CMOS Power Amplifiers for Wireless Communications

King Chun Tsai

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2007-161
http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-161.html

December 18, 2007

CMOS Power Amplifiers for Wireless Communications

by

King Chun Tsai

B.S. (Boston University) 1994

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

Doctor of Philosophy
in
Engineering-Electrical Engineering and Computer Sciences
in the
GRADUATE DIVISION
of the
UNIVERSITY of CALIFORNIA, BERKELEY

Committee in charge:
Professor Paul R. Gray, Chair
Professor Seth R. Sanders
Professor Paul K. Wright

Fall 2007

The dissertation of King Chun Tsai is approved:

Chair _____ Date _____

_____ Date _____

_____ Date _____

University of California, Berkeley

Fall 2007

CMOS Power Amplifiers for Wireless Communications

by

King Chun Tsai

Abstract

CMOS Power Amplifiers for Wireless Communications

by

King Chun Tsai

Doctor of Philosophy in Electrical - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Paul R. Gray, Chair

The advancement of CMOS technology has enabled an unprecedented level of integration in modern low cost, small form-factor and low power wireless devices. While Power Amplifies (PAs) are key components in wireless transceivers, their realization and integration in standard CMOS technology is hindered by a number of technological challenges.

One fundamental challenge of high efficiency CMOS PA realization is the low breakdown voltage of thin gate oxide devices. It forces high output power CMOS PAs to operate under high-current, low-impedance levels where they are vulnerable to parasitic losses. Other challenges include the limited intrinsic gain and large parasitic capacitance and resistance of CMOS transistors, as well as the lack of high quality factor monolithic passive components.

This thesis addresses these challenges and demonstrates an RF CMOS power amplifier that is suitable for amplification of constant envelope modulated signals that are widely

used in cellular systems such as GSM. The key design innovations include (1) the use of a differential switch-mode Class-E structure to optimize power efficiency, extend power capacity, and minimize the impact of substrate noise injection; (2) the use of injection locking technique to significantly reduce the input driving requirement of large transistors; and (3) the design of a compact hybrid balun to interface the differential PA with any conventional signal-ended RF load. The effectiveness of these techniques is demonstrated in a PA prototype that is fabricated in a standard 0.35um CMOS process. The prototype operates up to 2GHz and is capable of delivering 1 Watt of output power with a 48% power-added efficiency (41% including the balun). Comparing with its predecessors, this prototype demonstrates a new level of operational frequency, output power and power efficiency achievable by CMOS power amplifiers.

_____

Professor Paul R. Gray, Chair

Acknowledgements

First and foremost, I would like to express my deep and sincere gratitude to my supervisor Prof. Paul R. Gray for his guidance, patience and encouragement. His insist on excellence remains a beacon in my personal and professional lives. I also thank Professors Seth R. Sanders and Paul K. Wright for dedicating their time in reviewing this thesis.

The years in Berkeley have been the most inspiring and satisfying experience for me. I enjoyed every moment of it and I am grateful to the faculty of the EECS department for putting together such a fine program. The teachings of Professors Paul R. Gray, Robert G. Meyer, Bernhard E. Boser, Chenming Hu, and Edward A. Lee, in particular have induced great interest and continue to be a source of aspiration. I also appreciate the friendship and altruism of the comrades at Cory Hall and BWRC.

I am deeply indebt to the C.W. Chu foundation whose generous scholarship gave me the opportunity to pursue higher education in the United States. I owe my sincere gratitude also to Prof. Ted Moustakas of Boston University who plays a pivotal role in my education.

Last but certainly not least, I am grateful to my parents Kuang Sun Tsai and Ho Chi Lau for their sacrifice and supports. My special gratitude also goes to my wife Stella Tse whose patient love has accompanied me through the toughest times, and I am looking forward to a long and joyful life with her and our family.

King Chun Tsai

Table of Contents

List of Figures

List of Tables

Chapter 1 - Introduction

1.1 Background

The success of a wireless technology relies on the availability of affordable, compact and power efficient mobile devices capable of delivering robust performance using small batteries [Neuv96]. Over time, tremendous improvements are made in these desirable product features and particularly in consumer-oriented products such as cellular phones, Bluetooth and WiFi devices. Many factors contribute, but the key one is accredited to advanced CMOS technology which has allowed an unprecedented level of integration and cost reduction in modern wireless systems [Issa00].

A modern wireless device can generally be partitioned into the three parts shown in Figure 1-1: 1) a pure digital backend (including the microprocessor, memory and digital signal processor etc.) is the device's central intelligence that synthesizes and analyses signals in accordance to the system protocol; 2) a heavily analog/RF transceiver takes care of digital/analog data conversion, frequency translation, filtering and gain control; and lastly 3) a RF front-end, which includes the power amplifier, ensures proper interface between the transceiver and the antenna.

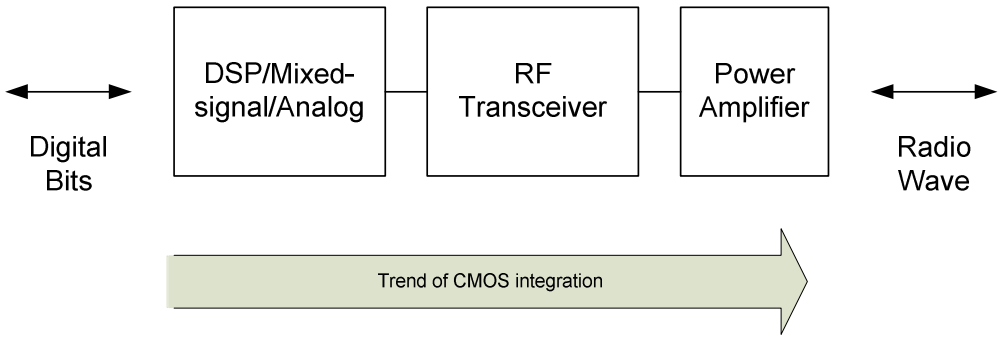


Figure 1-1 A typical partition of wireless system

The extreme cost sensitivity of consumer IC industry has made low-cost standard CMOS the de-facto technology of all digital and majority of mixed-signal and analog wireless

building blocks [Song86][Cave98][Huan98]. In fact, as the high-frequency performance of the technology continue to improve, core RF circuits in wireless transceivers are rapidly migrating to standard CMOS process [Cami94][Gray95][Stey96]. Table 1-1 shows the consistent enhancement of some important transistor performance metrics with technology scaling. These include f_T, f_{max}, minimum Noise Figure, and Flicker noise.

In addition to active components, high quality monolithic passive components are also essential to most RF transceivers [ITRS03]. These components include metal-insulator-metal (MIM) capacitors and ultra-thick metal inductors; and they too, are becoming readily available in standard CMOS technology with relatively insignificant cost implication. These favorable aspects in both active devices and passive components have resulted in impressive levels of CMOS integration achieved in today's wireless transceiver ICs [Sheng96][Rude97] [Rofo98] [Rofo98b] [Cho99] [Dara01].

Year	1997	1999	2002	2005	2008	2011	Unit
Min Channel Length	250	180	130	100	70	50	Nm
f_{max}/f_T	25/20	35/30	50/40	65/55	90/75	100/130	GHz
Min Noise Figure	2	1.5	1.2	<1	<1	<1	dB
1/f Noise	5	2	.5	.05	.02	.02	10-12 V2/Hz-um2

Table 1-1 Data from International Technology Roadmap of Semiconductor

With so many advances in integrated CMOS RF transceivers, one may be surprised to find that CMOS integration of the power amplifier has so far been relatively unsuccessful. The PA is a key transceiver building block that delivers radiation power to the transmit antenna; and for applications requiring moderate to high output power (such as cellular handset and wireless LAN), the PA often dominates the total transmit current consumption. According to [Matt97], as much as 90% of the total current consumption during active transmits in a GSM phone is attributed to the PA. Power amplifier efficiency is indisputably crucial to the overall system performance. Yet at least for now, excellent PA performance in the low-GHz range can only be met in production by discrete modules or MMICs using GaAs and other specialized technologies.

To realize high-efficiency RF power amplifiers in standard CMOS [Gupt99], one needs to understand the technology's limitations. We might also need to reconsider the effectiveness of traditional design approaches (e.g. Classes-A, B, AB and C) that optimize efficiency *only* at the maximum output power – a statistical rarity in many wireless applications. Pioneering efforts have shown the promise of CMOS power amplifiers [Rofo94] [Rofo95] [Su97]. However, much effort is still on-going to improve efficiency, output power and operation frequency before the ultimate goal of full CMOS, single-chip radio, can be practically achieved [Yoo01] [Gupt01] [Mert02] [Aoki02].

1.2 The challenges of CMOS PA

The most fundamental challenges in the realization of a CMOS PA can be traced to the low breakdown voltage, limited gain and parasitic capacitances and resistances of CMOS transistors. To achieve high efficiency, power amplifiers are often inductively loaded which results in large drain-to-gate voltage swing (beyond V_{DD}) during normal operation. Such high voltage swings can cause stress and damage to the thin gate oxide which in modern technologies consists of only a few layers of dielectric molecules. As CMOS technology scales down to finer geometries, the gate oxide breakdown voltage continues to drop and forces the PA to operate under increasingly lower supply voltage. Yet in order to deliver the same output power, the transistors now have to be larger in size in order to deliver more current into a low impedance load.

Operation under low impedance level can be detrimental to power efficiency because it increases the circuit's vulnerability to parasitic losses caused by on-chip routing, bond wire and bond pad contact resistance and other losses. This is by the way the same reason by electrical companies deliver power in high voltage in order to reduce cable losses. To make matter worse, power amplifiers are almost always designed to drive a standard load (typical 50 Ohm). A power amplifier with a low internal impedance level thus has to rely on a high ratio impedance transformation network to interface with the standard load, and this often results in additional power loss and bandwidth reduction.

Another big challenge related to low voltage operation is that the overdrive voltage (V_{GS}-V_T) drops as the threshold voltage becomes a significant portion of the available voltage swing; and this combines with the relatively low current gain of CMOS devices (in comparison with e.g. bipolar transistors) causes the transistor size to increase at a rate that is *faster* than that of the supply voltage reduction. Powerful driver circuits are now a necessity, and their power consumption becomes a big part of the overall dissipation.

1.3 Pioneering Developments in CMOS PAs

In 1994, Rofougaran [Rofo94] reported a power amplifier implemented in 1um CMOS technology for use in an all-CMOS spread-spectrum transceiver operation at 902-928 MHz. The PA was capable of delivering 20mW output power from a 3V supply at 25% efficiency. This is the first CMOS PA reported for 900MHz operation despite its relatively modest output power and efficiency.

While 20mW is adequate for some short-range wireless systems, applications such as cellular phones which operate over a much wider range typically require output of 1-W and beyond. Furthermore, the Class-C power amplifier introduces a fundamental trade-off between efficiency and output power, making it difficult to enhance them simultaneously for the same design.

In 1997, Su [Su97] argued that for wireless systems that do not require any linearity in the power amplifiers, much higher output power and efficiency can be achieved by using

a switch-mode power amplifier topology. These systems form an important category since they include GSM, the prevailing cellular technology. In GSM, transmit signals are modulated with a constant-envelope modulation scheme known as GMSK which is extremely tolerant to PA non-linearity. Su reported a Class-D CMOS power that was capable of delivering 1-W of output power at 824-849MHz with drain efficiency (η_{drain}) of 62% and power added efficiency (PAE) of 42%. Such performance was a milestone for CMOS PA, although it did suffer from a few potential drawbacks: The operation frequency was relatively low and did not include the important GSM and PCS bands at 1.8-1.9GHz. In addition, the relatively large gap between drain efficiency and PAE suggests that significant power was lost in the driver stages – and this could degrade further rapidly at higher operation frequency when the capacitive loading effect of device parasitic multiplies. In addition, the fully single-ended topology used in Su's PA injects large amount of disturbance into the substrate at the operating frequency, and this could be detrimental if the PA were integrated into a complete transceiver comprising sensitive circuits operating at the same frequency.

1.4 Research Goal

The goal of this research [Tsai99] is to understand the fundamental challenges of CMOS PA realization and to develop appropriate circuit techniques to effectively overcome these challenges. The criteria of our proposed techniques should include the capability of delivering useful amount of output power at 1-2GHz range and with efficiency

competitive with more expensive main-stream, non-CMOS technologies. We will also need to validate our proposal with experimental data. With the acquired knowledge through the research process, we hope to contribute in pushing the envelope of transceiver integration, and to develop technique that will ultimately be used in the lowest cost, most compact and highly power efficient single-chip wireless systems.

1.5 Thesis Organization

This chapter provides a background on wireless transceiver integration, and poses the motivation for CMOS PA implementation. We also established the research goal. For the rest of this thesis, we will report this research effort in the following organization:

Chapter 2 - Power Amplification Fundamentals

This chapter introduces the basic concepts of power amplification, and explores the fundamental roles taken by the active device and the impedance matching network within the amplifier. We also define the key PA performance parameters for comparing among various power amplifier design approaches.

Chapter 3 - Transconductor-based Power Amplifiers

Power amplifiers are traditionally classified into transconductor-based and switch-mode amplifiers. In this chapter we explore the distinction between these two types of PAs and

in particular focuses on operation principles and design trade-offs of the transconductor-based PAs.

Chapter 4 - Switch-mode Power Amplifiers

Switch-mode PAs are generally much more efficient than their transconductor-based counterparts. This type of PAs is very promising for applications that relies constant-envelope modulation such as our target applications. In this chapter, we will examine the operation principle of two main types of switch-mode PA design approaches.

Chapter 5 - Power Amplifier Implementation Technologies

Currently mainstream power amplifiers are predominantly manufactured in three prevailing semiconductor technologies. This chapter examines the operational principle and device characteristics of these technologies and compares them with CMOS in the context of PA realization. We will also review the oxide breakdown mechanism of CMOS devices.

Chapter 6 - Analysis of Class-E PA

Class-E operation is chosen as the basis of our CMOS PA design because of its inherently high efficiency. In chapter we will examine the defining characteristic of

Class-E power amplifier and provide an analytical background of the circuit operation and design trade-off.

Chapter 7 - Design of a 1-W 1.9GHz CMOS Class-E PA

In this chapter we will review the detail design considerations of the CMOS Class-E PA. Classical single-ended Class-E PA has drawbacks regarding input driving requirement, low operation impedance level, sensitivity to parasitic and potential noise-coupling. These drawbacks can seriously impact the achievable performance the PA implemented in a low-voltage integrated CMOS environment. We will explore how these effects can be alleviated by incorporating a fully differential topology and with the injection locking technique.

Chapter 8 - A Compact Differential to Single-ended Converter

In order to make the differential CMOS PA a practical solution, one must derive an efficient way of converting the differential PA output back to a single-ended signal for filter or antenna interface. This chapter proposes a new compact hybrid balun structure to address this key signal conversion issue. A detailed analysis of this compact microstrip balun is provided in this chapter.

Chapter 9 - Experimental Prototype and Measured Results

The chapter reviews the measurement results on the CMOS PA and the compact microstrip balun. The discussion includes the experimental set-up, the method of evaluation, as well as techniques of high frequency measurements.

Chapter 10 - Conclusion

This chapter summarizes this research work on CMOS power amplifier design and discusses the perspective of applying our results to future research.

Chapter 2 - Power Amplification Fundamentals

2.1 A conceptual model for Power Amplifier

Shown in Figure 2-1 is a conceptual model of the operation of a power amplifier. Here we partition the PA into two main components — a passive load network and an active device. The passive load network consists of a resistive load and an impedance matching network. At the operation frequency of several hundred MHz and beyond, the load impedance is often standardized to 50 Ohm. This is the eventual sink of useful output power and it may represent, for example, an antenna or a filter which is driven by the PA. The matching network serves the important function of impedance transformation which ultimately allows the PA to deliver high output power at low operational voltage.

During operation, an RF input signal drives the active device and causes it to modulate the load network with periodic excitation. This process draws energy from the power supply and delivers it to the load in the form of an RF signal (the load network thus always contains a path that conducts current from supply, often through an inductor or an RF choke). It is during this process that information-bearing characteristic of the input signal, such as frequency and/or amplitude modulation, is transferred to the output.

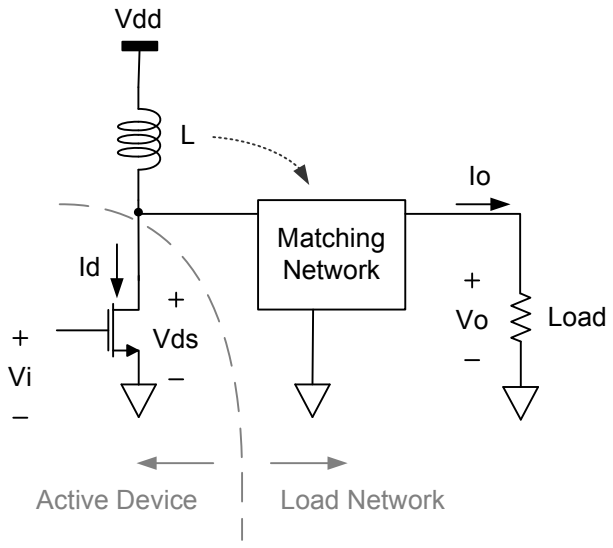


Figure 2-1 A generic PA model

One of the most critical concerns in PA design is the efficiency of the power conversion process. An efficient PA delivers most of the energy drawn from the supply to the load,

whereas an inefficient one dissipates much energy as heat in the active device or in the lossy components within the matching network.

