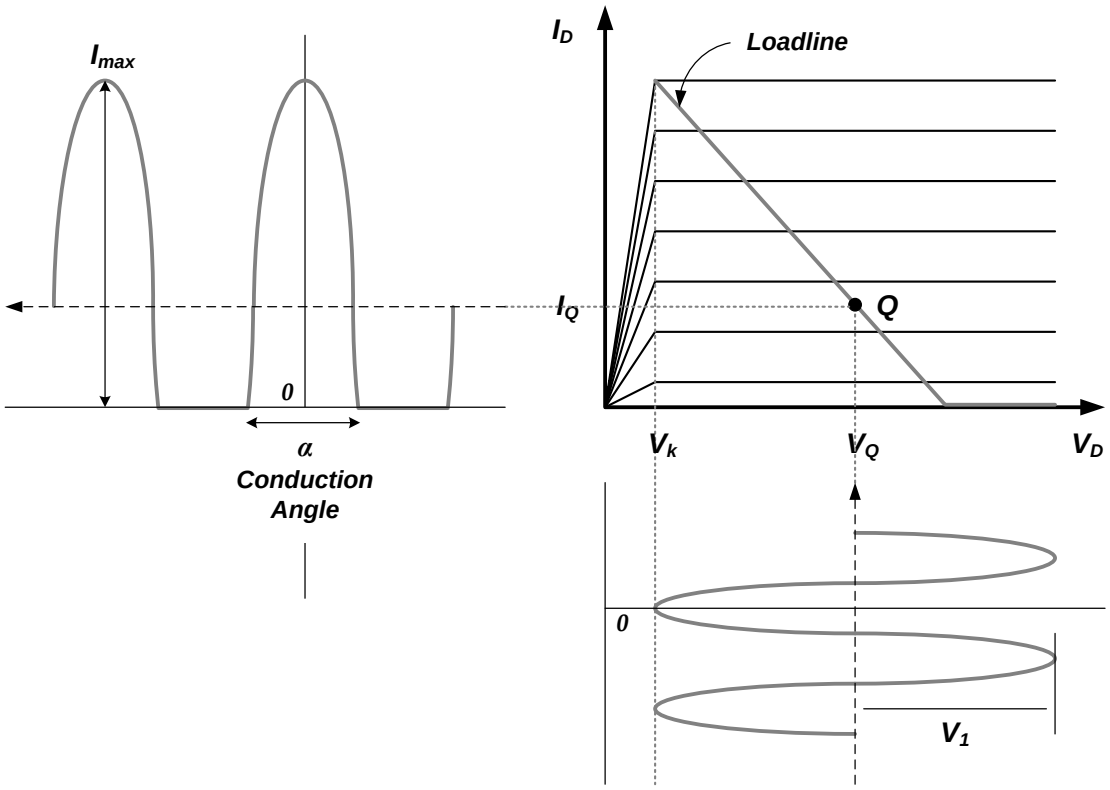


Figure 67. Loadline analysis of a PA at a reduced conduction angle

With Fourier analysis of the given current waveform, current components such as the DC, the fundamental, and harmonics can be generated as follows:

$$i_D(\theta) = I_Q + \left(I_{max} - I_Q\right) \cdot \cos\theta\,,\ -\frac{\alpha}{2} < \theta < \frac{\alpha}{2} \tag{5.1}$$

$$= 0,\ -\pi < \theta < -\frac{\alpha}{2} \text{ and } \frac{\alpha}{2} < \theta < \pi \tag{5.2}$$

From the relation $i_D(\alpha/2) = 0$, Equation (5.1) can be simplified as

$$i_D(\theta) = I_{max} \cdot \frac{\cos\theta - \cos\left(\frac{\alpha}{2}\right)}{1 - \cos\left(\frac{\alpha}{2}\right)},\ -\frac{\alpha}{2} < \theta < \frac{\alpha}{2} \tag{5.3}$$

Applying Fourier series expansion of the even function, Equation (5.3),

$$i_D(\theta) = I_{DC} + \sum_{n=1}^{\infty} I_n \cos(n\theta),\ -\frac{\alpha}{2} < \theta < \frac{\alpha}{2} \tag{5.4}$$

and the required terms are

$$I_{DC} = \frac{1}{2\pi} \cdot \int_{-\alpha/2}^{\alpha/2} I_{max} \cdot \frac{\cos\theta - \cos\left(\frac{\alpha}{2}\right)}{1 - \cos\left(\frac{\alpha}{2}\right)} \cdot d\theta \tag{5.5}$$

$$I_n = \frac{1}{\pi} \cdot \int_{-\alpha/2}^{\alpha/2} I_{max} \cdot \frac{\cos\theta - \cos\left(\frac{\alpha}{2}\right)}{1 - \cos\left(\frac{\alpha}{2}\right)} \cdot \cos(n\theta) \cdot d\theta \tag{5.6}$$

where $n = 1,2,3,$ ..., the positive integer numbers.

The higher-order current components can generally be represented as

$$I_n = \frac{1}{\pi} \cdot \frac{1}{1 - \cos\left(\frac{\alpha}{2}\right)} \cdot \left[\frac{1}{n+1}\sin\left(\frac{(n+1)\alpha}{2}\right) + \frac{1}{n-1}\sin\left(\frac{(n-1)\alpha}{2}\right) - \frac{2}{n}\cos\left(\frac{\alpha}{2}\right)\sin\left(\frac{n\alpha}{2}\right)\right] \tag{5.7}$$

The equation indicates that the higher order degrades the magnitudes of the current. Thus, the second and third harmonic terms have a more serious impact on PA design while the other higher-order terms naturally fade out.

For the fundamental output power and its efficiency, I_{DC} and I_1 can more conveniently be defined in a usable form as Equations (5.8) and (5.9):

$$I_{DC} = \frac{I_{max}}{2\pi} \cdot \frac{2 \cdot \sin\left(\frac{\alpha}{2}\right) - \alpha \cdot \cos\left(\frac{\alpha}{2}\right)}{1 - \cos\left(\frac{\alpha}{2}\right)} \tag{5.8}$$

$$I_1 = \frac{I_{max}}{2\pi} \cdot \frac{\alpha - \sin\alpha}{1 - \cos\left(\frac{\alpha}{2}\right)} \tag{5.9}$$

Figure 68 shows the current waveforms of various classes along the conduction angle. In this research, PAs with a conduction angle can be classified as class A ($\alpha = 2\pi$), shallow class AB ($3\pi/2 \leq \alpha < 2\pi$), deep class AB ($\pi < \alpha \leq 3\pi/2$), class B ($\alpha = \pi$), and class C ($\alpha < \pi$). Because of its high fundamental signals with reduced DC energy, the design of linear PAs focuses on class AB. For increased efficiency, a reduced conduction angle is better, but only in the case of increased second and third harmonic components.

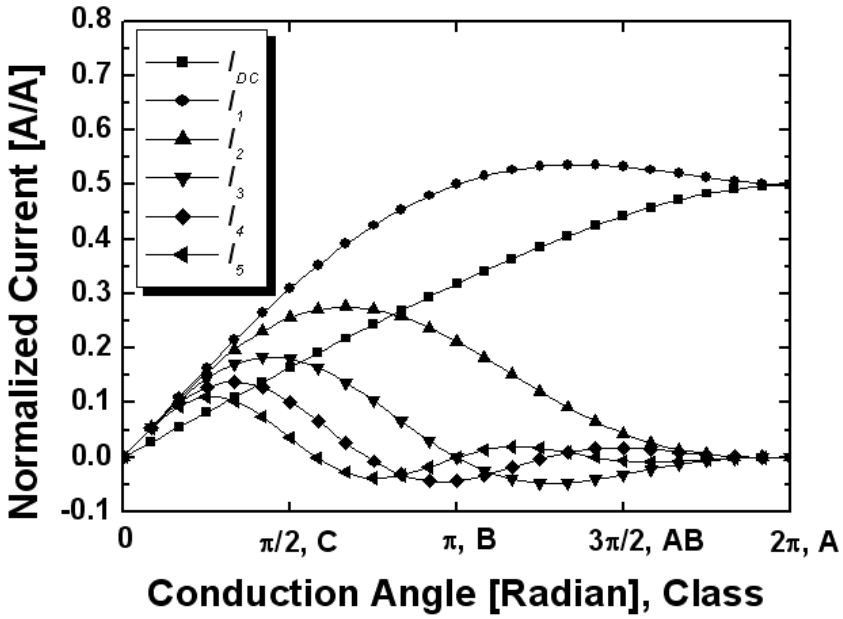


Figure 68. Current waveforms of Fourier components as a function of the conduction angle