The active device often takes the form of a single or an array of transistors; it controls the power delivering capability of the power amplifier and its characteristics are highly technology dependent. Later in this chapter, we will discuss the concept of Safe Operation Area (SOA), which defines the operation capability of the active device. The maximum achievable gain of a single-stage PA is mostly determined by the active device, and this is an important challenge for high efficiency CMOS PA realization. With low intrinsic gain, multiple cascaded gain stages are needed and the additional power consumption of the drivers can significantly degrade the overall PA efficiency. Besides, the parasitic elements of the transistor also play an important role in the achievable performance. We will review the prevailing PA transistor technologies and further explore these effects in Chapter 5 .

2.2 The Matching Network

Impedance Transformation

To understand the significance of the impedance matching network, let's consider a hypothetical but realistic situation in which a PA has to deliver a 1-W sinusoidal signal into a 50Ω load. The peak voltage swing across the load in this case will be

$$v_{out} = \sqrt{P_{out} \cdot 2R_L} = 10\text{V} \tag{2.1}$$

If the load were connected directly to the drain of the active device, then the device drain would need to be biased at 10V supply. Since the transistor output is inductively load and swings above supply, the device in this case will have to withstand a peak voltage of 20V. This arrangement presents two problems – first, 10V supply voltage is rarely available in battery powered consumer wireless devices; second and more importantly, the active device may not even survive the high voltage without catastrophic breakdown – the breakdown limit for 0.35um CMOS, for example, is less than 6V, and it gets even lower with more advanced CMOS technologies.

The matching network is therefore an essential component to deliver high output power with limited voltage swing. Consisting of only reactive components, a matching network is ideally lossless, and it transforms the load impedance to a level manageable by the active device. An example is shown in Figure 2-2. This matching circuit is formed by a simple L-C network and can match the load impedance R_L down to an impedance level equals to R_L / M , where $M = (1 + R_L^2 / X^2)$ is the matching ratio.

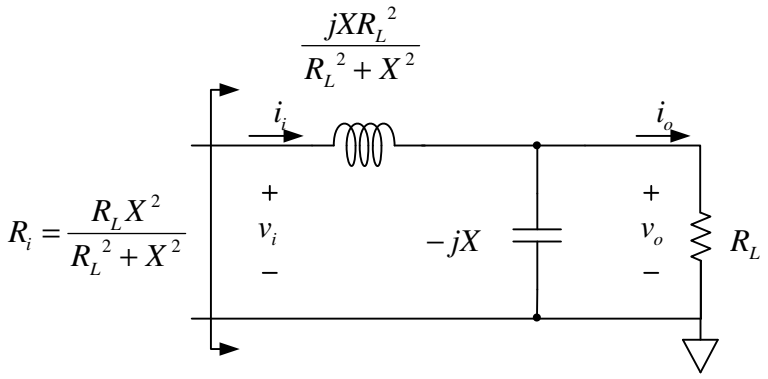


Figure 2-2 A step down matching network

With a lossless matching network, the input power equals to the power delivered to the actual load impedance. However, the voltage swing at the input port is now reduced by a factor of $\sqrt{M}$ due to the lowered input impedance. In our previous example, if we choose $M = 16$, then the maximum voltage appearing at the drain of the active device is reduced from 20V to 5V. Notice that the maximum current conducted by the active device is also increased by a factor of $\sqrt{M}$, but this increased current handling requirement can be easily overcome by using $\sqrt{M}$ devices in parallel. In essence, impedance transformation allows us to trade off the voltage handling capability of the device (which does not scale with device size) with its current handling capability, and therefore extends the power delivering capacity of the given device technology. However, the amount of extension will be eventually limited by parasitic effects.

Returning to the prior example where the impedance presented to the active device is reduced from 50Ω to 50/16=3.125Ω. Such low impedance allows high power to be delivered with low voltage swing, but it also increases the amplifier's sensitivity to parasitics. Consider if a parasitic resistance of R_p is present at the amplifier output due to on-chip metal routing and wire-bonding contact resistance, then the amplifier efficiency will be reduced by a factor *Rp/(Rp+Ri).* In this case, if R_p= 0.5Ω (a realistic value in practice), the loss due to *Rp* will increase 140-folds from 0.1% to 14% when *Ri* is transformed from 50Ω to 3.125Ω.

Waveform Shaping and Filtering

Besides the primary function of impedance transformation, matching networking can also enhance PA efficiency by properly shaping the current and voltage waveforms at the drain of the active device. This is the key-concept embraced in all switch-mode power amplifiers — when the active device is hard-driven, strong harmonics are present at multiples of the fundamental frequency. Since the impedance looking into the matching network is frequency dependent, it can be designed to provide proper terminations at the various harmonic frequencies so as to achieve a desirable voltage/current waveform at the drain. The usual goal is to minimize the conduction of transistor current when the drain voltage is high, as minimizing such overlap can greatly enhance the PA efficiency. This

concept will be explored in greater details when we discuss the switch-mode power amplifiers in Chapter 4 .

Although harmonics can be exploited at the drain node to optimize PA efficiency, they are generally undesirable at the output of the PA where they may interfere with other wireless systems. Regulatory bodies such as the FCC have strict regulations on how much out-of-band interference can be tolerated. When properly designed, the matching network can attenuate most of the harmonic components at the PA output and in some cases even eliminate the need of additional band-pass filter between the PA and the antenna.

2.3 Key Power Amplifier Performance Parameters

In this section we will review parameters generally associated with power amplifier performance, so as to establish a common baseline on which different design approaches can be compared [Abul01].

Power Efficiency

Power efficiency is one of the most important PA performance metrics. It measures how much power drawn from the energy source is delivered to the load as oppose to being dissipated within the PA. Three definitions are commonly used in the literature – the

drain efficiency, the power added efficiency and the overall efficiency. The *drain efficiency* (η_{drain}) of a PA is defined as

$$\eta_{drain} = \frac{P_{out}}{P_{DC}} \tag{2.2}$$

where P_{DC} is the power delivered from the DC power supply and P_{out} is the useful power of the output signal. In communication applications, P_{out} generally excludes the power of any undesirable harmonics present at the output. One may notice that the drain efficiency ignores the power delivered by the input signal, but the difference is often negligible if the PA gain is high or if the input impedance of the PA is predominately reactive. On the other hand, when the overall power gain is low, the input power may become a substantial portion of the output power. In this case, the *power added efficiency* and the *overall efficiency* provide a more accurate measure of the PA performance.

Power added efficiency (PAE) is defined as

$$PAE = \frac{P_{out} - P_{in}}{P_{DC}} \tag{2.3}$$

where P_{in} is the power of the input signal, while the *overall efficiency* ($\eta_{overall}$) is defined as:

$$\eta_{overall} = \frac{P_{out}}{P_{DC} + P_{in}} \tag{2.4}$$

Both PAE and overall efficiency capture the effect of input power on efficiency, and it is sometimes a matter of preference on which one to use.

Power Capacity

Power capacity measures the amount of output power deliverable from a PA when its transistor is operating under the technology's maximum reliability limits. It is an important consideration especially in designing CMOS PAs where the transistors' operations are subject to low breakdown voltage and low current gain. Passive components, such as metal interconnects, resistors and capacitors also have reliability limits due to phenomena such as electro-migration and oxide breakdown. However, these limits are usually more relaxed when compared with the active devices, and can often be mitigated with proper design and layout considerations.

The maximum operation voltage of an active device, whether it is BJT, CMOS, HBT or MESFET, is determined by its technology-dependent breakdown mechanism. Such voltage limit generally depends only on bias condition and device structure (e.g. device type, doping profile, dielectric thickness etc.) but independent of the device multiplicity. On the other hand, the maximum operation current of an active device generally scales directly with the number of instances used to constitute the device.

For a given transistor, the *Safe Operation Area (SOA)* is a region on the output current-voltage characteristic that is bounded by the current and voltage operation limits of the

device. It is shown as the shaded area in Figure 2-3 (b). As shown in the figure, the lower side of the SOA is bounded by the device going into cut-off region and creases to conduct current. The upper side of the SOA is bounded by the current driving capability of the device. In CMOS, this is typically due to limited transconductance and input voltage swing, whereas in bipolar, this limit is associated with the on-set of high-level injection and loss of f_T). The right-side boundary of the SOA is attributed to the device breakdown voltage which may also be dependent on the current level. In BJT, breakdown is typically induced by the on-set of avalanche breakdown in the collector-base junction, which can be followed by secondary breakdown due to, especially in large emitter area devices, current crowding and thermal instability. The breakdown voltage of MOSFET is dependent on the transistor layout: In devices where the substrate contact is placed close by the transistor to reduce local substrate resistance, breakdown is primarily induced by dielectric breakdown of the gate oxide. Otherwise, the large substrate resistance may allow the on-set of a parasitic bipolar device formed by the drain-bulk-source junctions of the primary MOS device and results in a lower-voltage but usually non-catastrophic breakdown. The former is the typical case for amplifying transistors (our main focus), while the latter situation is often found in ESD protection circuits.

The SOA is characterized by two parameters as shown in Figure 2-3 – the maximum current limit (I_{max}) and maximum voltage limit (V_{max}) of the device. When the PA device is driven by a periodic signal, its output I-V trajectory forms a closed loop on the I-V

plane. In order to maintain a reliable operation, the trajectory has to stay entirely within the SOA.

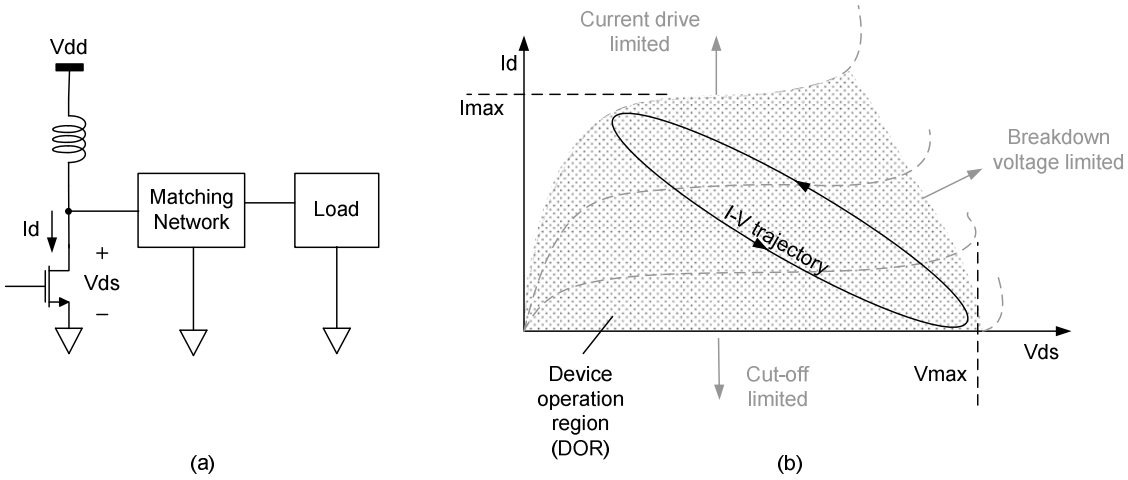


Figure 2-3 (a) a generic PA configuration, (b) current and voltage operation limits and I-V trajectory of the active device

For a given power amplifier design, the Power Capacity (*PC*) is a performance metric that measure how much power can be delivered by the PA with the given I_{max} and V_{max} associated with its active device. Power capacity is defined by

$$PC = \frac{P_{out}}{I_{\max} \cdot V_{\max}} \tag{2.5}$$

where P_{out} is the output power that is delivered to the load. Power capacity is thus a parameter that measures how well a power amplifier utilizes the device technology in delivering the required amount of output power. A high power capacity PA can deliver a given output power with smaller number of devices and is less prone to parasitic effects.

Linearity

In wireless systems, information is often embedded in both the amplitude and phase variations of the RF signals. Such signals are known as non-constant envelope signals since their amplitudes vary with time. Examples of non-constant envelope signals include the OFDM signals used in WLANs and the $\pi/4$ QPSK modulated signals used in some cellular systems.

Non-constant envelope signal exhibits high bandwidth efficiency (frequency bandwidth required to support a certain data rate), but they also pose stringent requirements on the PA since good linearity is needed to preserve the underlying amplitude variation [Ariy90] [Bocc95]. Since linear PAs are generally less efficient than non-linear ones, some wireless systems seek to relax linearity requirement utilizing constant envelope signals [Muro81] [Ande91]. These are fixed amplitude signals in which information is embedded only in the phase variation. Such schemes are widely used in prevailing cellular and cordless systems such as GSM and DECT. The key consequence of constant envelope is that PA nonlinearity results in negligible signal integrity degradation and therefore a

much broader class of nonlinear and highly efficient PAs can be utilized for these applications.

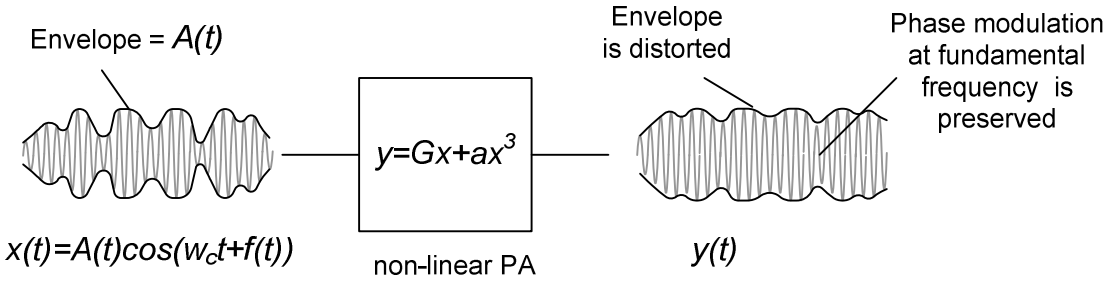


Figure 2-4 Effects of PA nonlinearity on non-constant envelope signal

Constant Envelope Modulation

The immunity of constant envelope signals towards PA nonlinearity can be illustrated by the example shown in Figure 2-4. First we assume the input signal is non-constant envelope, in which case it can be expressed as

$$x(t) = A(t)\cos(\omega_c t + \phi(t)) \tag{2.6}$$

where $A(t)$ is the time-varying envelope (or amplitude modulation), $\phi(t)$ is a phase modulation, and ω_c is the carrier frequency around which the bandpass input signal is

centered. In this example, we assume that PA suffers from weak 3rd order nonlinearity such that its input/output relationship is in the form of

$$y(x) = G \cdot x + \alpha \cdot x^3 \tag{2.7}$$

where G is the nominal gain of the PA, and α represents the nonlinearity. Substituting (2.6) into (2.7), it can be shown that the output signal contains one fundamental component and some higher order harmonics:

$$\begin{aligned} y(t) = &[G \cdot A(t) + \frac{3}{4} \cdot \alpha \cdot A(t)^3] \cdot \cos(\varpi_c t + \phi(t)) \\ &+ \quad \textit{higher order harmonics} \end{aligned} \tag{2.8}$$

Here we focus on the fundamental component only, assuming higher order harmonics can be properly removed by the matching network of the PA. It is clear from (2.8) that the phase modulation of the input signal is preserved at the output, but the signal envelope is distorted by the PA non-linearity. In addition, taking Fourier transform of *y(t)* reveals that the $A(t)^3$ term in (2.8) spreads the output spectrum around the fundamental frequency by three times when compared with the input signal. Such phenomenon is known as *spectral re-growth*. It is an important benchmark in evaluating PA linearity since in communication systems, large spectral re-growth can cause the transmitted signal to interfere with its neighboring channel. If the input signal is constant-envelope, then

$$x(t) = A \cdot \cos(\varpi_c t + \phi(t))$$
$$y(t) = [G \cdot A + \frac{3}{4} \cdot \alpha \cdot A^3] \cdot \cos(\varpi_c t + \phi(t)) \quad + \quad \textit{higher order harmonics} \tag{2.9}$$

Since the term within the bracket in (2.9) is a constant, the fundamental component of the output becomes a non-distorted replica of the input signal. As a result, the PA nonlinearity causes neither spectral re-growth nor degradation in the signal integrity.

Although a weakly non-linear PA is assumed in the illustration above, the immunity of constant-envelope signal towards PA nonlinearity is generally true even for PAs with much stronger nonlinearity.

Chapter 3 - Transconductor-based Power Amplifiers

3.1 Introduction

Power amplifiers can generally be categorized into two basic types – the transconductor-based (gm-based) PAs and the switch-mode PAs [Soka95]. The distinction between them lies in the operation mode of the active devices. The active device in a gm-based PA acts as a high output impedance transconductor throughout the operation cycle, and its output current is controlled solely by the input signal and is independent of the load impedance. In contrast, the active device of a switch-mode PA behaves as a switch that is toggling between an on-state and an off-state. In this case, the output current of the transistor is

almost solely controlled by the load impedance, and only depends weakly on the on-resistance of the transistor.

Gm-based PAs are in general much more linear than the switch-mode ones. This is because the input signal in a gm-based PA has direct control on the output current, and hence can easily transfer the input signal amplitude variation to the output. On the other hand, the input signal of a switch-mode PA is discretized into two levels, (either above or below the switching threshold) and information embedded in the variations of the input signal amplitude is essentially lost in the process. Linear amplification with switch-mode PAs is however still possible if linearization schemes such as Envelope Elimination and Restoration (EER) and Envelope Tracking (ET) are used. These schemes feed-forward the amplitude modulation to the output and achieve an overall linear PA system, even though the core amplifiers embedded within is non-linear.

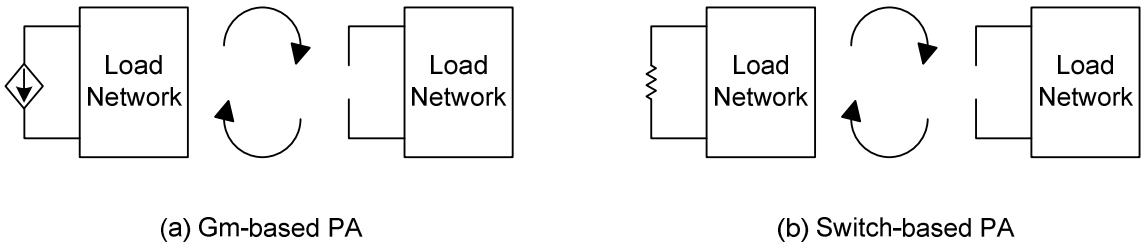


Figure 3-1 Illustration of (a) gm-based PA and (b) switch-mode PA during one conduction cycle

The reason for us to classify power amplifiers into gm-based and switch-mode is that it greatly simplifies the analyses and provides valuable insights into the circuit operation, efficiency limits and reliability constrains. In reality, power amplifiers often operate in a mixed-mode conditions in which case the active devices behave as transconductors for one part, but as switches for the other part of a cycle. These PAs are often products of deliberate attempt to improve efficiency, but sometimes they are results of uncompensated parasitic or unexpected signal overdrive. Close-form analyses of these PAs are in general very complex and designers often rely heavily on simulations and bench experimentation. Nonetheless, even for these more complex forms of power amplifiers, familiarity on the pure gm-based and pure switch-mode operations still offers valuable insights. In the remaining part of this chapter, we will examine the key forms of gm-based PAs. The discussion of switch-mode PAs will be deferred to Chapter 4 and Chapter 5 .

3.2 Class-A, AB, B and C Power Amplifiers

Class-A, AB, B and C PAs are the classical forms of gm-based power amplifiers. They are also the first PAs ever built since the era of vacuum tubes. This family of PAs is closely related, and their distinction lies on the bias condition of the active device rather than on the circuit topology. In fact, even though different implementations exist, the operation principles of these PAs can be examined from a common circuit model shown in Figure 3-2.

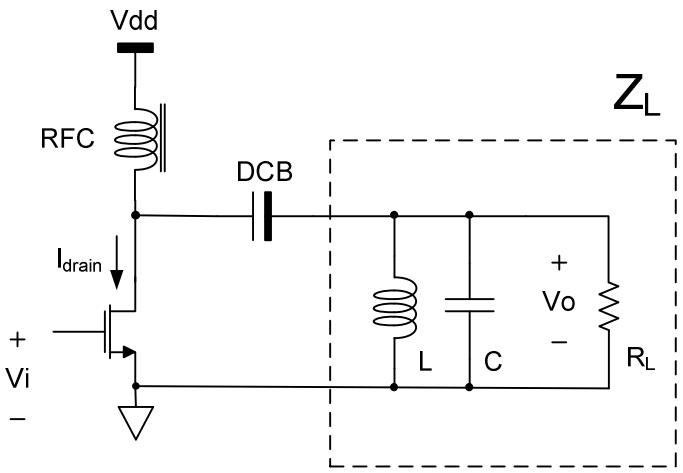


Figure 3-2 A common circuit model for a Class-A, AB, B and C power amplifiers

Let us first highlight the basic assumptions: we assume that the active device in these gm-based PAs behaves as a high output impedance transconductor whose output current depends only on the input voltage. For a MOSFET, this implies that the transistor never enters the triode region. As we will see, such criterion requires an appropriate choice of the load resistance R_L and input signal amplitude. The transistor may however cease to conduct current if the input voltage drops below the device's threshold voltage. The portion of an operation cycle where the device conducts current is known as the *conduction angle* and it defines the class of operation.

In Figure 3-2, the LC tank in the load network forms a resonant circuit at the operation frequency. An RF choke (RFC) is shown here to provide a DC path from the supply to

bias the transistor. In practice, a finite inductor may be used in place of the RFC and in that case, its inductance will be lumped into the LC tank. Similarly, the tank will also absorb the output capacitance of the active device.

We assume further that the quality factor (Q) of the LC tank is sufficiently high, such that the output voltage waveform remains a sinusoid even if high frequency harmonics are present in the output current waveform of the active device. Such filtering effect is generally desirable since it prevents the high frequency harmonics from leaking into the output and causing out-of-band spurious emission. The flip side is that the high-Q tank also limits the operational bandwidth of PA. In practice this is rarely a concern because in prevailing wireless systems, the band of operation is typically only a few percent of the carrier frequency (consider for example, the TX band in Personal Communication Systems(PCS) is only 60MHz, or 3% of the 1.9GHz carrier frequency).

3.3 Class-A Power Amplifiers

The invention of Class-A amplifier was generally attributed to Fritz Lowenstein who discovered in 1911 that the current through a triode tube could be effectively modulated from the grid terminal if the grid is negatively biased with respect to the cathode (see Figure 3-3 for an illustration of the triode tube) [Espe59]. In many ways, a Class-A PA is similar to a standard small-signal amplifier except for its large signal swing is relatively large with respect to the bias level.

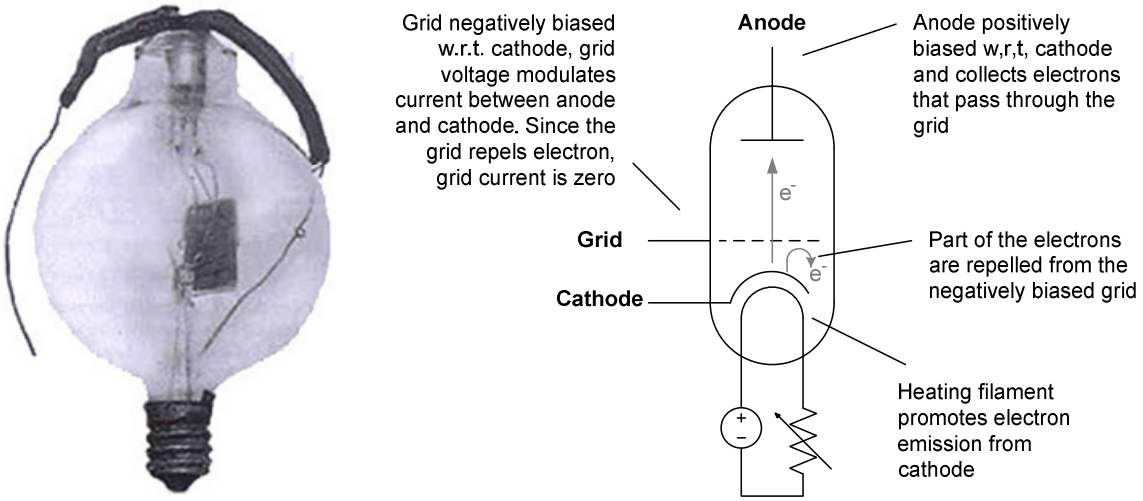


Figure 3-3 A triode tube similar to that used in Lowenstein's Class-A amplifier

The chief advantage of a Class-A PA is its good linearity since the device stays in one operation region throughout the entire operation cycle; this is in contrast to other types of PA where the device may go in an out of triode and cut-off regions and thus results in gross cross-over distortion. The overall linearity of a Class-A PA is generally only limited by the inherent linearity of the device transconductance.

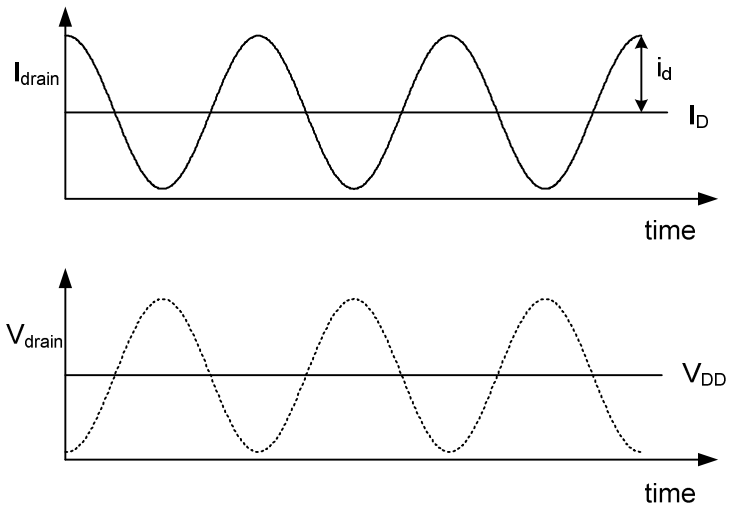


Figure 3-4 Typical drain current and voltage waveforms of a Class-A PA

The main drawback of a Class-A PA is its low efficiency. Consider a typical Class-A drain and current waveforms as shown in Figure 3-4. Assuming the device gm is linear, the drain current can be modeled as

$$i_{drain} = I_D + i_d \cdot \cos(\varpi_o t) \tag{3.1}$$

Here I_D is the current bias level, i_d is the signal swing, and ω_o is the operating frequency. Note that the DC portion of i_{drain} is shunt to the supply through the RFC, and only the AC portion is injected into the tuned load to establish an output voltage that is equal to

$$v_o = -i_d R_L \cdot \cos t(\varpi_o t) \tag{3.2}$$

Due to the DC blocking capacitor (DBC), the drain voltage is identical to the output voltage except it is DC shifted by V_{DD}. i.e.

$$v_{drain} = V_{DD} - V_o \tag{3.3}$$

Furthermore, for Class-A operation, we need to ensure that the device is always conducting current and the drain voltage remains positive so that the device is not pushed into the triode region. This results in the following bounds on the drain current swing and load resistance value:

$$\begin{aligned} i_d &\le I_D \\ i_d \cdot R &\le V_{DD} \end{aligned} \tag{3.4}$$

The output power and the DC power dissipation from the supply can therefore be expressed as

$$P_{out} = \frac{i_d^{\,2} \cdot R_L}{2} \quad \le \quad P_{o,\max} = \frac{I_D \cdot V_{DD}}{2} \tag{3.5}$$

$$P_{DC} = I_D \cdot V_{DD}$$

The drain efficiency of the PA can therefore be calculated as

$$\eta_{drain} = \frac{P_{out}}{P_{DC}} = \frac{i_d^{\,2} \cdot R_L}{2 \cdot V_{DD} I_D} \quad \le \quad \eta_{\max} = \frac{P_{o,\max}}{P_{DC}} = 50\% \tag{3.6}$$

From (3.5), it is clear that the DC power dissipation from the supply is independent of the signal level. The drain efficiency therefore is worse at low signal level. Even at the highest possible signal level, the maximum efficiency of a Class-A PA is only 50%. In practice, the actual achievable efficiency will be even lower because of inevitable parasitic and finite knee voltage (i.e. V_{dsat} for MOS devices) which reduces the maximum signal swing.

Power capacity of a Class-A PA can be calculated from the max drain voltage and current:

$$\begin{aligned} I_{drain,\max} &= I_D + I_{,\max} = 2I_D \\ v_{drain,\max} &= V_{DD} + v_{o,\max} = 2V_{DD} \\ PC_{classA} &= \frac{P_{o,\max}}{I_{drain,\max} \cdot v_{drain,\max}} = \frac{1}{8} \end{aligned} \tag{3.7}$$

3.4 Reduced Conduction Angle Power Amplifiers

It is clear from the simple model in Figure 3-2 that the only power dissipating elements exist in the PA are the load resistor R_L and the active device. While the power delivered to R_L is desirable, the rest is wasted as internal heat dissipation. A key approach to improve efficiency is therefore to reduce the power loss in the transistor [Crip99].

The instantaneous power dissipation in the transistor is the product of its drain current and voltage. One way to minimize this loss is to reduce the conduction angle such that the device only draws current when the drain voltage is at its minimum. According to the conduction angle, the resulting PA can be classified into Class-AB (50% < duty cycle < 100%), Class-B (duty cycle = 50%) and Class-C (duty cycle $\leq$ 50%). Figure 3-5 illustrates the voltage and current waveforms for the three classes of PAs. Note that, in these waveforms we assume the device size is maintained at constant. Therefore, the maximum drain current is set at a fixed level, ideally just below the reliability limit for the given device size.

It is obvious from Figure 3-5 that the overlapping area between the drain current and voltage (and hence the device power loss) decreases with the conduction angle. However, due to the constraint of the fixed maximum drain current, the amount of charge that can be injected into the load network also diminishes with reduced conduction angle. We therefore expect the deliverable output power to drop as the PA is biased towards Class-C. In addition, in order to shrink the conduction angle while maintaining a fixed peak drain current, we will have to simultaneously reduce the bias level and increase the amplitude of the input signal. Since the output power does not increase accordingly, it results in an overall reduction of the PA gain as the conduction angle is reduced.

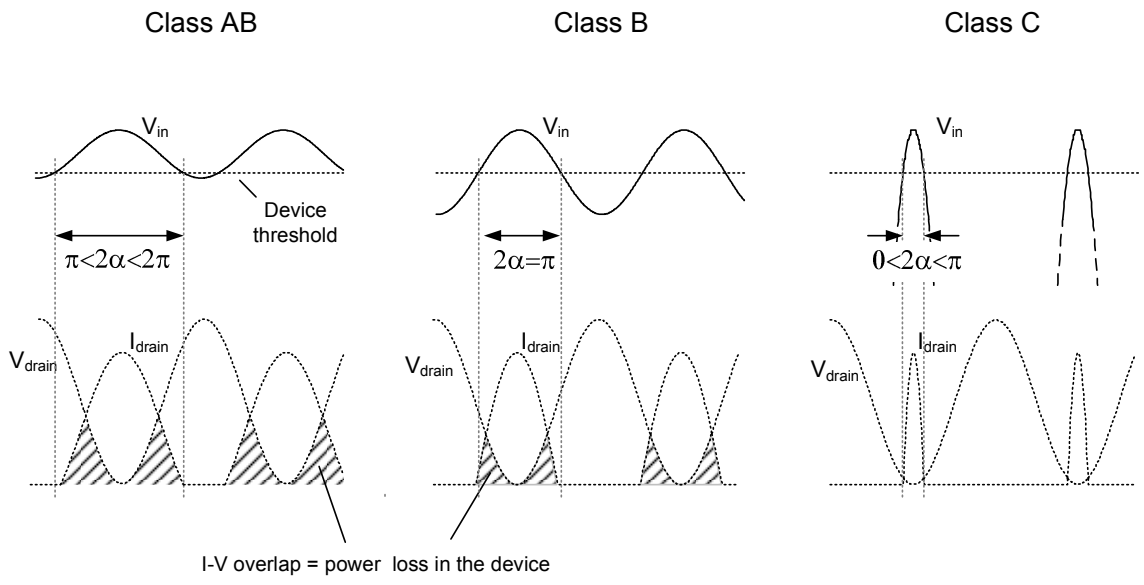


Figure 3-5 Characteristic device input and output waveforms of a Class-AB, Class-B and Class-C PA

Drain Current Harmonic Composition

To understand the operation of reduced conduction angle PAs as a function of the conduction angle, we first need a simple model for the drain current waveform. Here we assume the gm of the device is constant, and the drain current is approximated as a bottom-clipped sine wave:

$$i_{drain}(\theta,\alpha)=\begin{cases}\dfrac{\cos(\theta)-\cos(\alpha)}{1-\cos(\alpha)} & if \quad |\theta-2N\pi|<\alpha \\ 0 & otherwise\end{cases} \tag{3.8}$$

Here, 2α is the conduction angle in radian, θ is the phase angle ($=\omega_o t$), and N is any integer. Note the i_{drain} resembles a cosine function when θ is between $+\alpha$ and $-\alpha$, but is zero otherwise. The waveform of i_{drain} is shown in Figure 3-6.