Figure 69 shows that the highest output power can be expected for the class-AB operation range, but it has an intermediate efficiency performance based on the following equations

$$\eta = \frac{P_1}{P_{DC}} = \frac{\frac{V_1}{\sqrt{2}} \cdot \frac{I_1}{\sqrt{2}}}{V_Q \cdot I_Q} = \frac{\alpha - \sin\alpha}{2\left[2 \cdot \sin\left(\frac{\alpha}{2}\right) - \alpha \cdot \cos\left(\frac{\alpha}{2}\right)\right]} \tag{5.10}$$

where P_1 is the fundamental output power, P_{DC} is the DC power supplied to the PA, and V_1 is the amplitude of the fundamental signal. For a narrower conduction angle, the power capability deteriorates, seriously marring the narrow conduction angle in PA applications even with very good efficiency characteristics.

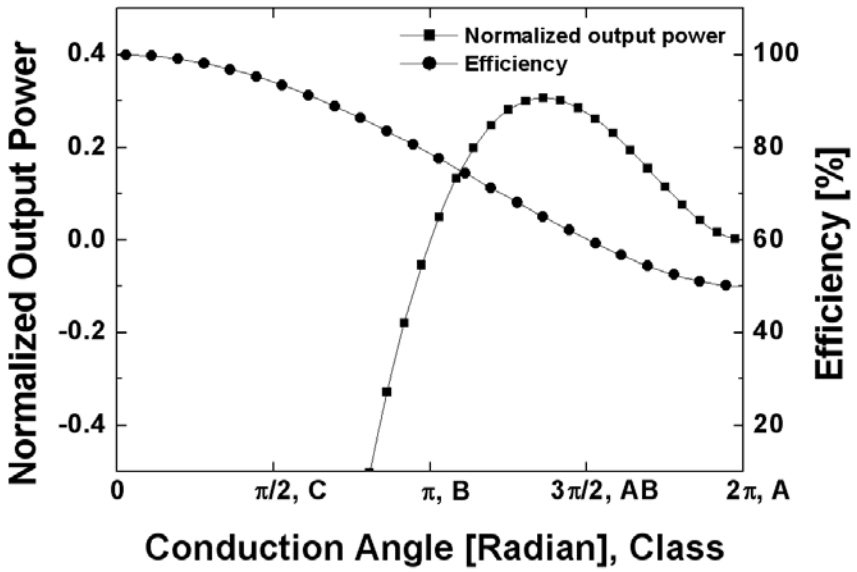


Figure 69. Output power and efficiency capability as a function of the conduction angle

For maximum efficiency characteristics of a PA, it would be beneficial to run the PA in class B or deep class AB for low input power while the PA is operating as shallow

class AB for strong power capability for high input power, as conceptually represented in Figure 70. Such a class shift is also beneficial for linearity enhancement, but at a cost of efficiency in the high output power range.

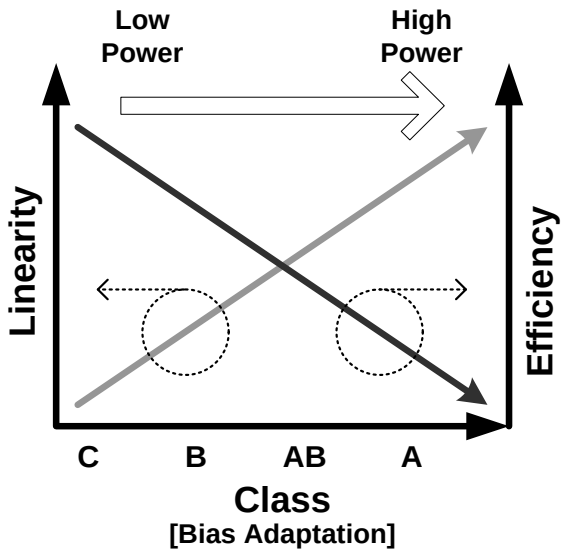


Figure 70. Conceptual class shift by bias adaptation

For optimum matching with a various conduction angle, the optimum load impedance can be derived as follows:

$$R_{OPT} \equiv \frac{V_1}{I_1} = \frac{V_1}{\frac{I_{max}}{2\pi} \cdot \frac{\alpha - \sin\alpha}{1 - \cos\left(\frac{\alpha}{2}\right)}} \tag{5.11}$$

The optimum load impedance is normalized to that of class-A PA, $R_{OPT} = V_1/(I_{max}/2)$, shown as Figure 71. The optimum load is not constant for conduction angles so the adjustment of the loadline along the conduction angle can improve the output power capability and efficiency.

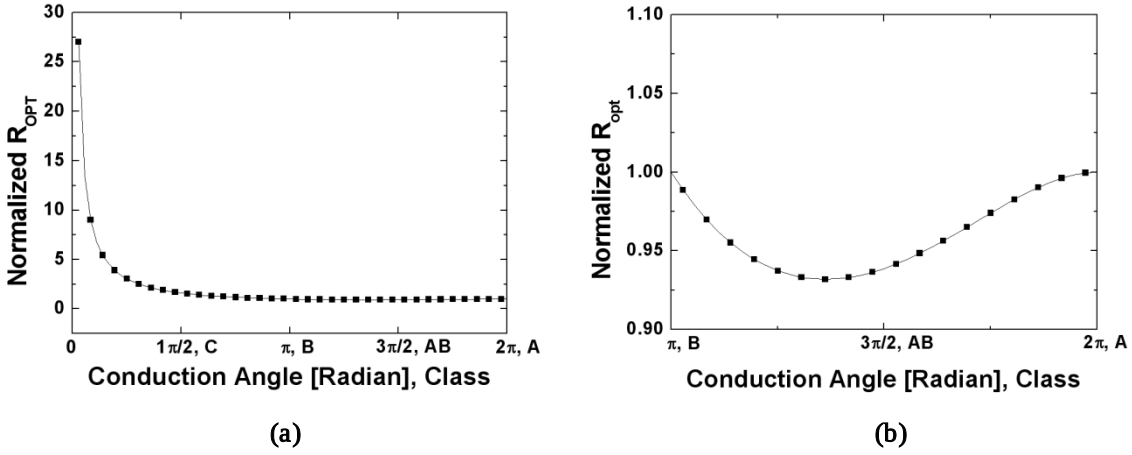


Figure 71. Optimal loadline as a function of the conduction angle (a) the 0~2π range and (b) zoomed view for the π ~2π range

5.2.2 Efficiencies in the Back-off Area

When a back-off of an output power is defined as BO_P in dB, the back-off of the current and voltage magnitude, BO_M, in this analysis can be defined as

$$BO_M(\text{V/V or A/A}) = \sqrt{10^{\frac{BO_P(\text{dB})}{10}}} \tag{5.12}$$

Then, the fundamental signal in the back-off area with the optimum loadline can be represented as

$$I_{1,BO} = \frac{I_1}{BO_M} \tag{5.13}$$

$$V_{1,BO} = I_{1,BO} \cdot R_{OPT} = \frac{V_1}{BO_M} \tag{5.14}$$

From these two relations, the fundamental output power is

$$P_{1,BO} = \frac{V_{1,BO}}{\sqrt{2}} \cdot \frac{I_{1,BO}}{\sqrt{2}} = \frac{V_1}{2 \cdot {BO_M}^2} \cdot \frac{I_{max}}{2\pi} \cdot \frac{\alpha - \sin\alpha}{1 - \cos\left(\frac{\alpha}{2}\right)} \tag{5.15}$$

and the DC power is

$$P_{DC,BO} = V_Q \cdot \frac{I_{DC}}{BO_M} = \frac{V_1}{BO_M} \cdot \frac{I_{max}}{2\pi} \cdot \frac{2 \cdot \sin\left(\frac{\alpha}{2}\right) - \alpha \cdot \cos\left(\frac{\alpha}{2}\right)}{1 - \cos\left(\frac{\alpha}{2}\right)} \tag{5.16}$$

Thus, the efficiency in the back-off area with an optimum impedance condition can be summarized as follows:

$$\eta_{BO} = \frac{P_{1,BO}}{P_{DC,BO}} = \frac{\alpha - \sin\alpha}{2\left[2 \cdot \sin\left(\frac{\alpha}{2}\right) - \alpha \cdot \cos\left(\frac{\alpha}{2}\right)\right]} \cdot \frac{1}{\sqrt{10^{BO_P/10}}} \tag{5.17}$$

and plotted in Figure 72.