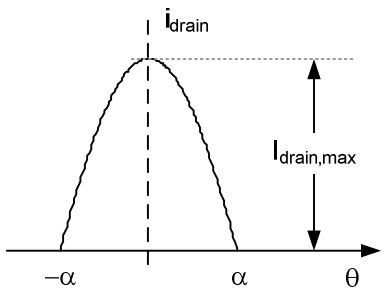


Figure 3-6 Bottom-clipped sinusoidal waveform of i_{drain}

The periodic function i_{drain} can be decomposed into its Fourier components:

$$i_{drain}(\theta,\alpha)=\sum_{n=0}^{\infty} idn(\alpha)\cdot\cos(n\theta)$$

$$where \tag{3.9}$$

$$idn(\alpha)=\frac{1}{\pi}\cdot\int_{-\pi}^{\pi} i_{d\,rain}(\theta,\alpha)\cdot\cos(n\theta)d\theta\cdot\begin{cases}1/2 & if\ n=0\\ 1 & if\ n>0\end{cases}$$

Evaluating the DC and first harmonics of i_{drain} yields

$$id0(\alpha) = \frac{i_{drain,\max}}{\pi} \cdot \frac{\sin(\alpha) - \alpha \cdot \cos(\alpha)}{1 - \cos(\alpha)}$$

$$id1(\alpha) = \frac{i_{drain,\max}}{\pi} \cdot \frac{\alpha - \sin(\alpha) \cdot \cos(\alpha)}{1 - \cos(\alpha)} \tag{3.10}$$

The first few harmonic components of i_{drain} are computed and plotted in Figure 3-7

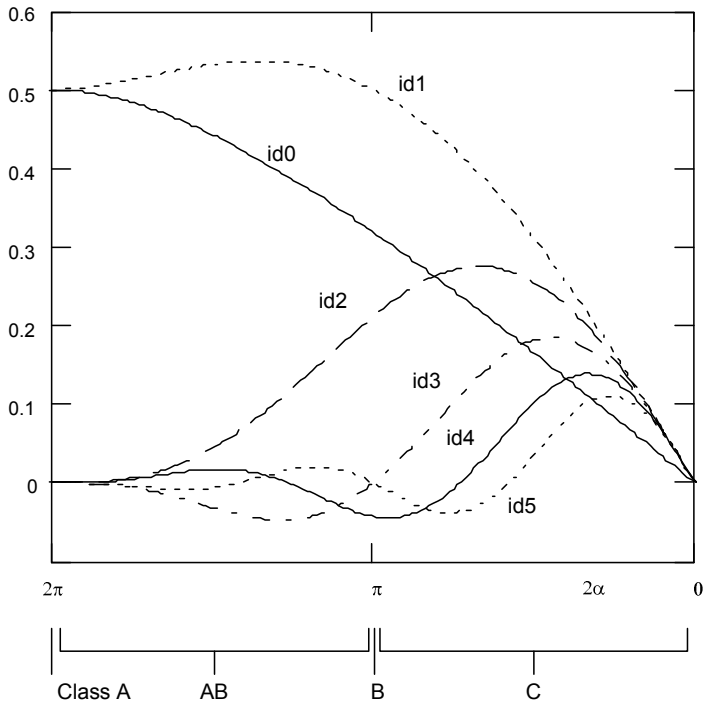


Figure 3-7 Fourier composition of i_{drain} as a function of the conduction angle (2α)

As shown in Figure 3-7, the DC component (*Id0*) is highest at Class-A bias and it decreases monotonically with the conduction angle. The fundamental component however increases slightly initially and then drops with the conduction angle. Evaluating (3.10) shows that the fundamental component at Class-B bias reaches the same value as Class-A bias:

$$id1(\pi) = id1(\pi/2) = \frac{I_{drain,\max}}{2} \tag{3.11}$$

A Class-B PA thus is capable of delivering the same amount of output power with a substantially lower DC drain current, which results in improved efficiency.

Since we have ignored the nonlinearity in the device gm, Figure 3-7 shows perfect linearity at Class-A bias. This is because the only non-zero Fourier components in the drain current waveform is the fundamental and DC components. As the conduction angle decreases, the high frequency components start to grow. As the conduction angle approaches zero, all harmonic components converges to the same level, which is consistent with the Fourier composition of an infinitely narrow pulse train. One may notice the simultaneous null of the 3rd and the 5th harmonic at the Class-B bias. This is however an artifact of the half sine wave I_{drain} waveform assumption and is not guaranteed in practice when the non-linear device gm is considered.

The Fourier composition of the drain current and voltage is a very important concept in optimizing PA efficiency. Figure 3-8 shows in time domain the Fourier decomposition of the drain current for both a Class-A bias and a Class-B bias PA. As we can see, the net result of a substantial presence of 2^{nd} and 4^{th} order harmonics in the Class-B bias is that it flattens the bottom and sharpens the peak of the overall i_{drain} waveform. The final result is a drain current waveform that has the same amount of fundamental component (and hence the same output power), but with a substantially lower DC level. The resulting PA is thus more efficient.

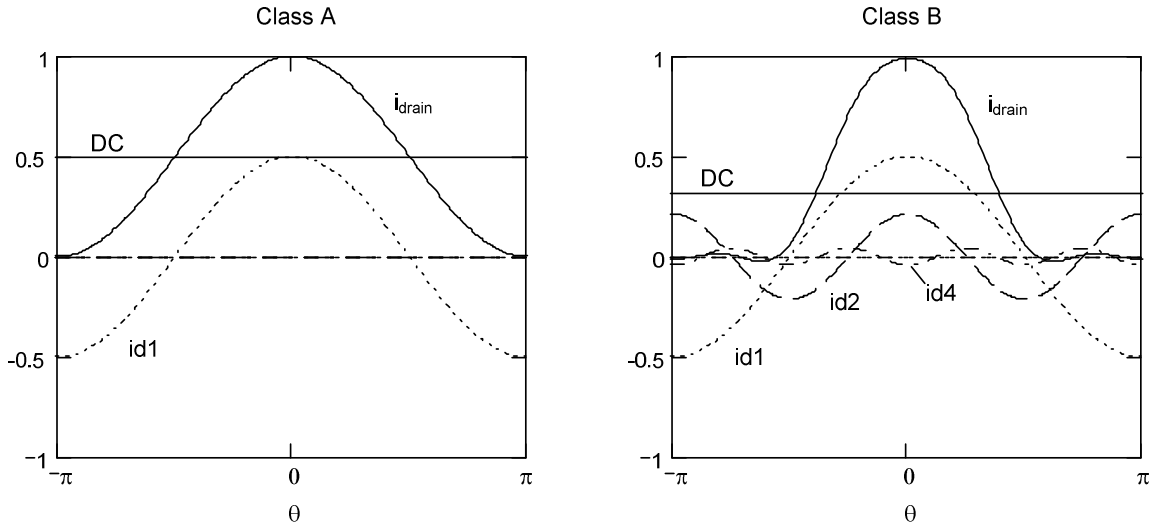


Figure 3-8 Time domain illustration of the I_{drain} harmonic composition for a Class-A and Class-B PA

Efficiency and Power Capacity vs. Conduction Angle

In the reduced conduction angle PA, each drain current harmonic component is terminated by well-defined impedance at the corresponding frequency: the termination impedance is zero at DC (due to the inductor), R_L at the fundamental frequency, and zero again at all higher order harmonics. Figure 3-9 depicts the load termination at each harmonic frequency.

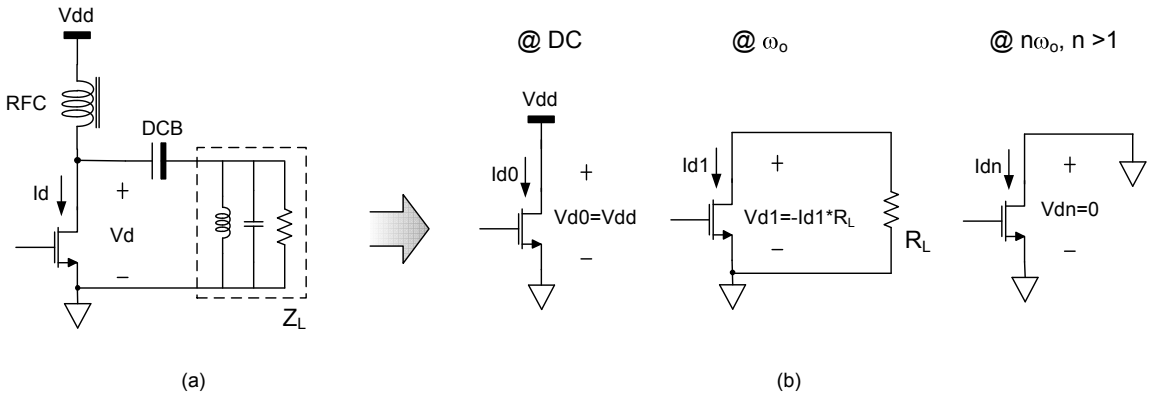


Figure 3-9 load termination at each critical frequencies for a Class-A,AB,B or C power amplifier

It is clear from Figure 3-9 that the DC component of the drain voltage is always V_{DD}, while the voltage swing at the fundamental frequency is determined by the resistive load R_L. In order to avoid the drain voltage from swinging below ground, we need to limit the

swing of v_{drain} by V_{DD}. In other word, R_L needs to be chosen according to the conduction angle such that it is bounded by the following limit:

$$R_L = \frac{V_{DD}}{id1} \tag{3.12}$$

In our discussion so far, we have implicitly assumed that the device knee voltage is negligible when compared with V_{DD}. If the knee voltage is high compared to V_{DD}, then R_L will have to be further reduced in order to avoid driving the transistor into triode. In general, the efficiency will be lowered by a factor of (1-V_{knee}/V_{DD}). This factor, unfortunately, is often quite significant in low voltage CMOS designs.

If R_L is chosen according to(3.12), then the DC power consumption, the output power delivered, and the drain efficiency can be expressed as:

$$P_{DC} = V_{DD} \cdot id0 \tag{3.13}$$

$$P_{out} = \frac{1}{2} id1^2 R_L = \frac{1}{2} V_{DD} id1 \tag{3.14}$$

$$\eta_{drain} = \frac{P_{out}}{P_{DC}} = \frac{1}{2} \frac{\alpha - \sin(\alpha)\cos(\alpha)}{\sin(\alpha) - \alpha\cos(\alpha)} \tag{3.15}$$

These quantities are plotted against the conduction angle in Figure 3-10. Note that efficiency increases monotonically with the decrease in conduction angle. The theoretical

efficiency is 50% at Class-A, 78.5% at Class-B and approaching 100% as the conduction angle approaches zero. Unfortunately, the apparent high efficiency at the Class-C bias is also at the expense of a diminishing deliverable output power. This is consistent with our observation from Figure 3-5 that reducing drain current pulse width at low conduction angle limits the amount of charge injected into the load network and hence reduces the deliverable output power.

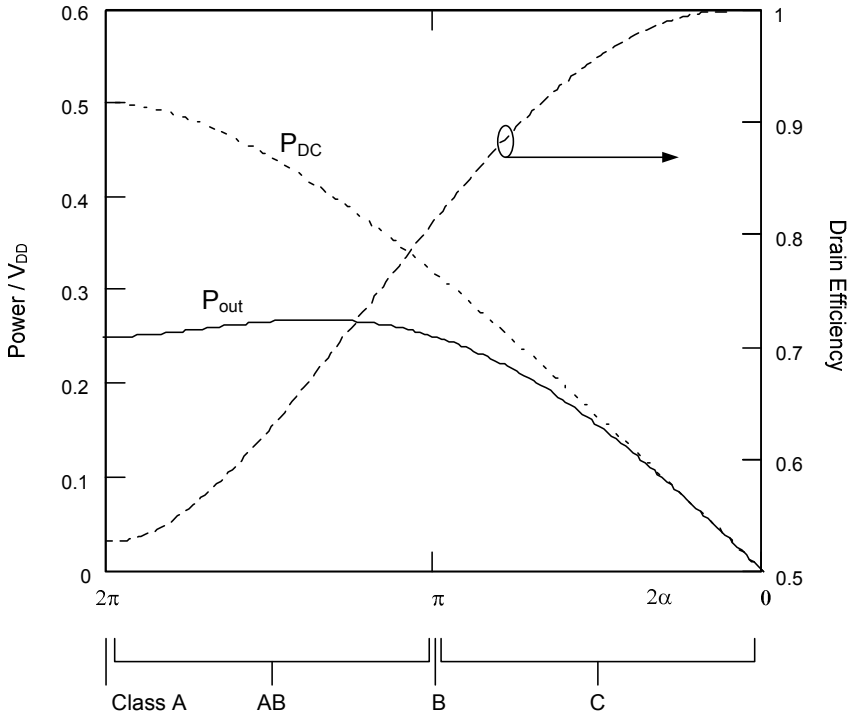


Figure 3-10 Output Power, DC power consumption, and efficiency vs. conduction angle

The diminishing deliverable output power at low conduction angle is also evident from calculating the power capacity as a function of conduction angle:

$$v_{drain,\max} = 2 \cdot V_{DD}$$
$$PC = \frac{P_{out}}{v_{drain,\max} \cdot i_{drain,\max}} = \frac{1}{4\pi} \cdot \frac{\alpha - \sin(\alpha) \cdot \cos(\alpha)}{1 - \cos(\alpha)} \tag{3.16}$$

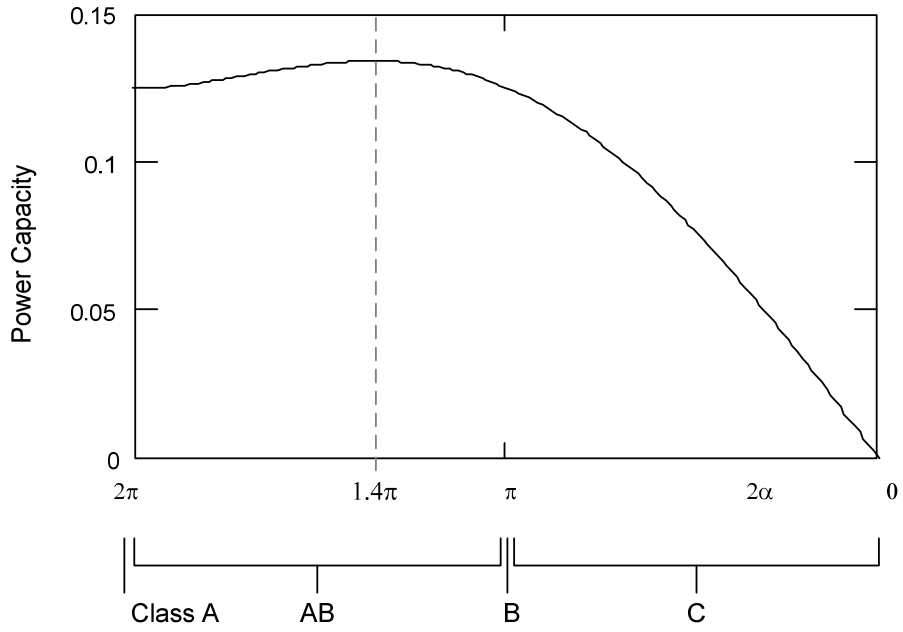


Figure 3-11 Power capacity vs. conduction angle

The power capacity is plotted against the conduction angle in Figure 3-11. Note that the power capacity is 1/8 at Class-A bias, and it reaches a peak at Class-AB bias with a

conduction angle of 1.4π. Beyond this point the power capacity drops monotonically with conduction angle. In other word, for a given supply voltage and device size (i.e. with fixed $V_{dain,max}$ and $i_{drain,max}$), the deliverable output power drops rapidly with conduction angle once it is biased beyond Class-B.

We have so far assumed a fixed device size in our comparison among different conduction angle bias. In principle, at small conduction angle, the device size can be scaled up to boost the peak drain current value. Doing this will partially compensate the diminishing power capacity. However, the degree of improvement is often as scaling up the device by a large factor inevitably introduces extra parasitic and additional losses. In addition, precise control of the conduction angle in the real circuit in the presence of process and temperature variation is quite challenging.

Effects of Reduced Input Amplitude

We have so far assumed that the PA is driven by the maximum input level that results in the maximum drain current ($I_{drain,max}$) allowable for the given device size. For constant-envelope modulated signals, this would be the preferred operation condition as it maximizes the efficiency and fully utilizes the available power capacity. There are, however, situations in which the PA needs to operate at reduced signal level, such as when the input signal is amplitude modulated or suppressed for power control purpose.

We will explore in this section the impacts on power efficiency when a Class-A, AB, B or C PA is driven with reduced signal level.

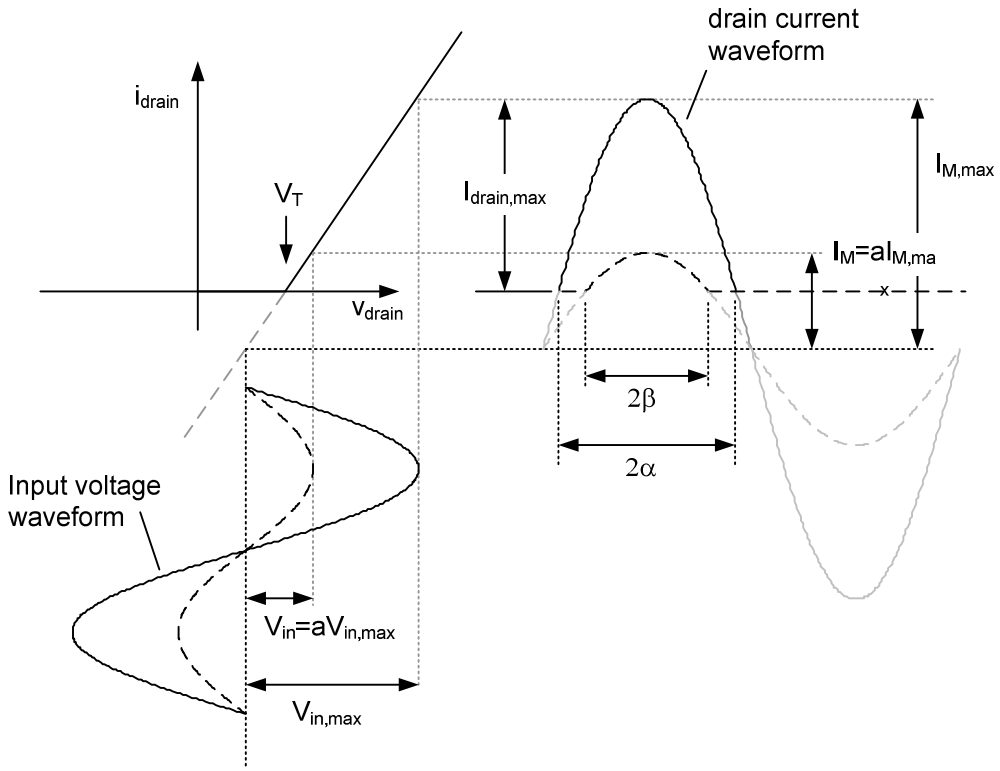


Figure 3-12 The active device input/output relationship at full and reduced signal levels

Figure 3-12 shows the transfer characteristic of the active device, and the input and output signal waveforms. For illustration purpose we have shown a Class-B bias, but the conclusion is independent of the conduction angle. The solid traces show the input

voltage and the drain current waveforms at their maximum level, while the dashed traces correspond to reduced input level. We defined the amplitude reduction fact (a) as:

$$a = \frac{V_{in}}{V_{in,\max}} = \frac{I_M}{I_{M,\max}} \tag{3.17}$$

where V_{in} is the input signal amplitude, and I_M is the drain current amplitude corresponding to the factitious case where there were no cut-off in the active device. From Figure 3-12, it is clear that the conduction angle is in fact a function of the input signal level. The actual conduction angle (β) as a function of the input signal level can be expressed as:

$$\beta = \begin{cases} \cos^{-1}\left(\frac{\cos(\alpha)}{a}\right) & if \quad a \geq \left|\cos(\alpha)\right| \\ \pi & if \quad -a > \cos(\alpha) \\ 0 & if \quad a < \cos(\alpha) \end{cases} \tag{3.18}$$

In Figure 3-13, β is plotted against the amplitude reduction factor, a, for a number of nominal conduction angles, 2α, (i.e. the conduction angle at maximum input level, in other word, α corresponds to the value of β at a=1.) As the input amplitude decreases (a:1→0), the conduction angle of a nominally Class-AB biased PA *increases* and eventual reaches Class-A bias at $a=\cos(\alpha)$. However, the conduction angle of a nominally Class-C biased PA *decreases* as the input signal is reduced until eventually it is

completely cut off at $a=cos(\alpha)$. On the other hand, the conduction angle of a Class-A or Class-B PA is independent of the input driving signal level.

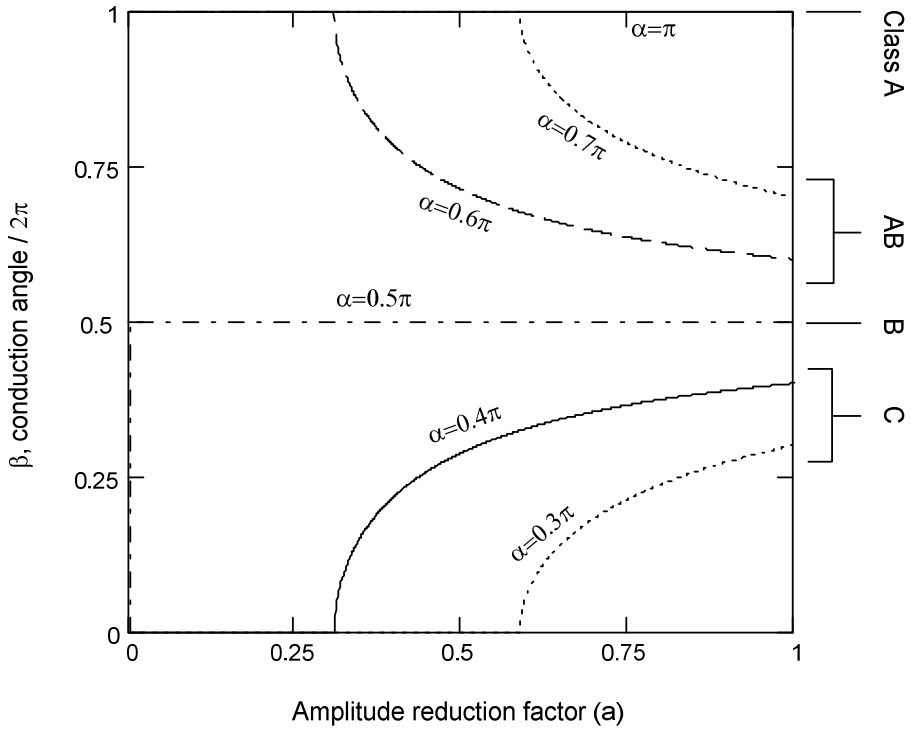


Figure 3-13 Conduction angle as a function of the amplitude reduction factor

The drain current waveform under reduced input signal level, $i_{drain,reduced}$, can be expressed as:

$$i_{drain,reduced} = i_{drain,\max} \cdot \frac{a \cdot \cos(\theta) - \cos(\alpha)}{1 - \cos(\alpha)} \tag{3.19}$$

Taking Fourier series expansion yields the DC and fundamental components of $i_{drain,reduced}$:

$$\begin{aligned} id0_{reduced}(\beta) &= \frac{I_{drain,\max}}{\pi} \cdot \frac{(a \cdot \sin(\beta) - \beta \cdot \cos(\alpha))}{(1 - \cos(\alpha))} \\ id1_{reduced}(\beta) &= \frac{I_{drain,\max}}{\pi} \cdot \frac{(a \cdot \beta - \sin(\beta)(2\cos(\alpha) - a \cdot \cos(\beta)))}{(1 - \cos(\alpha))} \end{aligned} \tag{3.20}$$

$Id0_{reduced}$ and $id1_{reduced}$ at a number of nominal conduction angles, 2α, are plotted against the amplitude reduction factor in Figure 3-14 and Figure 3-15. As shown in Figure 3-14, the DC components of $i_{drain,reduced}$ generally decreases with the input amplitude. As expected, Class-A bias draws a constant DC drain current regardless of the actual signal swing. The PAs with a nominal Class-AB bias eventually reaches a constant level when the amplitude reduction factor is below $\cos(\alpha)$ and become Class-A biased. The DC component for Class-B biased PA is linear with the input signal amplitude, and reaches zero at zero input signal swing – this is consistent with the Class-B bias in which the input bias point is set at the device threshold level. The DC drain current of nominal Class-C PAs decreases sharply and becomes zero when the amplitude reduction factor reaches $\cos(\alpha)$ where the input swing is so small that the device is always in the cut off region.

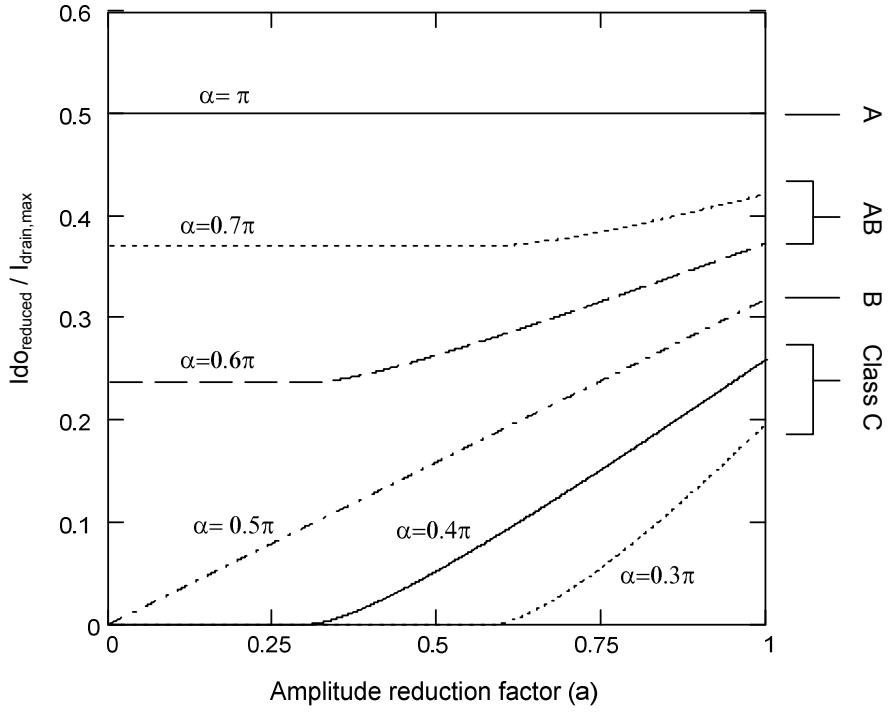


Figure 3-14 DC component of drain current as a function of input signal level

Even though we have ignored the inherent device non-linearity and assumed a constant device transconductance, it is obvious from Figure 3-12 that the input signal amplitude dependent conduction angle itself can cause significant non-linearity. This in fact is evident from Figure 3-15 which shows the fundamental component of $i_{drain,reduced}$ at a number of nominal conduction angle, 2α, as a function of the amplitude reduction factor.

As can be seen from Figure 3-15, PAs with a nominal Class-AB bias exhibits gain compression because the conduction angle decreases with input signal amplitude. On the

other hand, PAs with nominal Class-C bias shows gain expansion because the conduction angle increases with signal level. Since the conduction angle is constant for a Class-A or Class-B biased PA, the output signal swing is linear with the input.

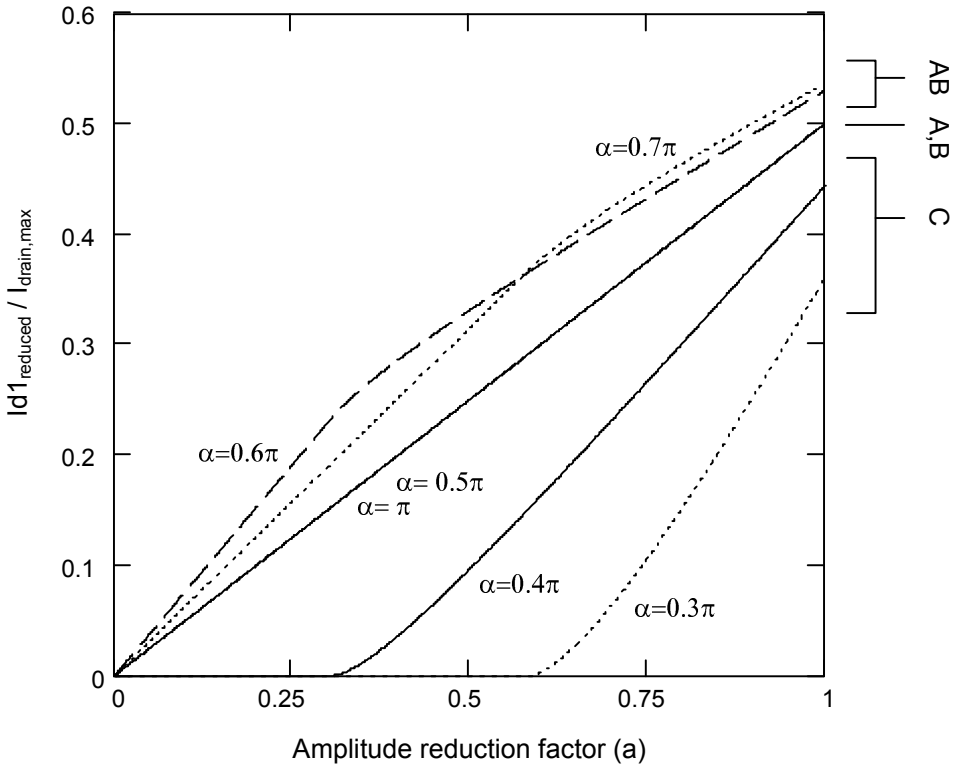


Figure 3-15 Fundamental component of drain current as a function of input signal level

The overall drain efficiency as a function of input signal level can now be evaluated as:

$$\eta_{drain} = \frac{1}{2} \cdot \frac{id1_{reduced}{}^2 \cdot R_L}{V_{DD} \cdot ido_{reduced}} = \frac{1}{2} \cdot \left(\frac{id1_{reduced}}{id0_{reduced}} \right) \cdot \left(\frac{id1_{reduced}}{id1} \right)$$
$$= \frac{1}{2} \cdot \left(\frac{id1_{reduced}}{id0_{reduced}} \right) \cdot \left(\frac{a \cdot \beta - \sin(\beta) \cdot (2 \cdot \cos(\alpha) - a \cdot \cos(\beta))}{\alpha - \sin(\alpha) \cdot \cos(\alpha)} \right) \quad (3.21)$$

Note that in the above expression, we assume the same value of the load resistance R_L as previously defined in (3.12). Clearly R_L has to be fixed at a value to ensure that drain voltage swing is confined within 0-2V_{DD} at the maximum possible signal level. The drain efficiency at a number of conduction angles, 2α, as a function of the amplitude reduction factor is plotted in Figure 3-16. The efficiency value at a=1 corresponds to the maximum possible drain efficiency at the corresponding nominal bias condition, and is consistent with what is shown previously in Figure 3-10. Although nominal Class-C bias achieves the highest efficiency at maximum signal level, Class-B biased PA offers higher efficiency over a wide range of input signal level. This together with the better linearity of a Class-B PA suggests that it is more preferable in applications that process signals with deep peak-to-average ratios, and over wide range of output power levels.

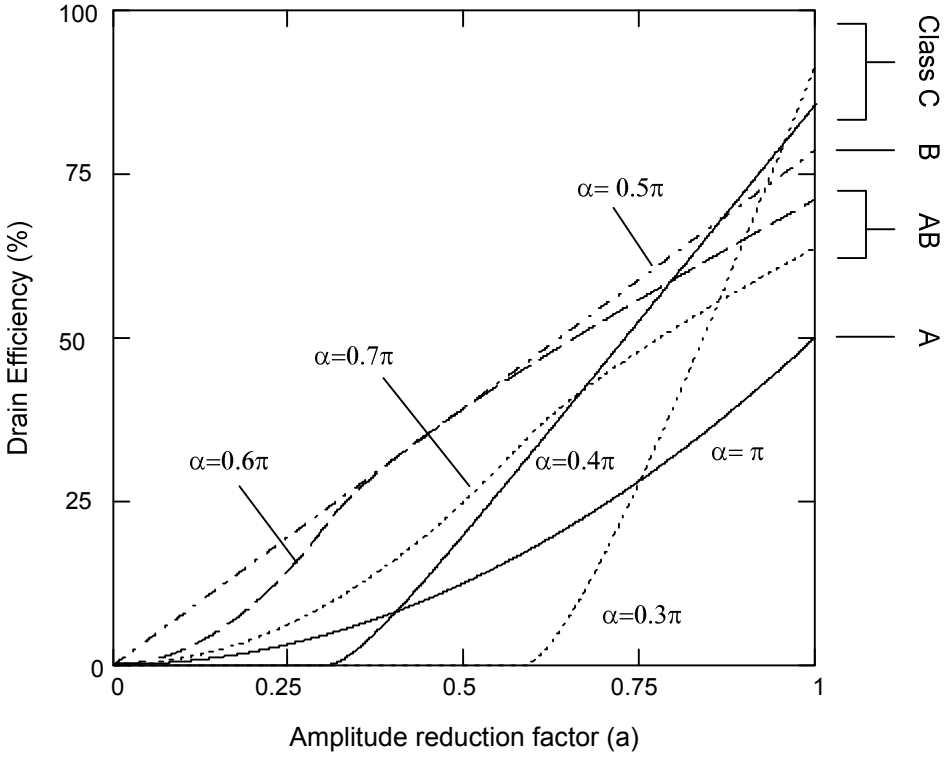


Figure 3-16 Drain efficiency dependency on input signal level

3.5 Class-F1 Power Amplifier

Class-A and reduced conduction angle PAs adopt a single-resonant load network that shunts all non-fundamental harmonics of the drain current to ground. As a result, the drain voltage is confined to a simple sinusoid. Although efficiency can be improved by reducing the conduction angle, the actual enhancement is quickly hampered by the severe drop in power capacity. One way to break away from this direct trade-off between

efficiency and power capacity is to introduce additional degree of freedom in the harmonic terminations of the load network; and such is the principle behind the Class-F PAs.

By adopting a multiple-resonant load network, a Class-F PA controls the harmonic contents of the drain voltage and current, and shapes them in a way to minimize the device power dissipation. The key advantage is that both efficiency and power capacity can be enhanced simultaneously. The term Class-F operation generally refers to the use of multiple-resonant load network. Multiple basic types exist, and we will discuss the Class-F1 type in this chapter, and bring up the Class-F2 type in the next chapter when we discuss switch-mode PAs.