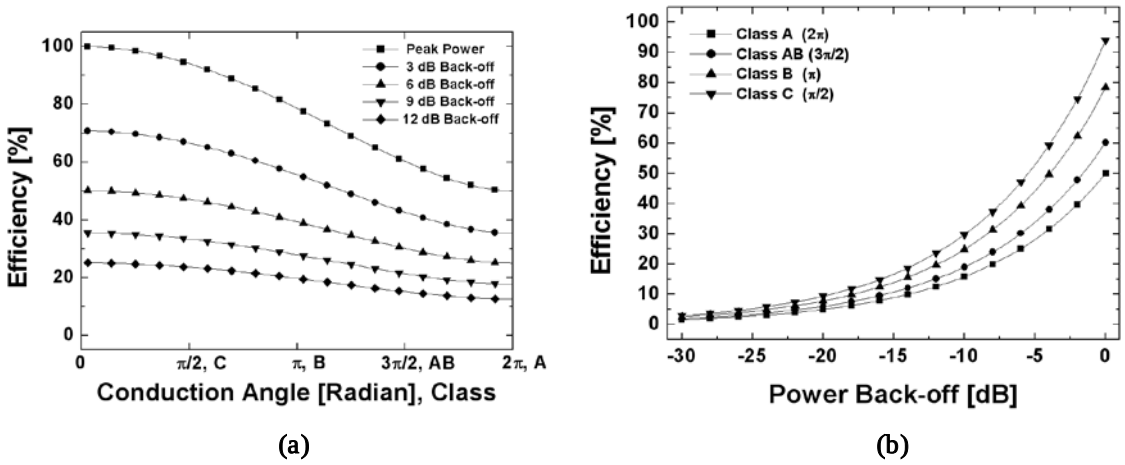


Figure 72. Efficiency in the power back-off (a) efficiency as a function of the conduction angle and (b) efficiency as a function of the power back-off

If the optimum load line condition cannot be satisfied, mismatch between the device and the matching block occurs. The mismatch can be symbolized as a reflection coefficient, assuming that no reactive mismatch elements are present.

$$\Gamma_x = \frac{R_{OPT} - R_{loadline}}{R_{OPT} + R_{loadline}} \tag{5.18}$$

Thus, $R_{loadline}$ can be expressed in terms of R_{OPT} and Γ_x as follows:

$$R_{loadline} = R_{OPT} \cdot \frac{1 - \Gamma_x}{1 + \Gamma_x} \tag{5.19}$$

The fundamental components of the back-off signals with mismatch conditions are

$$I_{1,BOx} = \frac{I_1}{BO_M} \tag{5.20}$$

$$V_{1,BOx} = I_{1,BOx} \cdot R_{loadline} \tag{5.21}$$

and the output power is

$$P_{1,BOx} = \frac{V_{1,BOx}}{\sqrt{2}} \cdot \frac{I_{1,BOx}}{\sqrt{2}} = \frac{{I_1}^2}{2 \cdot {BO_M}^2} \cdot R_{loadline} \tag{5.22}$$

The efficiency for this condition is derived as follows:

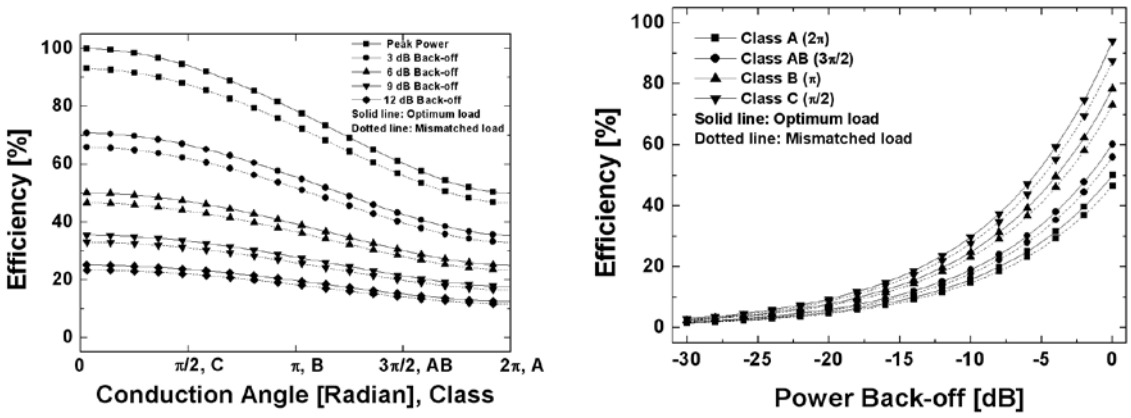
$$\eta_{BOx} = \frac{P_{1,BOx}}{P_{DC,BOx}} = \frac{\alpha - \sin\alpha}{2\left[2 \cdot \sin\left(\frac{\alpha}{2}\right) - \alpha \cdot \cos\left(\frac{\alpha}{2}\right)\right]} \cdot \frac{1}{\sqrt{10^{BO_P/10}}} \cdot \frac{1 - \Gamma_x}{1 + \Gamma_x} \tag{5.23}$$

$$= \eta_{BO} \cdot \frac{1 - \Gamma_x}{1 + \Gamma_x} \tag{5.24}$$

$$= \eta \cdot \frac{1}{BO_M} \cdot \frac{1 - \Gamma_x}{1 + \Gamma_x} \tag{5.25}$$

$$= \text{peak efficiency} \cdot \text{power back-off} \cdot \text{load mismatch} \tag{5.26}$$

The mismatched load can degrade the efficiency performance following Equation (5.25). For example, if a PA is originally matched to a class-A condition and the load is mismatched by running the PA in the class-AB condition, the optimum loadline would deviate from the original optimum loadline by 7%, as shown in Figure 71(b). Thus, $\Gamma_x = 0.036$ is the measure of the reflection caused by the mismatch. The results in Figure 72 are then adjusted by the mismatch condition plotted in Figure 73. While efficiency degradation is relatively insignificant, it is often critical when meeting a target specification.



(a) (b)

Figure 73. Efficiency degradation by a mismatched load (Γ_x=0.036) at the power back-off
(a) efficiency as a function of the conduction angle and
(b) efficiency as a function of the power back-off

From Equation (5.25), it can be derived that the back-off efficiency with mismatch is a direct multiplication of the peak efficiency, η, the inverse of the magnitude back-off, BO_M, and the voltage standing wave ratio at the matching node, $(1-\Gamma_x)/(1+\Gamma_x)$. Usually, BO_M is determined by the signal characteristics, and designers should focus on maximizing peak efficiency and minimizing the mismatch at the matching node for a maximum efficiency.

Thus, for a linear PA design, class-AB PA is a reasonable choice for high output power and reasonably high efficiency. Furthermore, a mismatch condition of a matching network can cause an additional efficiency drop such that the bias condition of the PA should be shifted to a deep class-AB area further than what is expected for a perfect matching condition.

5.3 Linear Power Amplifier Design Using Discrete Power Control

5.3.1 Discrete Power Control of Parallel Amplification

Studies pertaining to the topologies of PAs combined in parallel have been conducted for the purpose of increasing efficiency in low output power range [36, 51, 56]. As shown in Figure 74, the unit PAs are turned on independently and the turned-off PAs are isolated from the output at the power-combining block. In [56], a PMOS AC ground path and a quarter-wave transmission line are used for matching and isolating each unit PA. In [51], *LC* baluns are effectively exploited for matching and power combining. However, both schemes require bulky and complicated structures that consume large die areas. If a transformer is used instead, impedance-matching and power-combining functions can be concurrently incorporated in the transformer. Transformers in series for parallel amplification and discrete power control are shown in [36].

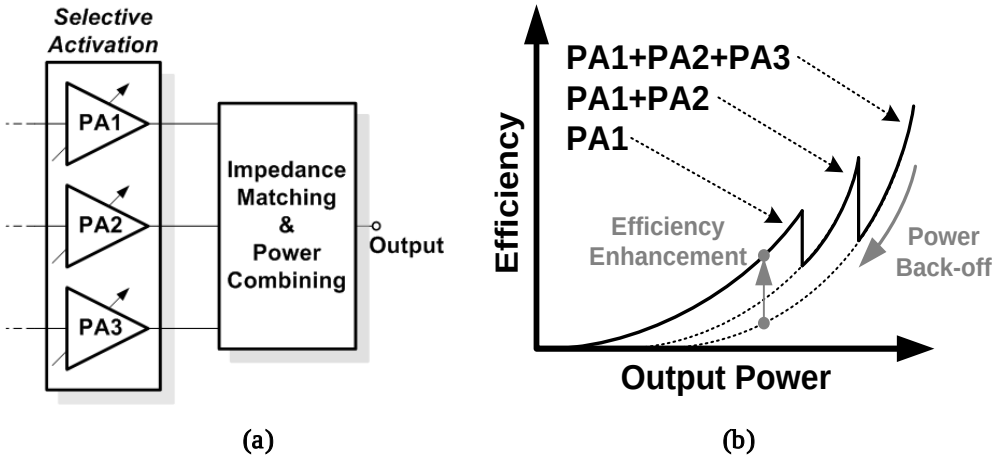


Figure 74. Design concept of the proposed PA: (a) block diagram of the PA and (b) efficiency enhancement of the PA

For the same purpose, transformers in parallel can also be used to condense the layout of the transformers as shown in Figure 75. The primary and secondary pairs are magnetically coupled and physically detached to isolate the turned-off unit PAs while also performing impedance matching and power combining [55]. In addition to the reduction in the current consumption by controlling the number of turned-on unit PAs, the bias adaptation of V_{G1} in Figure 75 can also help the current reduction by shifting the operating class of the PA from shallow class AB to deep class AB until the quiescent current of the PA is minimized.

5.3.2 Linear Power Amplifier Design

The overall schematic diagram of the proposed PA is shown in Figure 75. The PA consists of the input balun, the driver stage, interstage matching, the power stage, and PCT. The two-stage topology is adopted to provide enough gain. To minimize ground bouncing and the bond wire effect, a differential topology is used. In addition, the driver and power stages are designed in a cascode topology so that excessive voltage stress on CMOS devices, which are prone to breakdown under high voltages [27], can be prevented. Each power stage is a stack of 0.35 μm thick-oxide transistors for the common-gate (CG) stages, M2, and 0.18 μm thin-oxide transistors for common-source (CS) stages, M1, distributing the voltage stress [55]. The widths of each CG and CS stages are 4096 μm (8 × 64 fingers × 8 μm width) and 3072 μm (24 × 16 fingers × 8 μm width), respectively. Both the driver and power stages are biased in class AB for linearity and good efficiency. According to the concept of discrete power control, three identical unit PAs are combined in parallel using a PCT [55]. Each PA unit is independently

controlled by the ports for the gate of the CS and CG stages, V_{G1} and V_{G2}, respectively. The PCT is composed of three primary windings and one secondary winding all with 1:2 turn ratios. The windings are 30 μm wide and 2.34 μm thick, and the gap between them is 5 μm. They are interwoven as shown in the layout in Figure 75. The total transformer size is 900 μm × 600 μm. When it is excited differentially, the simulated insertion loss is 1.05 dB (transformer efficiency of 78.5%) at 2.4 GHz. For optimization, the capacitors C_{IN} (2.6 pF) and C_{OUT} (1.4 pF) are tuned for the primary and secondary windings, respectively.

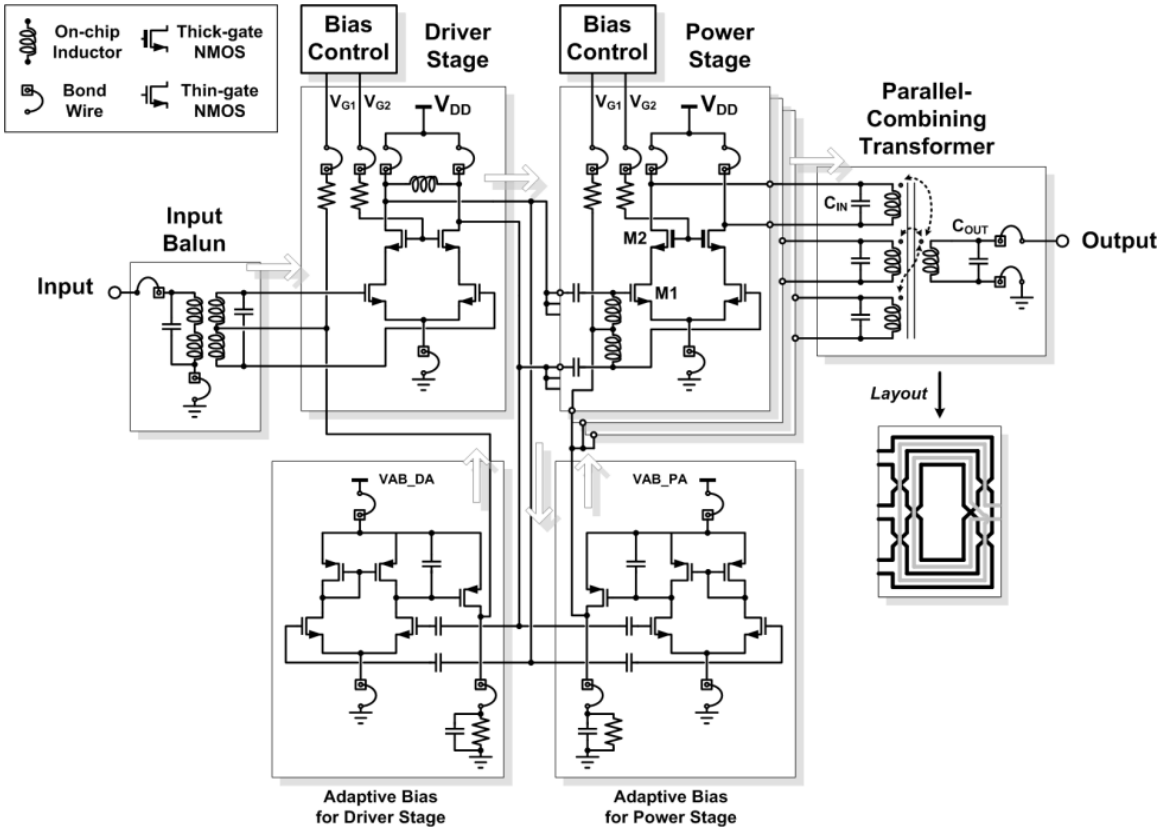


Figure 75. Circuit schematic of the proposed PA

5.3.3 Measurement Results

A photograph of the implemented PA using a standard 0.18 μm CMOS process is shown in Figure 76. The die area, including the wire-bonding pads, is 1200 μm × 1650 μm. For an accurate measurement, the die is attached onto a PCB board with 50-Ohm input and output terminations. The PCB and cable losses are de-embedded while the loss of the bond wires is included in the performance.