Figure 3-17 shows a possible realization of a Class-F1 PA; the active device is driven in the same way as a Class-B PA, but the load network now includes two resonant tanks. One tank (L1-C1) is tuned to the fundamental frequency, while the other one (L3-C3) is tuned to the 3^{rd} harmonic. The L3-C3 resonator also serves to decouple the drain node from the output and prevent high frequency harmonics from leaking into the output and causes interference. The resulting characteristic waveforms are also shown in Figure 3-17

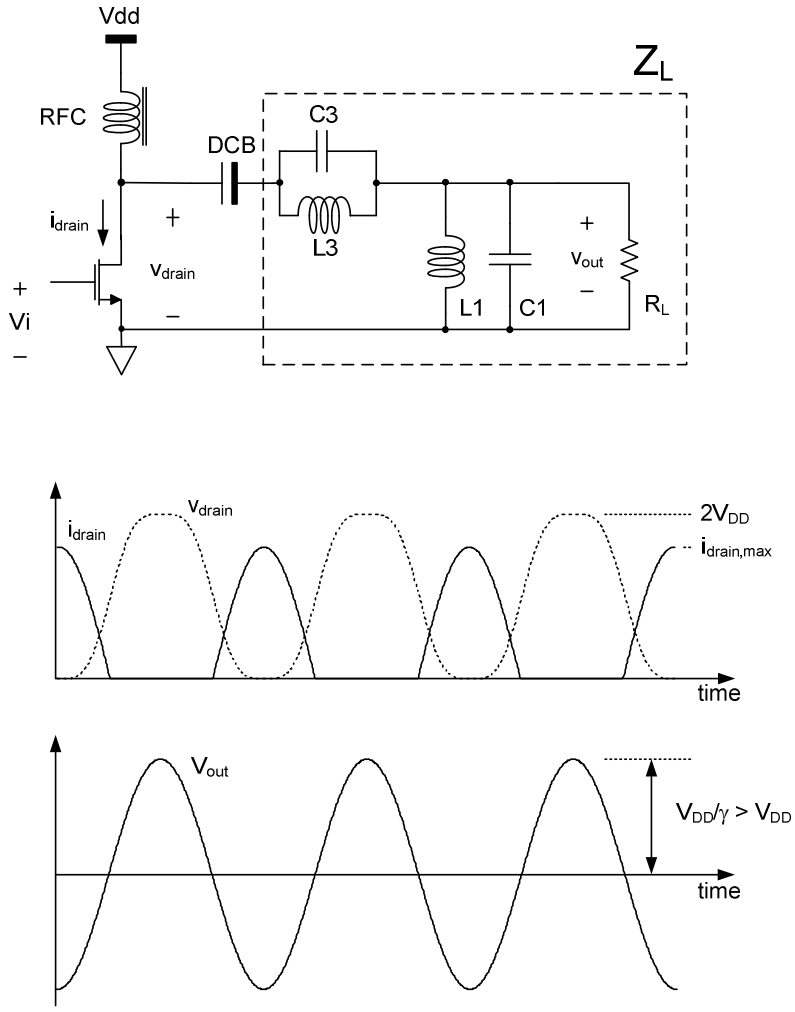


Figure 3-17 Class-F1 PA and its drain current/voltage and output waveforms

If the quality factors (Q) of the LC tanks are sufficiently high, the impedance across them is infinite at their resonant frequencies but zero anywhere else. In this case, the drain termination can be analyzed in the way shown in Figure 3-18. At DC, the drain is short to

V_{DD} through the RFC. At the operating frequency ω_o, tank L1-C1 is open, but tank L3-C3 is shorted, such that the drain is terminated with impedance R_L. At $3\omega_o$, tank L3-C3 is open and thus presents high impedance to the drain. Finally, at any other harmonics the drain node is effectively shorted to AC ground.

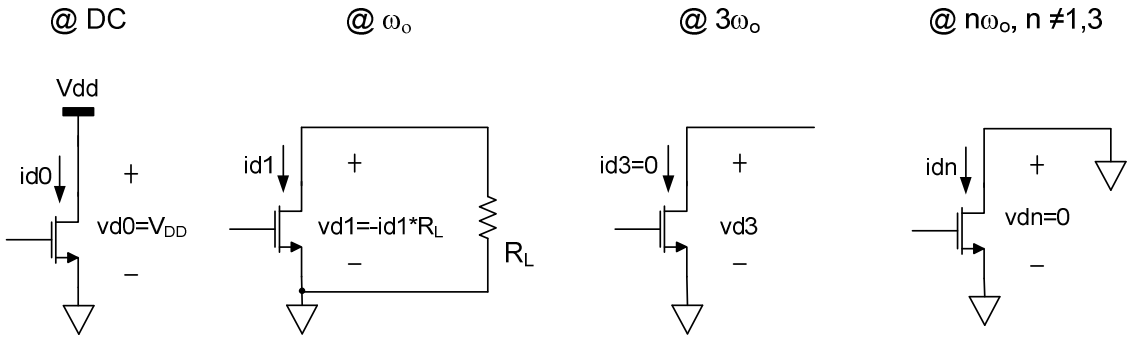


Figure 3-18 Drain termination of a Class-F1 PA at various harmonics

As a result of the load harmonic terminations, we know the DC component of the drain voltage will remain at $V_{DD,}$ and its fundamental component will be equal to $R_L \cdot id1$ *(id1* is the fundamental component of the drain current.) However, the distinction is that a small amount of 3rd harmonic component can now also be present in drain voltage because of the high impedance termination at that frequency. The drain voltage waveform can therefore be expressed in the following form:

$$v_{drain} = V_{DD} + vd1 \cdot (\cos(\theta) - x \cdot \cos(3\theta)) \tag{3.22}$$

where x is the ratio between the 3rd harmonic amplitude and the fundamental.

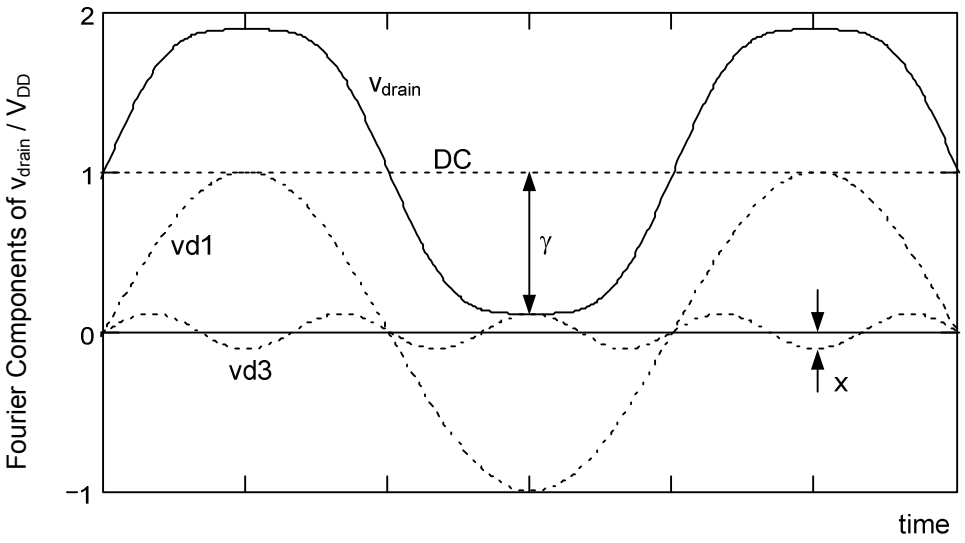


Figure 3-19 Fourier decomposition of the drain voltage waveform for a Class-F1 PA with x=1/9

Figure 3-19 shows an example of the composite v_{drain} waveform and its Fourier decomposition. This waveform corresponds to a 3rd harmonic component x = 1/9, which results in a maximally flat v_{drain}. As shown in the figure, the key effect of prescribing a small amount of 3rd order harmonic is that it compresses the v_{drain} waveform with respect to the amplitude of its fundamental component. The resulting v_{drain} waveform is flattened

on top and bottom, resulting in a relatively square appearance. The compression factor (γ) is defined as the ratio between the zero-peak v_{drain} voltage swing and the amplitude of its fundamental component. It can be defined as a function of the 3rd harmonic component x:

$$\gamma = \frac{v_{drain,z-pk}}{vd1} = \begin{cases} 1-x & if \quad x \le \frac{1}{9} \\ \frac{1}{3}\sqrt{\frac{(1+3x)^3}{3x}} & if \quad x > \frac{1}{9} \end{cases} \tag{3.23}$$

In Figure 3-20, we plot the value of the compression factor against the 3rd order harmonic component x. Note the x=1/9 provides the maximally flat v_{drain} waveform, whereas x=1/6 results in the minimum compression factor with a slight amount of overshoot.

Since the drain voltage swing is compressed with respect to the fundamental component in a Class-F1 PA, this PA can accommodate a fundamental component that is $1/\gamma$ times larger than V_{DD}. If we now scale up the value of R_L by a factor $1/\gamma$ with respect to the Class-B value, the resulting PA will reach the same maximum drain voltage limit, but produce an output power that is $1/\gamma$ times larger.

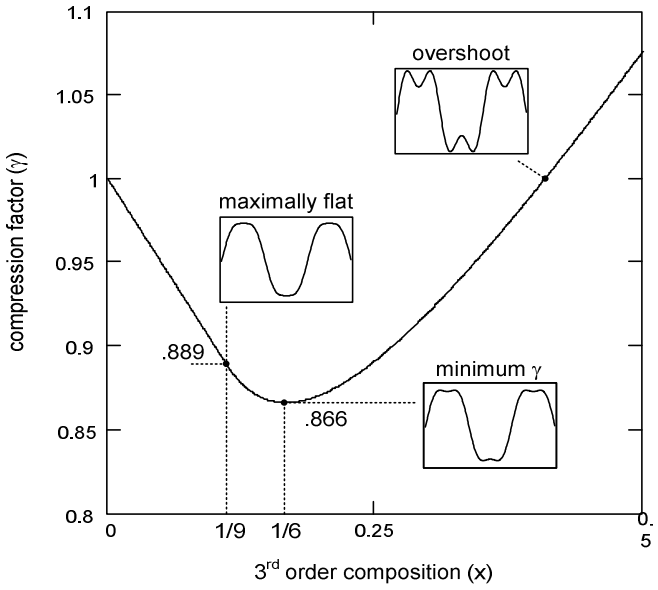


Figure 3-20 Compression factor vs. 3rd order composition

We expect the efficiency of the class F PA to be better than Class-B because the compressed drain voltage waveform has less overlap with the drain current and thus results in less power loss in the device. We can deduce the amount of efficiency enhancement by recognizing the following: since the i_{drain} waveform remains identical to that of the Class-B PA, the DC power consumption ($=V_{DD}\cdot ido$) remains unchanged. However, the output power is increased by a factor of $1/\gamma$. For γ=1/9, the overall efficiency thus becomes:

$$\eta_{classF1} = \frac{\eta_{classB}}{\gamma} = 88.3\% \tag{3.24}$$

In addition, power capacity of a Class-F1 PA is also increased by same factor with respect to the Class-B PA:

$$PC_{classF1} = \frac{PC_{classB}}{\gamma} = 0.141 \tag{3.25}$$

In conclusion, by introducing a small amount of 3rd order harmonic into the drain voltage waveform, we can simultaneously enhance both efficiency and power capacity. In fact, as we see in subsequent chapters, load harmonic termination engineering is the key concept behind many types of high efficiency PAs.

One potential drawback of the Class-F1 PA is the difficulty in precisely controlling the amount of 3rd harmonic that is introduced into the drain voltage waveform. In fact the harmonic composition of the drain current that is injected into the load network is usually quite sensitive to the transconductance characteristic of the active device and is not very well controlled. In next chapter, we will see how the harmonic tuning concept can be realized more robustly in the Class-F2 PA, where the desired waveform shaping is achieved by operating the transistor as a switching device rather than as a transconductor.

Chapter 4 - Switch-mode Power Amplifiers

4.1 Switch-mode vs. Transconductance-based PAs

It is evident from the discussion on Class-F1 PA that drain voltage waveform shaping is an effective way to enhance an amplifier's efficiency. By reducing the transistor's output voltage and current overlapping time, power dissipation at the transistor is minimized and supply power can be more efficiently delivered to the output load. Achieving such optimal waveform in a gm-based PA is however not straight forward. Since the transistor acting as a high-output-impedance transconductor, it does not have direct control over its drain voltage. In fact, the drain voltage is controlled by both the harmonic composition of the transistor output current and the harmonic termination of the load network. The

former is particularly difficult to control precisely as it is sensitive to the non-linear characteristic of the transistor.

If the input signal is constant-envelope modulated and features fixed amplitude with only time varying frequency or phase, a much more effective way to achieve the optimal drain voltage waveform is to incorporate a switch-mode operation. In a switch-mode PA, the transistor acts as a switching device which toggles between an ON-state and an OFF-state with approximately 50% duty cycle. During the ON-state the device is in the triode region and it can be modeled as a small on-resistance R_{ON}. During the OFF-state the device is cut off and can be modeled as an open circuit.

One key distinction from gm-based PAs is that, in a switch-mode PA the only function of the input signal is to synchronize switching of the transistor so as to pass the frequency and phase modulation to the output. When the transistor is turned on, it shorts the drain to ground - the transistor thus directly controls the drain voltage for about half of the operation cycle. R_{ON} is relatively insensitive to the input signal amplitude (as long as it is large enough to turn the transistor on) and much smaller than the input impedance of the matching network. This results in a robust operation in which the drain current and voltage waveforms are determined solely by the matching network termination but insensitive to either the transistor characteristic or the input signal amplitude.

In this chapter, we examine two basic switch-mode PAs, namely Class-F2 and Class-D [Crip99].

4.2 Class-F2 power amplifier

Figure 4-1 shows one possible implementation of a Class-F2 PA. It is similar to the Class-F1 PA except the transistor now acts as a toggling switch rather than a transconductor. In addition, the 3rd order harmonic resonator is replaced by a quarter wavelength ($\lambda/4$) transmission line serving as an impedance transformer between the output and the drain node. At DC the drain is shorted to the supply by the RFC. At the operating frequency ω_o, where the tank L1-C1 is open and the length of the transmission line is $\lambda/4$, the load impedance R_L is transformed to Z_o^2/R_L, where Z_o is the characteristic impedance of the transmission line.

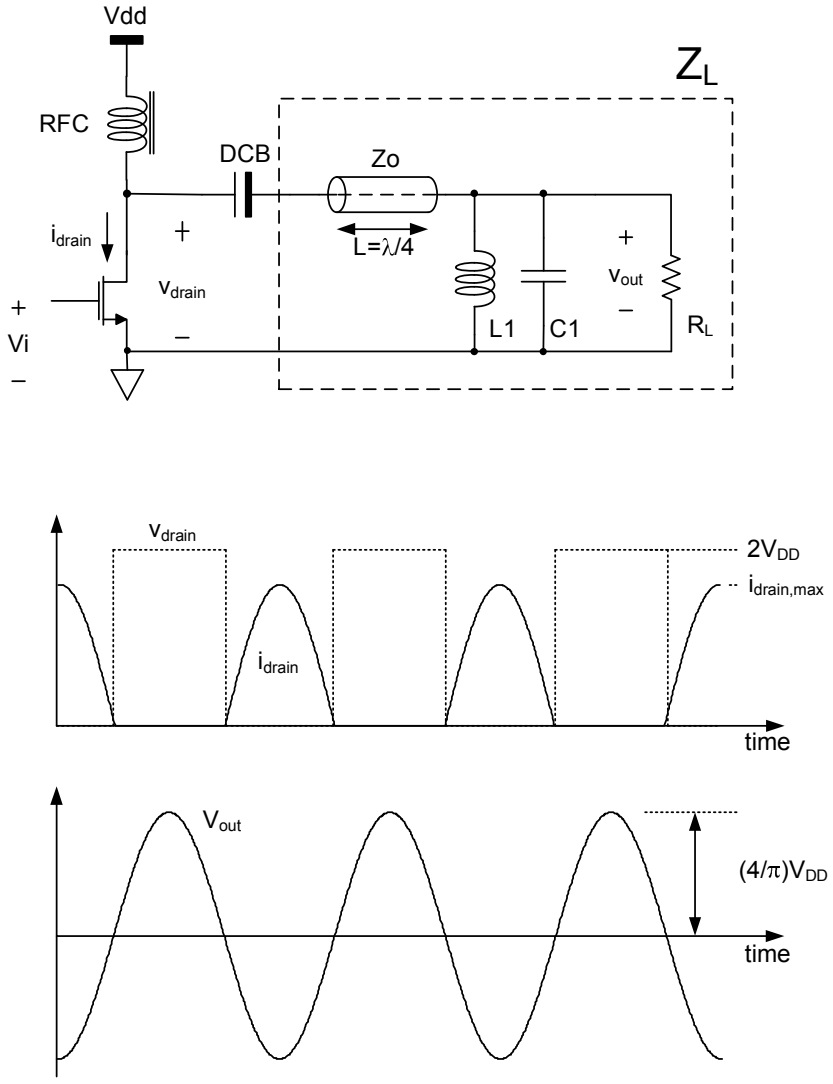


Figure 4-1 A Class-F2 PA and its characteristic drain and output waveforms

With the usual high-Q assumption, the resonator L1-C1 is shorted at all non-fundamental harmonics. However, the short at output is transformed to different drain termination

depending on the order of the harmonic. At even harmonics where the length of transmission line is a multiple of λ/2, the drain is terminated to ground. However, at any odd harmonics where the length of transmission line is an odd multiple of λ/4, the short at the output is transformed to an open termination at the drain. In summary,

$$R_{drain,n} = \begin{cases} 0 & if \quad n = 0 \\ Z_o^{\,2} / R_L & if \quad n = 1 \\ 0 & if \quad n = 2k \\ \infty & if \quad n = 2k+1 \end{cases} \tag{4.1}$$

From (4.1), we know that the DC component of v_{drain} is V_{DD}, and all even harmonics are zero because of the short termination at these frequencies. In addition, all odd order harmonics of the drain current i_{drain} are also zero because of the open termination. However, at the fundamental components of both v_{drain} and i_{drain} are finite and they are related by the ratio of $R_{drain,1} = Z_o^2/R_L$. Considering these properties, we can express v_{drain} and i_{drain} in the following functional forms:

$$\begin{aligned} v_{drain}(\theta) &= V_{DD} + \sum_{n \in 2k+1} vdn \cdot \cos(n \cdot \theta + \phi_n) \\ i_{drain}(\theta) &= ido + id1 \cdot \cos(\theta + \phi_1) + \sum_{n \in 2k} idn \cdot \cos(n \cdot \theta + \phi_n) \end{aligned} \tag{4.2}$$

where $\theta = \omega_o t$ is the phase angle, and k is any non-negative integer. The fundamental component of i_{drain} is

$$id1 = -vd1 \cdot (R_L / Z_o^{\,2}) \tag{4.3}$$

Note that the negative sign indicates that direction of the current is out of the load network and into the drain of the active device.