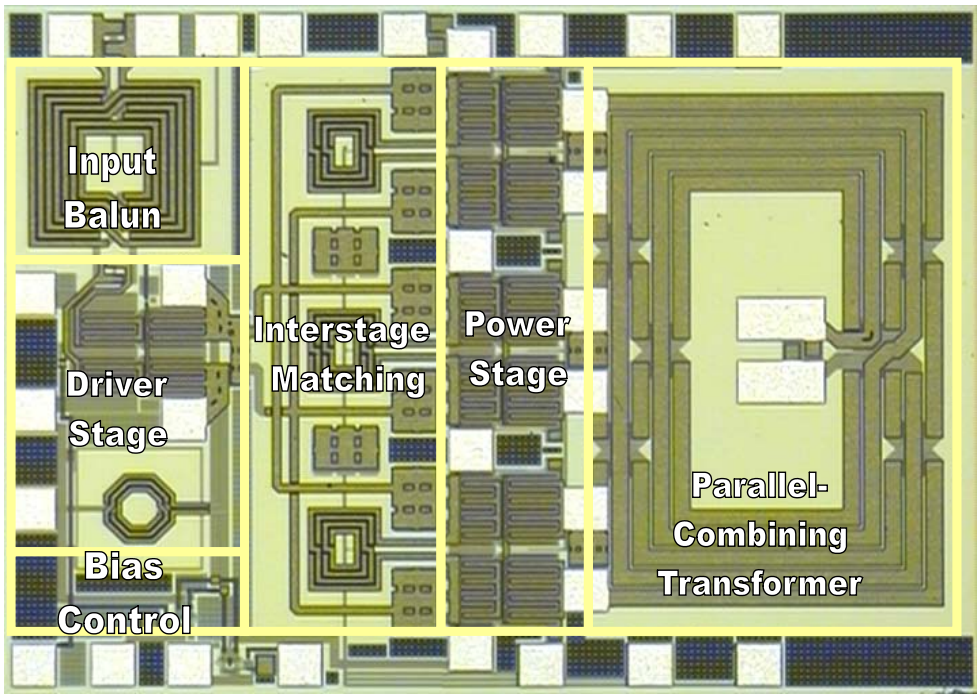


Figure 76. Layout of the proposed PA.

With VDD = 3.3 V, VG2 = 2.4 V, and varying VG1 as 0.6 V (PA1 only), 0.65 V (PA1+PA2), and 0.7 V (PA1+PA2+PA3) by external power supplies, an example of efficiency enhancement can be plotted as shown in Figure 77. The efficiency improvement in the back-off range can be observed between the 10 dBm and 25 dBm

output powers. The peak output power is 31 dBm, and the drain efficiency of the power stage is 33% (with an overall PAE of 27%). If the operation mode is changed, the gain can vary in steps, as indicated in Figure 77, but the gain is still more than 20 dB even for the PA1 mode, so the effect on efficiency is insignificant.

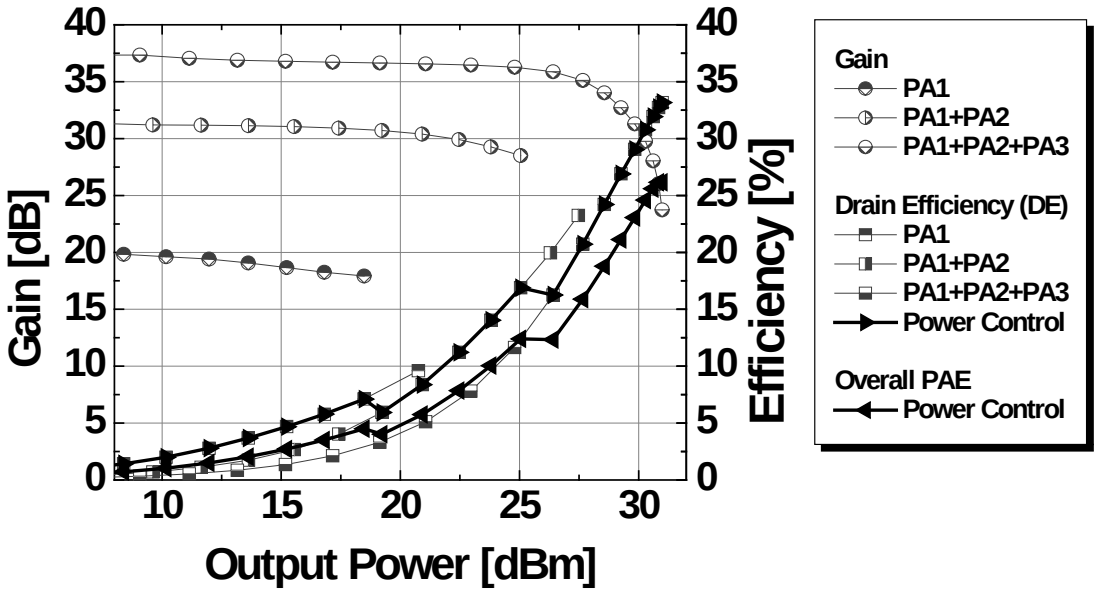


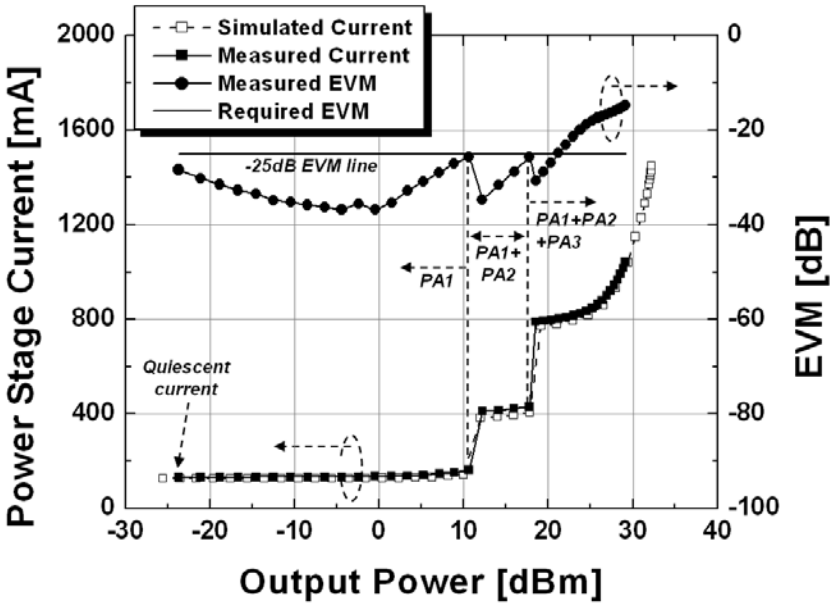
Figure 77. Gain and efficiency variation according to discrete power control

The linearity performance of the PA depends on the requirements of the input modulation signals. The performance for digital modulations was evaluated by applying WLAN 802.11g 54Mbps 64QAM OFDM (EVM limit < –25 dB) and WiMAX 802.16e 54Mbps 64QAM OFDM (EVM limit < –31 dB) signals at 2.4 GHz, and discrete power control is executed according to the linearity specifications. Initially, only one unit PA, which guarantees the minimum quiescent current of 130 mA and the minimum linearity requirement, is turned on. Compared to the quiescent current 780 mA of the fully turned-on PA, one unit PA can save more than 650 mA of the current consumption in the low

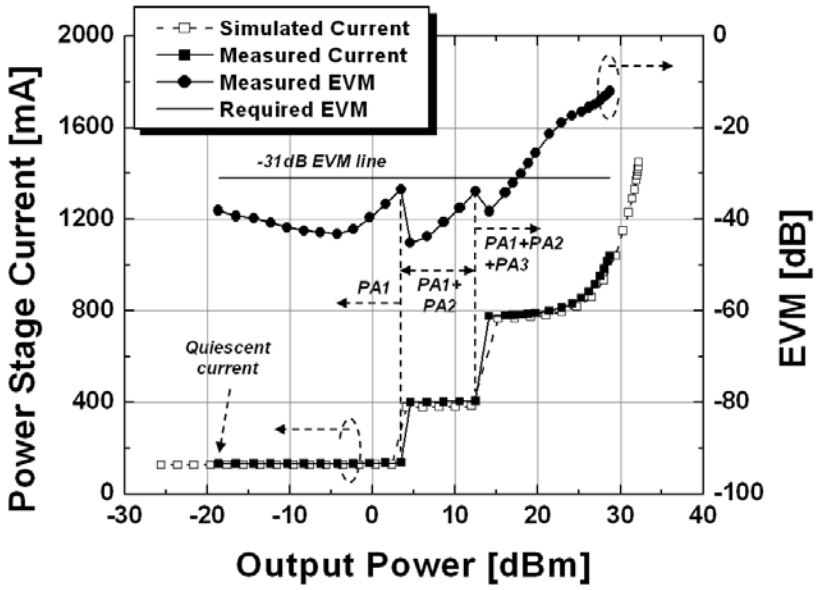
output power range. When the PA that is first turned on reaches the point at which the linearity requirement is violated with the input increase, the second unit PA is turned on. In this manner, the low quiescent current can be maintained in the low power range while the linearity requirement is satisfied in the high power range by supplying more current. In accordance with the power control, the EVM tendency, exhibiting a saw-like shape, does not exceed the limit governed by the requirements.

The maximum linear output power is 22 dBm for the WLAN and 18 dBm for the WiMAX signals, shown in Figure 78(a) and (b). The stringent linearity requirement of the WiMAX signal forces the second unit PA to turn on in the lower output power range of close to 4 dBm while the WLAN signal extends the turn-on point of the second unit PA to 10 dBm output power. For maximum current reduction at the power back-off, the requirements of linearity should be a criterion for discrete power control.

The measured performance results of the fully-integrated CMOS PAs with discrete power control agree well with the simulation results. In Table 10, the CMOS PAs in this study are compared with other reported fully-integrated CMOS PAs both with parallel amplification, [36] and [51], and without it [13]. This work achieves comparable efficiency performance while generating the highest output power.



(a)



(b)

Figure 78. Discrete power control with EVM (2.4 GHz) (a) the 802.11g WLAN 54 Mbps 64 QAM OFDM signal and (b) the 802.16e WiMAX 54 Mbps 64 QAM OFDM signal

Table 10. Performance summary and comparison with other fully-integrated CMOS PAs

Reference	CMOS Tech. [μm]	Frequency [GHz]	V_{DD} [V]	P_{1dB}/P_{SAT} [dBm]	DE [†] (-6dB/ P_{SAT})[%]	Size [mm^2]	PA Class	Application
This work	0.18	2.4	3.3	27/31	10/33	2.0	AB	WLAN, WiMAX
	No mode control				5/33	--		
[13]	0.18	2.4	3.3	--/24.5	12/31 [‡]	1.7	AB	WLAN
[36]	0.13	2.4	1.2	--/27	--/32	2.0	AB	WLAN
[51]	0.13	2.45	1.5	--/23	32/42	5.5	BE	Bluetooth

[†] Drain efficiency, [‡] PAE

5.4 Conclusion

This chapter presents a fully-integrated linear PA in a 0.18 μm CMOS process with discrete power control. By using a transformer with multiple primaries in parallel coupled with a single secondary and an adaptive gate bias control scheme, it efficiently implements the discrete power control of the PA. The efficiency at power back-off is enhanced by reducing the current consumption by more than 650 mA. A maximum output power of 31 dBm and a peak drain efficiency of 33% are observed with a 3.3 V power supply at 2.4 GHz. Maximum linear output power can be achieved of up to 22 dBm and 18 dBm for WLAN 802.11g and WiMAX 802.16e, respectively.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1. Technical Contributions and Achievements

Over the past several years, we have observed improvements in CMOS RF PAs in both academic and wireless market arenas. Although the profits garnered by CMOS RF PAs are still insignificant when compared to those derived from compound semiconductors, the products of powerhouses in the market, the possibilities for integration and low-cost advantages are inducing more and more design efforts in this field.

For the purpose of developing CMOS RF PAs in wireless mobile communications, this research shows both the theoretical contributions and the successful implementation of fully-integrated CMOS RF PAs and a fully-integrated CMOS RF front-end. The achievements can be summarized as follows:

- A highly efficient and compact power-combining method has been developed. The power-combining mechanism was successfully analyzed based on simple transformer models, and the dominant loss factor of the structure, specifically the primary winding resistances of transformers, was also presented. The theoretical backgrounds of both series- and parallel-combining methods were compared and some design parameters for required output power were suggested as guidelines.

- A successful implementation of two fully-integrated switching PAs for a GSM application was demonstrated using a 0.18-μm CMOS process, yielding PAs with 31.2 dBm output power with 41% PAE, and 31.9 dBm output power with 30% PAE. Through the measured PAs, the feasibility of the suggested power-combining method was experimentally proven. The technique, the most compact method ever reported with regard to output power, efficiency, and the form factor of output matching and the power-combining structure, achieves the best figure-of-merit among all reported CMOS RF PAs.
- The first fully-integrated CMOS RF front-end ever reported for a watt-level application was realized using a 0.18-μm CMOS process. The front-end consists of an RF PA, an RF switch, and an LNA. The front-end can generate 30 dBm output power and good harmonic suppression of more than 50 dBc at 2 GHz operation.
- Comprehensive efficiency characteristics in linear PAs, including the conduction angle, power back-off, and load mismatch, have been analyzed. Throughout this analysis, the dominant efficiency degradation caused by power back-off and the allowable ranges of mismatch were presented. This analysis forms the basis of efficiency enhancement techniques in the power back-off area.
- Fully-integrated linear PAs for WLAN and WiMAX applications were successfully implemented using a 0.18-μm CMOS process. This PA achieved the highest output power, 31 dBm, for 2.4 GHz applications reported, and the low-power current saving of more than 650 mA was successfully demonstrated while achieving the linearity specifications of WLAN and WiMAX EVM below -25 dB and -31 dB, respectively.

6.2. Future Research Directions

The design techniques of CMOS RF PAs for wireless communications are not yet fully mature. Several issues such as low power, efficiency, and linearity of CMOS RF PAs compared to those of their counterparts, compound semiconductor PAs, must be addressed. To overcome these problems, researchers in both academia and industry must devote considerable effort. Moreover, such efforts should create the impetus for constructive endeavors that respond to the requirements of future communication needs.

The first challenge in the development of a future RF PA is to ensure a sufficient supply of power. The current power supply of mobile terminals is approximatley 3 V. However, the change in power management techniques will require a lower supply voltage for all components with no exception allowed for PAs. Since wireless communication standards require a certain amount of power for each application, a lower supply leads to a higher current driving capability to obtain the same output power as that in the original supply. While the breakdown issue may be alleviated, large power cells can increase parasitic elements and lower efficiency due to the limited voltage swings.

Second, for high-data rate communications, highly-linear PAs will be welcomed in the future. Such a trend can already be observed for WLAN, WiMAX, and LTE standards, in which OFDM signals with high PAR are used. Thus, PAs should show high linearity performance with high efficiency. For a more thorough understanding of nonlinearity characteristics, rigorous analysis of nonlinearity in CMOS devices should be carried out. In addition, linearity and efficiency enhancement techniques must be in a compact form so that they can be easily adapted to compact mobile terminals.

Moreover, for easy roaming service, multi-mode and multi-band standards should be covered by either one PA with adaptive matching techniques or an array of PAs. Both cases necessitate near-perfect switching techniques in bulk CMOS technology [57].

Finally, the integration of a PA with a transceiver that contains all functions must be realized in cellular applications. Typically, the higher output power of cellular applications requires better isolation techniques, particularly for receiving front-end and other noise-susceptive blocks.

On a final note, with CMOS technology fostered by the demands of highly-linear performance, the PA industry will achieve unparalleled sophistication. Thus, despite all the changes and demands, PA designers, rather than limiting their knowledge and understanding to the very basic RF area, should expand their sights to the analog and digital areas in an effort to redefine the design of RF PAs as one of a transmitting system.

PUBLICATIONS

[1] **K. H. An**, D. H. Lee, O. Lee, H. Kim, J. Han, W. Kim, C.-H. Lee, H. Kim, and J. Laskar, "A 2.4 GHz fully integrated linear CMOS power amplifier with discrete power control," *IEEE Microw. Wireless Compon. Lett.*, vol. 19, no. 7, Jul. 2009.