Due to the ON/OFF switching operation of the transistor, we already have knowledge on the i_{drain} and v_{drain} waveforms for part of the cycle, namely

$$\begin{aligned} &during\ OFF\ state && i_{drain} = 0 \\ &during\ ON\ state && v_{drain} = i_{drain} \cdot R_{ON} \end{aligned} \tag{4.4}$$

Base on (4.2), we can derive one half cycle of the drain voltage and current waveforms from the other halve:

$$\begin{aligned} v_{drain}(\theta - \pi) &= V_{DD} + \sum_{n \in 2k+1} vdn \cdot \cos(n \cdot \theta + \phi_n - n \cdot \pi) \\ &= V_{DD} - \sum_{n \in 2k+1} vdn \cdot \cos(n \cdot \theta + \phi_n) \\ &= 2V_{DD} - v_{drain}(\theta) \end{aligned} \tag{4.5}$$

$$\begin{aligned} i_{drain}(\theta - \pi) &= ido + id1 \cdot \cos(\theta + \phi_1 - \pi) + \sum_{n \in 2k} idn \cdot \cos(n \cdot \theta + \phi_n - n \cdot \pi) \\ &= ido - id1 \cdot \cos(\theta + \phi_1) + \sum_{n \in 2k} idn \cdot \cos(n \cdot \theta + \phi_n) \\ &= i_{drain}(\theta) - 2 \cdot id1 \cdot \cos(\theta + \phi_1) \end{aligned} \tag{4.6}$$

If i_{drain}=0 during the OFF-state, then (4.6) dedicates i_{drain} will be a half sinusoid during the ON state. Based on this we can also deduce v_{drain} during the ON-state according to (4.4). With v_{drain} during the ON-state known, we can now deduce v_{drain} during the OFF state from equation (4.5). Finally, the amplitude of i_{drain} during ON state is related to the fundamental component of v_{drain} according to (4.3). Following this procedure, we can derive the complete drain voltage and current waveforms as shown in Figure 4-2.

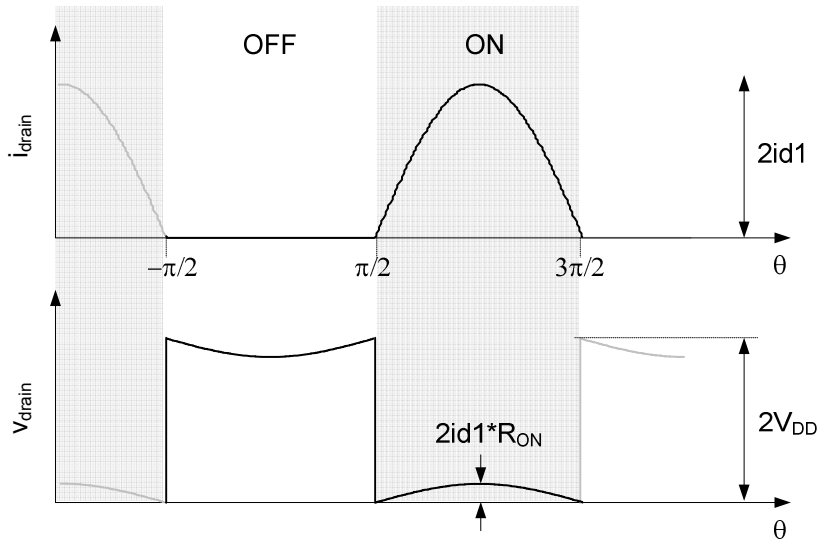


Figure 4-2 Characteristic waveforms of a Class-F2 PA

Taking Fourier transforms of the above waveforms, the fundamental component of v_{drain}, and i_{drain}, as well as the DC component of i_{drain} can be expressed as:

$$\begin{aligned} vd1 &= \frac{4V_{DD}}{\pi}\left(\frac{1}{1+2R_{ON}R_L / Z_o^{\,2}}\right) \\ \phi_1 &= 0 \\ id1 &= vd1 \cdot \left(\frac{R_L}{Z_o^{\,2}}\right) \\ id0 &= \frac{2}{\pi} \cdot id1 \end{aligned} \tag{4.7}$$

Following equation (4.7), output power, DC power consumption and the drain efficiency can be expressed as:

$$P_{out} = \frac{1}{2} vd1^2 \cdot \frac{R_L}{Z_o^{\,2}} = \frac{8V_{DD}^{\,2}}{\pi^2}\frac{R_L}{Z_o^{\,2}}\eta^2 \tag{4.8}$$

$$P_{DC} = V_{DD} \cdot ido = \frac{8V_{DD}^{\,2}}{\pi^2}\frac{R_L}{Z_o^{\,2}}\eta \tag{4.9}$$

$$\eta = \frac{P_{out}}{P_{DC}} = \left(\frac{1}{1+2R_{ON}R_L / Z_o^{\,2}}\right) \tag{4.10}$$

Note that the efficiency depends on the ratio between the R_{ON} of the device and the load impedance as seems at the drain (Z_o^2/R_L). This in fact is generally true for all switch-mode power amplifiers. At R_{ON}=0, that the theoretical drain efficiency is 100%. This is

consistent with Figure 4-1 since there would be no overlap between the drain voltage and current waveforms, resulting in zero power lost in the device. Furthermore, given that $I_{drain,max}=2id1$ and $V_{drain,max}=2V_{DD}$, the power capacity becomes

$$PC = \frac{P_{out}}{v_{drain,\max} \cdot i_{drain,\max}} = \frac{\eta^2}{2\pi} \leq 0.159 \tag{4.11}$$

where the upper bound corresponds to the limiting case where $R_{ON}=0$.

The device R_{ON} is by far the largest determining factor for the overall efficiency, but other effects may also degrade the achievable efficiency. One such effect is the power loss due to charging and discharging of the drain capacitance (C_{drain}), which can be expressed as

$$P_{C_D} = \frac{1}{2} C_{drain} (2V_{DD})^2 f_o \tag{4.12}$$

where f_o is the operation frequency. Equation (4.12) shows that the power loss in the drain capacitance is $2mW / pF \cdot V^2 \cdot GHz$. At multi-GHz frequency operation, with typical V_{DD} and C_{drain} values, this power can easily become a significant portion of the total power loss.

Equation (4.12) seems to suggest that the drain capacitance loss would scale down quickly with the supply voltage. Unfortunately, this is usually not the case in reality:

Lowering V_{DD} often demands a reduction in the load impedance (proportional to $1/V_{DD}^2$) so as to preserve the deliverable output power. From (4.8), it is clear that, in order to preserve the efficiency, R_{ON} needs to be scaled down accordingly. For any given device technology, $R_{ON} \cdot C_{drain}$ is approximately constant since C_{drain} scales proportionally, while R_{ON} is inverse proportional to device size. As a result, the effect of reducing V_{DD} is eventually cancelled by the increase C_{drain}, and in the end the power loss due to drain capacitance switching is roughly unaffected. Besides introducing switching additional loss, C_{drain} also shunts the drain node at high frequency, and thus undermines the high odd order harmonic termination established by the load network. Achievable efficiency at high frequency is thus further degraded.

4.3 Class-D Power Amplifier

A Class-D PA directly controls the drain voltage at both phase of the operation cycle. A single-ended version can be realized with complementary devices (e.g. PMOS and NMOS), but they are rarely used at high frequency due to the relatively poor high-speed performance of p-type devices.

At RF frequencies, transformer coupled differential Class-D PA using only n-type devices typically offers better performance. One possible implementation is shown in Figure 4-3. Here transistors M1 and M2 are driven by a differential input signal and toggle alternately between ON and OFF states with 50% duty cycle. The center-tapped

transformer converts the differential drain signal into the single-ended output. It also helps to shape the drain voltage into a square wave, and acts as an impedance transformer between the drain and the load network. The series resonant tank (L1-C1) is tuned to the operation frequency. For now we assume that this tank is high Q so only the fundamental component of drain voltage will pass to the output. We will subsequently relax this assumption and examine its effect on efficiency.

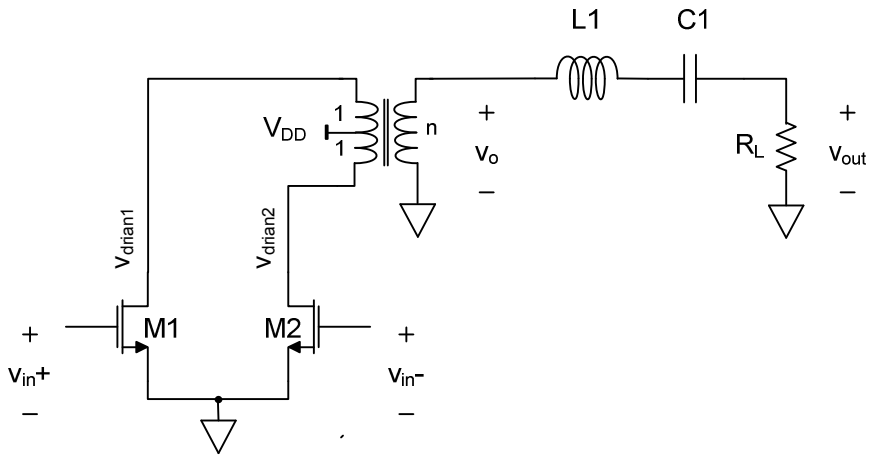
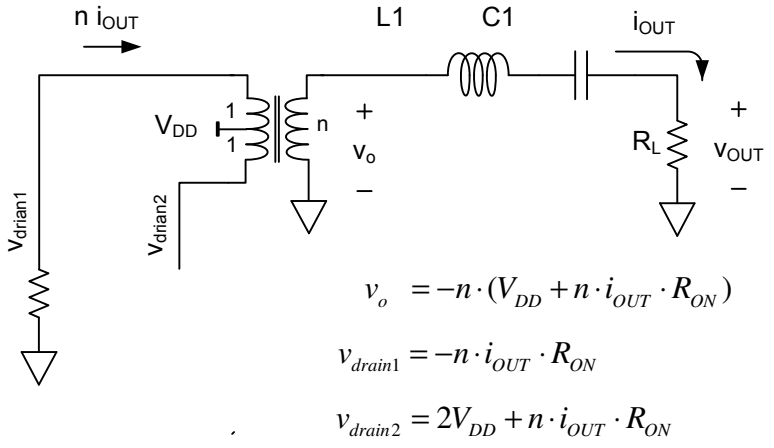


Figure 4-3 One possible implementation of a Class-D PA

The behavior of the circuit can be analyzed by breaking the operation cycle into two phases. In each phase, one transistor is turned on while the other is off. We model the on

device as a constant resistor R_{ON}, and the off device simply as an open circuit. The current and voltage relations during the two phases are illustrated in Figure 4-4.

(a) Phase I ($-\pi<\theta<0$), M1=ON, M2=OFF

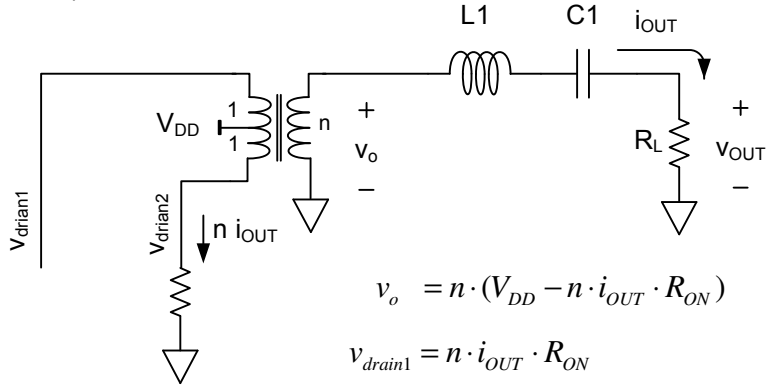


$$v_o = -n \cdot (V_{DD} + n \cdot i_{OUT} \cdot R_{ON})$$

$$v_{drain1} = -n \cdot i_{OUT} \cdot R_{ON}$$

$$v_{drain2} = 2V_{DD} + n \cdot i_{OUT} \cdot R_{ON}$$

(b) Phase II ($0<\theta<\pi$), M1=OFF, M2=ON



$$v_o = n \cdot (V_{DD} - n \cdot i_{OUT} \cdot R_{ON})$$

$$v_{drain1} = n \cdot i_{OUT} \cdot R_{ON}$$

$$v_{drain2} = 2V_{DD} - n \cdot i_{OUT} \cdot R_{ON}$$

Figure 4-4 I-V relations of a Class-D PA during the two opposite phases

During each phase, one of the two drain nodes is pull low to a level = $n \cdot i_{OUT} \cdot R_{ON}$, where R_{ON} is the on-resistance, i_{OUT} is the output current, and n is the transformer turn ratio. Assuming the transformer is lossless so that voltage across the primary coils is equal, the other drain node will be pushed to $2V_{DD} - n \cdot i_{OUT} \cdot R_{ON}$. In addition, the voltage establish on the primary side of the transformer is reflected to the secondary side with gain of n.

With a high Q tank L1-C1, the output current can be expressed in the form

$$i_{out} = I_o \cdot \sin(\theta + \phi) \tag{4.13}$$

where $\theta = \omega_o t$ is the phase angle. From (4.13) and the equations listed in Figure 4-3, we can derive the voltage across the secondary coil of the transformer as:

$$v_o(\theta) = n \cdot V_{DD} \cdot \Pi(\theta) - n^2 I_o R_{ON} \sin(\theta + \phi) \tag{4.14}$$

where $\Pi(\theta)$ is a square wave function:

$$\Pi(\theta) = \begin{cases} +1 & if\ (2n-1)\pi < \theta < 2n\pi \quad (\text{phase I}) \\ -1 & if\ 2n\pi \le \theta < (2n+1)\pi \quad (\text{phase II}) \end{cases} \tag{4.15}$$

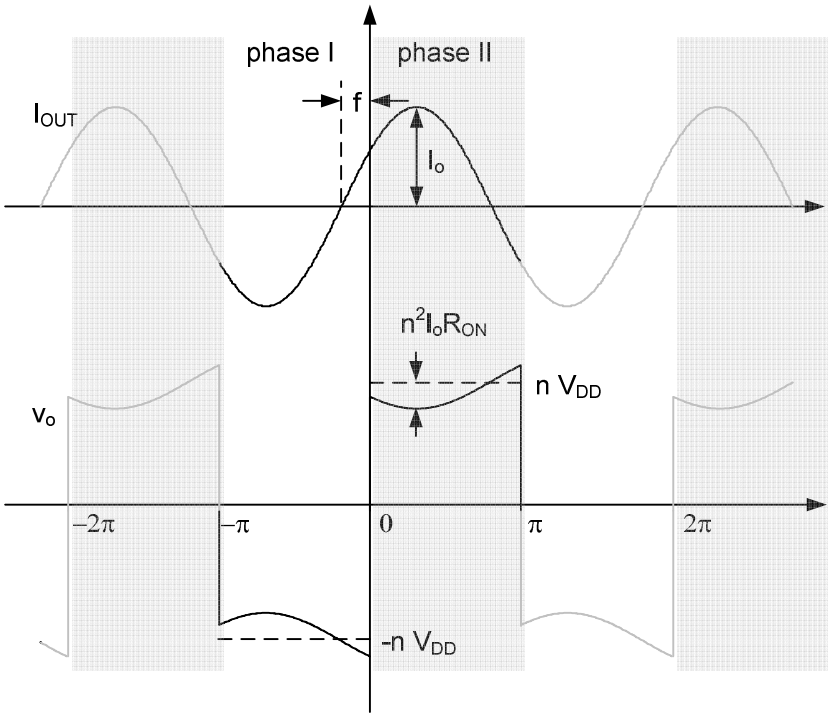


Figure 4-5 waveform of output current i_{OUT} and secondary coil voltage v_o

The waveforms of i_{OUT} and v_o are plotted in Figure 4-5. The fundamental component of v_o can be obtained by taking Fourier transform of (4.14):

$$v_{o1}(\theta) = \frac{4nV_{DD}}{\pi}\sin(\theta) - n^2 I_o R_{ON}\sin(\theta+\phi) \tag{4.16}$$

Since the tank L1-C1 is shorted at the fundamental frequency, we can relate fundamental component of v_o with the output current as:

$$v_{o1}(\theta) = R_L i_{out}(\theta) \tag{4.17}$$

From equations (4.13),(4.16) and (4.17), we can now determine the output current amplitude and phase:

$$\begin{aligned} I_o &= \frac{4nV_{DD}}{\pi R_L}\left(\frac{1}{1+n^2 R_{ON}/R_L}\right) \\ \phi &= 0 \end{aligned} \tag{4.18}$$