[2] **K. H. An**, O. Lee, H. Kim, D. H. Lee, J. Han, K. S. Yang, Y. Kim, J. J. Chang, W. Woo, C.-H. Lee, H. Kim, and J. Laskar, "Power-Combining Transformer Techniques for Fully-Integrated CMOS Power Amplifiers," *IEEE J. Solid-State Circuits*, vol. 43, pp. 1064-1075, May 2008. **(*IEEE top 100 documents accessed #55 : May 2008*)**

[3] **K. H. An**, Y. Kim, O. Lee, K. S. Yang, H. Kim, W. Woo, J. J. Chang, C.-H. Lee, H. Kim, and J. Laskar, "A Monolithic Voltage-Boosting Parallel-Primary Transformer Structures for Fully Integrated CMOS Power Amplifier Design," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symp. Dig. Papers*, 2007, pp. 419-422.

[4] D. H. Lee, **K. H. An**, O. Lee, H. Kim, C.-H. Lee, and J. Laskar, “A partially dynamic biased power amplifier for WCDMA applications,” in review with *IEEE Trans. Microw. Theory Tech.*

[5] O. Lee, J. Han, **K. H. An**, K. S. Lee, C.-H. Lee, and J. Laskar, “A charging acceleration technique for highly efficient CMOS cascade class-E power amplifiers,” in review with *IEEE J. Solid-State Circuits*.

[6] O. Lee, **K. H. An**, H. Kim, D. H. Lee, J. Han, K. S. Yang, C.-H. Lee, H. Kim, and J. Laskar, "Analysis and design of fully integrated high power parallel-circuit class-E CMOS power amplifiers," *IEEE Trans. Circuits Systems,* Accepted 2009.

[7] C.-H. Lee, J. J. Chang, K. S. Yang, **K. H. An**, I. Lee, K. Kim, J. Nam, Y. Kim, and H. Kim, "A highly efficient GSM/GPRS quad-band CMOS PA module," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symp. Dig. Papers*, 2009, pp. 229-232.

[9] M. Ahn, **K. H. An**, C.-H. Lee, and J. Laskar, "Fully integrated RF front-end circuits in 2 GHz using 0.18 um standard CMOS process," in *APMC. Dig. Papers*, 2008. ***(Best paper award)***

[9] O. Lee, K. S. Yang, **K. H. An**, Y. Kim, H. Kim, J. J. Chang, W. Woo, C.-H. Lee, and J. Laskar, "A 1.8-GHz 2-Watt Fully Integrated CMOS Push-Pull Parallel-Combined Power Amplifier Design," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symp. Dig. Papers*, 2007, pp. 435-438.

[10] W. Woo, **K. H. An**, O. Lee, J. J. Chang, C.-H. Lee, K. Yang, M. J. Park, H. Kim, and J. Laskar, "A novel linear polar transmitter architecture using low-power analog predistortion for EDGE applications," in *APMC Dig. Papers*, 2006, pp. 1102-1105.

REFERENCES

[1] J. R. Long, "Monolithic transformers for silicon RF IC design," *IEEE J. Solid-State Circuits,* vol. 35, pp. 1368-1382, 2000.

[2] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[3] M. Ahn, C.-H. Lee, B. S. Kim, and J. Laskar, "A high-power CMOS switch using a novel adaptive voltage swing distribution method in multistack FETs," *IEEE Trans. Microw. Theory Tech.,* vol. 56, pp. 849-858, Apr. 2008.

[4] A. Hajimiri, "Fully integrated RF CMOS power amplifiers - a prelude to full radio integration," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symp. Dig. Papers*, 2005, pp. 439-442.

[5] H. S. Bennett, R. Brederlow, J. C. Costa, P. E. Cottrell, W. M. Huang, A. A. Immorlica, Jr., J. E. Mueller, M. Racanelli, H. Shichijo, C. E. Weitzel, and B. Zhao, "Device and technology evolution for Si-based RF integrated circuits," *IEEE Trans. Electron Devices,* vol. 52, pp. 1235-1258, Jul. 2005.

[6] B. Razavi, *RF Microelectronics*. Upper Saddle River, NJ: Prentice-Hall. Inc., 1998.

[7] S. Yamanouchi, Y. Aoki, K. Kunihiro, T. Hirayama, T. Miyazaki, and H. Hida, "Analysis and Design of a Dynamic Predistorter for WCDMA Handset Power Amplifiers," *IEEE Trans. Microw. Theory Tech.,* vol. 55, pp. 493-503, 2007.

[8] B. Razavi, *Design of Integrated Circuits for Optical Communications*, 1st ed.: McGRAW-HILL, 2003.

[9] R. W. Jackson, "Rollett proviso in the stability of linear microwave circuits-a tutorial," *IEEE Trans. Microw. Theory Tech.,* vol. 54, pp. 993-1000, 2006.

[10] Skyworks, "http://www.skyworksinc.com." (November/2009)

[11] RFMD, "http://www.rfmd.com." (November/2009)

[12] R. Kruger and H. Mellein, *UMTS - Introduction and measurement*: Rhode&Schwarz, 2004.

[13] J. Kang, A. Hajimiri, and B. Kim, "A single-chip linear CMOS power amplifier for 2.4 GHz WLAN," in *IEEE ISSCC Dig. Tech. Papers*, 2006, pp. 761-769.

[14] V. Aparin, "Analysis of CDMA signal spectral regrowth and waveform quality," *IEEE Trans. Microw. Theory Tech.,* vol. 49, pp. 2306-2314, 2001.

[15] A. Suarez and R. Quere, *Stability Analysis of Nonlinear Microwave Circuits*. Norwood, MA: Artech House, 2003.

[16] I. Aoki, S. D. Kee, D. B. Rutledge, and A. Hajimiri, "Distributed active transformer-a new power-combining and impedance-transformation technique," *IEEE Trans. Microw. Theory Tech.,* vol. 50, pp. 316-331, Jan. 2002.

[17] H. Gan, "On-chip transformer modeling, characterization, and applications in power and low noise amplifiers." vol. Ph.D Dissertation: Stanford University, 2006.

[18] C. Yoo and Q. Huang, "A common-gate switched 0.9-W class-E power amplifier with 41% PAE in 0.25-um CMOS," *IEEE J. Solid-State Circuits,* vol. 36, pp. 823-830, 2001.

[19] K.-C. Tsai and P. R. Gray, "A 1.9-GHz, 1-W CMOS class-E power amplifier for wireless communications," *IEEE J. Solid-State Circuits,* vol. 34, pp. 962-970, Jul. 1999.

[20] D. Heo, A. Sutono, E. Chen, Y. Suh, and J. Laskar, "A 1.9-GHz DECT CMOS power amplifier with fully integrated multilayer LTCC passives," *IEEE Microw. Wireless Compon. Lett.,* vol. 11, pp. 249-251, 2001.

[21] J. Aguilera and R. Berenguer, *Design and Test of Integrated Inductors for RF Applications*: Kluwer Academic Publishers, 2003.

[22] C. P. Yue and S. S. Wong, "On-chip spiral inductors with patterned ground shields for Si-based," *Solid-State Circuits, IEEE Journal of,* vol. 33, pp. 743-752, 1998.

[23] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. New York: Wiley, 2001.

[24] J. Kang, J. Yoon, K. Min, D. Yu, J. Nam, Y. Yang, and B. Kim, "A highly linear and efficient differential CMOS power amplifier with harmonic control," *IEEE J. Solid-State Circuits,* vol. 41, pp. 1314-1322, 2006.

[25] S. Kang, B. Choi, and B. Kim, "Linearity analysis of CMOS for RF application," *IEEE Trans. Microw. Theory Tech.,* vol. 51, pp. 972-977, 2003.

[26] J. Kang, D. Yu, Y. Yang, and B. Kim, "Highly linear 0.18-um CMOS power amplifier with deep n-Well structure," *IEEE J. Solid-State Circuits,* vol. 41, pp. 1073-1080, 2006.

[27] I. Aoki, S. D. Kee, D. B. Rutledge, and A. Hajimiri, "Fully integrated CMOS power amplifier design using the distributed active-transformer architecture," *IEEE J. Solid-State Circuits,* vol. 37, pp. 371-383, Mar. 2002.

[28] S. Kim, K. Lee, J. Lee, B. Kim, S. D. Kee, I. Aoki, and D. B. Rutledge, "An optimized design of distributed active transformer," *IEEE Trans. Microw. Theory Tech.,* vol. 53, pp. 380-388, 2005.

[29] C. Wang, M. Vaidyanathan, and L. E. Larson, "A capacitance-compensation technique for improved linearity in CMOS class-AB power amplifiers," *IEEE J. Solid-State Circuits,* vol. 39, pp. 1927-1937, Nov. 2004.

[30] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, 2nd ed. Norwood, MA: Artech House, 2006.

[31] W. H. Doherty, "A New High Efficiency Power Amplifier for Modulated Waves," *Proc. IRE,* vol. 24, pp. 1163-1182, 1936.

[32] L. R. Kahn, "Single-sideband transmission by envelope elimination and restoration," *Proc. IRE,* vol. 40, pp. 803-806, Jul. 1952.

[33] H. Chireix, "High power outphasing modulation," *Proc. IRE,* vol. 23, pp. 1370-1392, 1935.

[34] C. Park, Y. Kim, H. Kim, and S. Hong, "A 1.9-GHz CMOS power amplifier using three-port asymmetric transmission line transformer for a polar transmitter," *IEEE Trans. Microw. Theory Tech.,* vol. 55, pp. 230-238, 2007.

[35] Y. Kim, C. Park, H. Kim, and S. Hong, "CMOS RF power amplifier with reconfigurable transformer," *Electron. Lett.,* vol. 42, pp. 405-407, 2006.

[36] G. Liu, P. Haldi, T.-J. K. Liu, and A. M. Niknejad, "Fully Integrated CMOS power amplifier with efficiency enhancement at power back-off," *IEEE J. Solid-State Circuits,* vol. 43, pp. 600-609, Mar. 2008.

[37] K. H. An, Y. Kim, O. Lee, K. S. Yang, H. Kim, W. Woo, J. J. Chang, C.-H. Lee, H. Kim, and J. Laskar, "A monolithic voltage-boosting parallel-primary transformer structures for fully integrated CMOS power amplifier design," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symp. Dig. Papers*, 2007, pp. 419-422.

[38] O. Lee, K. S. Yang, K. H. An, Y. Kim, H. Kim, J. J. Chang, W. Woo, C.-H. Lee, and J. Laskar, "A 1.8-GHz 2-watt fully integrated CMOS push-pull parallel-combined power amplifier design," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symp. Dig. Papers*, 2007, pp. 435-438.

[39] D. K. Su and W. J. McFarland, "An IC for linearizing RF power amplifiers using envelope elimination and restoration," *IEEE J. Solid-State Circuits,* vol. 33, pp. 2252-2258, Dec. 1998.

[40] N. Srirattana, A. Raghavan, D. Heo, P. E. Allen, and J. Laskar, "Analysis and design of a high-efficiency multistage Doherty power amplifier for wireless communications," *IEEE Trans. Microw. Theory Tech.,* vol. 53, pp. 852-860, 2005.

[41] M. Elmala and R. Bishop, "A 90nm CMOS Doherty power amplifier with integrated hybrid coupler and impedance transformer," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symp. Dig. Papers*, 2007, pp. 423-426.

[42] I. Cendoya, J. de No, B. Sedano, A. Garcia-Alonso, D. Valderas, and I. Gutierrez, "A new methodology for the on-wafer characterization of RF integrated transformers," *IEEE Trans. Microw. Theory Tech.,* vol. 55, pp. 1046-1053, 2007.

[43] P. Haldi, D. Chowdhury, L. Gang, and A. M. Niknejad, "A 5.8 GHz Linear Power Amplifier in a Standard 90nm CMOS Process using a 1V Power Supply," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symp. Dig. Papers*, 2007, pp. 431-434.

[44] H. Solar, R. Berenguer, I. Adin, U. Alvarado, and I. Cendoya, "A Fully Integrated 26.5 dBm CMOS Power Amplifier for IEEE 802.11a WLAN Standard with on-chip "power inductors"," in *Microwave Symposium Digest, 2006. IEEE MTT-S International*, 2006, pp. 1875-1878.

[45] J. Paramesh, R. Bishop, K. Soumyanath, and D. J. Allstot, "A four-antenna receiver in 90-nm CMOS for beamforming and spatial diversity," *IEEE J. Solid-State Circuits,* vol. 40, pp. 2515-2524, 2005.

[46] S. D. Kee, I. Aoki, A. Hajimiri, and D. Rutledge, "The class-E/F family of ZVS switching amplifiers," *IEEE Trans. Microw. Theory Tech.,* vol. 51, pp. 1677-1690, 2003.

[47] N. O. Sokal and A. D. Sokal, "Class E-A new class of high-efficiency tuned single-ended switching power amplifiers," *IEEE J. Solid-State Circuits,* vol. 10, pp. 168-176, 1975.

[48] R. Zulinski and J. Steadman, "Class E Power Amplifiers and Frequency Multipliers with finite DC-Feed Inductance," *IEEE Trans. Circuits Systems,* vol. 34, pp. 1074-1087, Sep. 1987.

[49] ITRS, "System Drivers," International Technology Roadmap for Semiconductors, 2001.

[50] R. Brama, L. Larcher, A. Mazzanti, and F. Svelto, "A 30.5 dBm 48% PAE CMOS class-E PA with integrated balun for RF applications," *IEEE J. Solid-State Circuits,* vol. 43, pp. 1755-1762, Aug. 2008.

[51] P. Reynaert and M. Steyaert, "A 2.45-GHz 0.13-um CMOS PA with parallel amplification," *IEEE J. Solid-State Circuits,* vol. 42, pp. 551-562, Mar. 2007.

[52] M. Ahn, "Design and analysis of high power and low harmonic RF front end for multi band wireless application," in *Electrical and Computer Engineering*: Georgia Institute of Technology, 2007.

[53] M.-C. Yeh, Z.-M. Tsai, R.-C. Liu, K. Y. Lin, Y.-T. Chang, and H. Wang, "Design and analysis for a miniature CMOS SPDT switch using body-floating technique to improve power performance," *IEEE Trans. Microw. Theory and Tech.,* vol. 54, pp. 33-39, 2006.

[54] M. Ahn, K. H. An, C.-H. Lee, J. Laskar, and H. Kim, "Fully integrated RF front-end circuits in 2 GHz using 0.18 um standard CMOS process," in *APMC Dig. Papers*, Hong Kong, 2008.

[55] K. H. An, O. Lee, H. Kim, D. H. Lee, J. Han, K. S. Yang, Y. Kim, J. J. Chang, W. Woo, C.-H. Lee, H. Kim, and J. Laskar, "Power-combining transformer techniques for fully-integrated CMOS power amplifiers," *IEEE J. Solid-State Circuits,* vol. 43, pp. 1064-1075, May 2008.

[56] A. Shirvani, D. K. Su, and B. A. Wooley, "A CMOS RF power amplifier with parallel amplification for efficient power control," *IEEE J. Solid-State Circuits,* vol. 37, pp. 684-693, Jun. 2002.

[57] ITRS, "Radio Frequency and Analog/Mixed-Signal Technologies for Wireless Communications," International Technology Roadmap for Semiconductor, 2007.

VITA

Kyu Hwan An was born in Incheon, Korea, in 1974. He received a B.S. degree (Magna Cum Laude) in electrical engineering from Pohang University of Science and Technology (POSTECH), Pohang, Korea, in 1997, and an M.S. degree in electrical engineering from the Korea Advanced Institute of Technology (KAIST), Daejeon, Korea, in 1999. He is currently working toward a Ph.D. degree in electrical and computer engineering at the Georgia Institute of Technology, Atlanta.

From 1999 to 2005, he was a research engineer with the RF Lab at Samsung Electo-Mechanics Co. Ltd., Suwon, Korea, where he was involved with the development of power amplifier modules for CDMA, GSM applications, and physical layers for IEEE 802.15.3a/4 standards. His current interests include CMOS RF power amplifiers for cellular applications and implementation of passive circuits on silicon substrates.