The output current, the secondary coil voltage, and the drain current and voltage of each device are plotted in Figure 4-6. As shown in the figure, output current i_{OUT}, after being reflected to the primary side of the transformer, is alternately sunk by one of the devices. The drain current waveform of each of the devices thus contains a series of half sinusoid pulses. Note that the peak value of the drain current is n times larger than the amplitude of i_{OUT} due to the transformer turn ratio. The average value of each of the drain current can be obtained by integrating the half sine wave:

$$i_{drain1,DC} = i_{drain2,DC} = \frac{nI_o}{\pi} \tag{4.19}$$

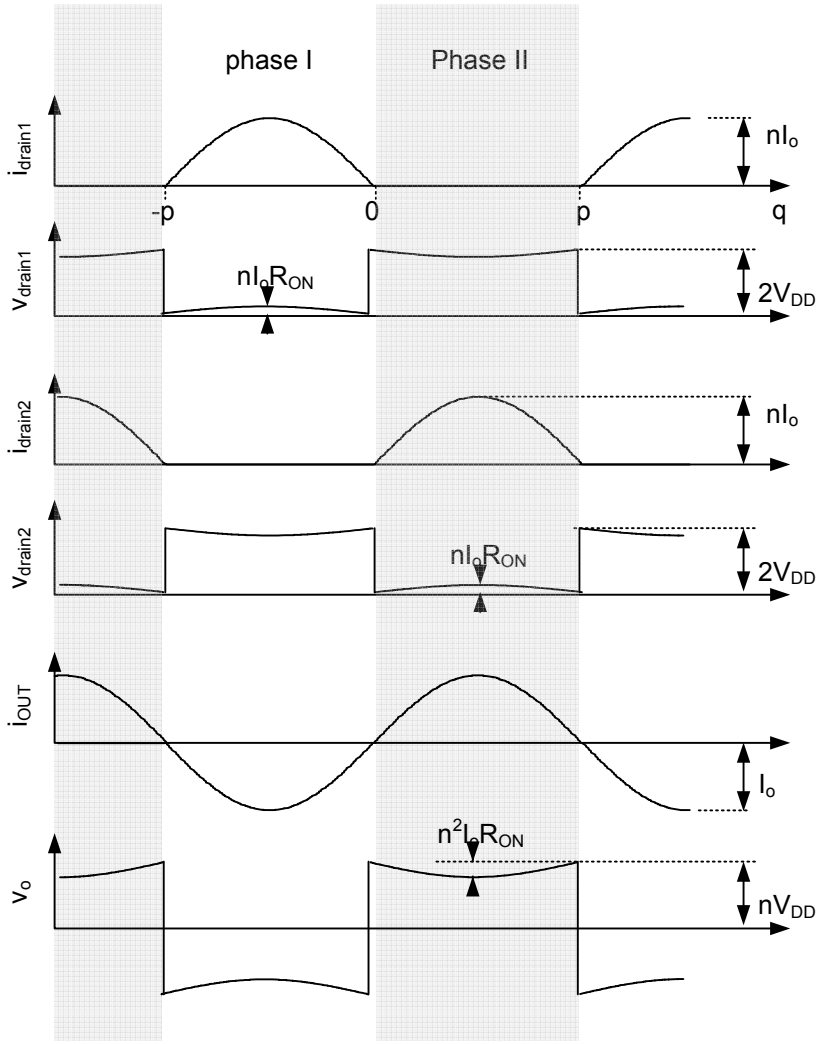


Figure 4-6 Key current and voltage waveforms for a Class-D PA

By evaluating equations (4.18) and (4.19), we can obtain the DC power consumption, the output power, and the drain efficiency of the amplifier,

$$P_{DC} = V_{DD}(i_{drain1,DC} + i_{drain1,DC}) = \frac{8n^2 {V_{DD}}^2}{\pi^2 R_L}\left(\frac{1}{1+n^2 R_{ON}/R_L}\right) \tag{4.20}$$

$$P_{out} = \frac{1}{2} {I_o}^2 R_L = \frac{8n^2 {V_{DD}}^2}{\pi^2 R_L}\left(\frac{1}{1+n^2 R_{ON}/R_L}\right) \tag{4.21}$$

$$\eta = \frac{1}{1+n^2 R_{ON}/R_L} \le 100\% \tag{4.22}$$

The upper bound in (4.22) correspond to the ideal case where $R_{ON}=0$. Note the similarity between (4.22) and the drain efficiency of Class-F2 PA in (4.10), where they both contains a term (R_L/n^2 and $Z_o^2/2R_L$ respectively) that correspond to the load impedance referred back to the drain after impedance transformation. The missing factor of two in the denominator of (4.22) is due to the differential structure.

The power capacity of the PA can be calculated as follow:

$$\begin{aligned} v_{drain,\max} &= 2V_{DD} \\ i_{drain,\max} &= nI_o \\ PC &= \frac{P_{out}}{2 \cdot v_{drain,\max} \cdot i_{drain,\max}} = \frac{\eta}{2\pi} \le 0.159 \end{aligned} \tag{4.23}$$

where the factor of two in the denominator is attributed to the number of devices used in the circuit, and the upper bound on the power capacity again corresponds to the ideal case

where $R_{ON}=0$. Comparing (4.23) with (4.11) shows that Class-D and Class-F2 PAs achieve the same power capacity.

4.4 Effects of finite load network Q on Harmonic Distortion

We will now relax the high Q assumption on the load network and examine its impact on harmonic distortion of the amplifier. First we realize that the load impedance measured from the secondary coil at the mth harmonic frequency can be expressed as

$$Z_m = \underbrace{R_L}_{R_m} + j \cdot \underbrace{R_L Q(m - \frac{1}{m})}_{X_m} \tag{4.24}$$

where the real and imaginary parts of Z_m are defined as R_m and X_m respectively. Note that in this particular case of a series resonant tank, $R_m = R_L$ for all m. The loaded Q of the load network is defined as follow, where ω_o is the operation frequency:

$$Q = \frac{\omega_o L}{R_L} = \frac{1}{\omega_o R_L C} \tag{4.25}$$

Due to the finite load network Q, there will be multiple harmonic components in the output current. However, due to the differential topology, transformer still prevents even harmonics from entering the load. The output current can thus be expressed as

$$i_{out} = \sum_{n \in 2k+1} I_m \sin(m\theta + \phi_m) \tag{4.26}$$

where k is any non-negative integer. The secondary coil voltage in equation (4.14) will be modified as

$$v_o(\theta) = n \cdot V_{DD} \cdot \Pi(\theta) - n^2 R_{ON} \sum_{m \in 2k+1} I_m \sin(\theta + \phi) \tag{4.27}$$

Taking the m[th] Fourier component of (4.27), and equating that to the product of the output current and load network impedance at the corresponding frequency, we obtain:

$$\begin{aligned} \frac{4nV_{DD}}{m\pi} \sin(m\theta) - n^2 R_{ON} I_m \sin(m\theta + \phi_m) = R_m I_m \sin(m\theta - \phi_m) \\ + X_m I_m \cos(m\theta - \phi_m) \end{aligned} \tag{4.28}$$

Since $\Pi(\theta)$ is an odd function, v_o has no even harmonic component. Also note that reactive part of the load impedance introduces a phase shift to the output current. The amplitude and phase of the output current for each even harmonic m must satisfy (4.28) which can be rewritten as

$$\begin{aligned} &\left[\frac{4nV_{DD}}{m\pi \cdot I_m} - (R_m + n^2 R_{ON}) \cos(\phi_m) + X_m \sin(\phi_m) \right] \sin(m\theta) = \\ &\left[(R_m + n^2 R_{ON}) \sin(\phi_m) + X_m \cos(\phi_m) \right] \cos(m\theta) \end{aligned} \tag{4.29}$$

In order for both sides of (4.29) to be equal at any phase angle θ, the expression within each bracket must be zero, which results in

$$\begin{cases} \phi_m = -\tan^{-1}(\eta Q(m - \frac{1}{m})) \\ I_m = \frac{4nV_{DD}\eta}{\pi R_L} \cdot \frac{\cos(\phi_m)}{m} \end{cases} \quad ; \quad m \in 2k+1 \tag{4.30}$$

where η is the nominal drain efficiency (with infinite load network Q) as expressed in (4.22). Since the power associated with the m[th] harmonic at the output is $P_{out,m} = I_m{}^2 R_m / 2$, we can calculate the m[th] output harmonic distortion from equation (4.30):

$$HD_m = \frac{1}{m^2\left[1+\left[\eta Q\left(m-\frac{1}{m}\right)\right]^2\right]} \tag{4.31}$$

The lower order harmonic distortion is plotted in Figure 4-7. We note that the distortion roll off fairly slowly, and Q>5 is needed to push the 3[rd] order harmonic to beyond -30dBc (dB below carrier). Many wireless applications require even lower harmonic distortion, so additional RF filtering is often required.

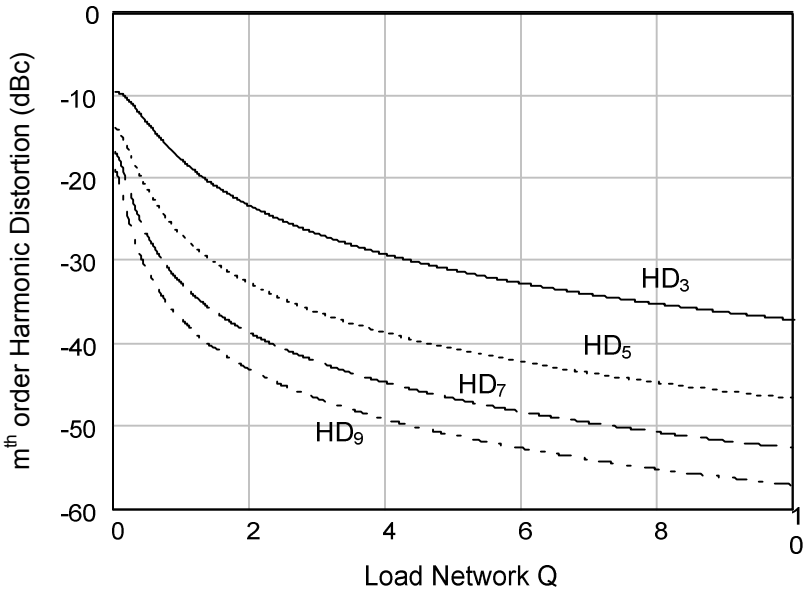


Figure 4-7 Harmonic distortion vs. load network Q

4.5 Effects of Finite Load Network Q on Efficiency

At low load network Q, more energy from the supply is used to generate the harmonic components of the output signal. Since we do not consider the non-fundamental components as wanted signal, the overall efficiency will degrade with the reduction in load network Q. The efficiency degradation factor can be defined at the ratio between the fundamental component ot the output power to the total output power. From (4.31), the degradation factor can be expressed as

$$\frac{\eta_{finiteQ}}{\eta_{Q=\infty}} = \left[\sum_{m \in 2k+1} \frac{1}{m^2 \left[1 + (\eta Q (m - \frac{1}{m}))^2 \right]} \right]^{-1} \tag{4.32}$$

The degradation factor is plotted in Figure 4-8, which shows that the efficiency degradation is fairly mild as long as Q is reasonably high (>2 or so). Besides, it is also not a strong function of the nominal drain efficiency (corresponding to infinite Q.)

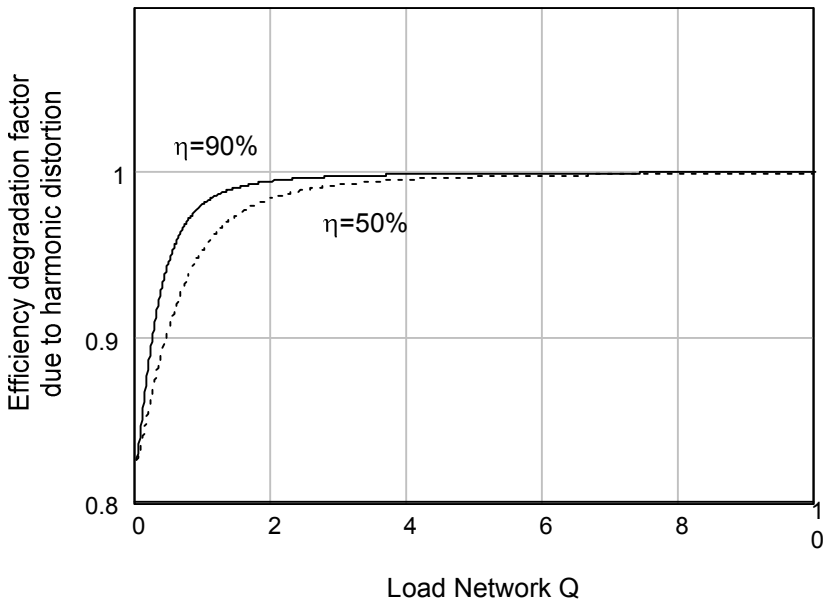


Figure 4-8 Efficiency degradation due to finite Q

Upper Limits of Realizable Load Network Quality Factor

Figure 4-8 seems to suggest that the load Network Q should always be maximize for optimized efficiency. However, the realizable quality factor in practice is limited by several factors. These factors include the flatness of the amplifier frequency response over the bandwidth of interest; as well as the compatibility of the required inductor and capacitor values with the realizing technology. Yet another factor is the intrinsic quality factors of the inductor and capacitor within the load network. Their effects can be modeled by the parasitic resistors r_L and r_C shown in Figure 4-9.

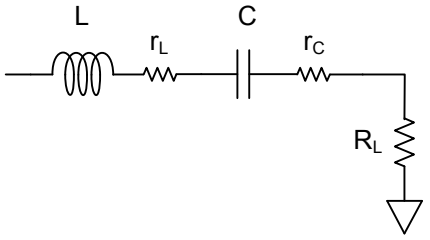


Figure 4-9 Model of parasitic loss in the passive components

In general r_L and r_C are frequency dependent because the underlying loss mechanism involves frequency dependent skin effect and substrate coupling. For the loss analysis, we concern mainly their values at the fundamental frequency. The parasitic resistance can be expressed as

$$r_L = \frac{\omega_o L}{Q_L} = R_L \frac{Q}{Q_L}$$
$$r_C = \frac{1}{\omega_o C Q_C} = R_L \frac{Q}{Q_C} \tag{4.33}$$

Q_L and Q_C are the intrinsic quality factors of the inductor and capacitor respectively. They depend only on the intrinsic resistance and reactance of the corresponding components. On the other hand, the load network quality factor Q in (4.33) includes the total loss (i.e. intrinsic loss of L and C, as well as the output load R_L). Since the parasitic resistors are in series with the load resistor, part of the output power is dissipated in the parasitics and this degrades the overall drain efficiency. The degradation fact can be expressed as:

$$\frac{\eta_{lossy\ passives}}{\eta_{lossless\ passives}} = \frac{R_L}{R_L + r_L + r_C} = \frac{1}{1 + Q(\frac{1}{Q_L} + \frac{1}{Q_C})} \tag{4.34}$$

Equation (4.34) shows that the efficiency degradation can be significant if Q approaches the intrinsic quality factors of the passive components. With on-chip components and GHz operations, Q_C can be in the hundreds, but the Q_L is usually much lower – often in the range of 10-20 depending on frequency, metal thickness, dielectric stack-up, substrate material etc. Off-chip inductors have lower loss, but still their quality factor rarely exceeds 50 in practice. Figure 4-10 shows the efficiency degradation at various combinations of Q_C and Q_L. Because of the strong influence of passive component loss

on the overall efficiency, the realizable load network Q of a typical integrated PA is often limited to low single digit values.

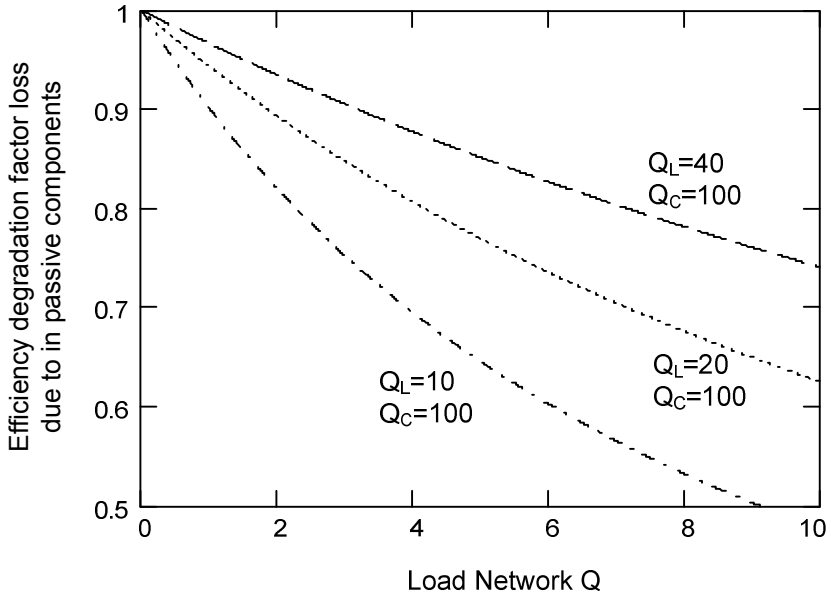


Figure 4-10 Efficiency degradation due to passive component loss

While the discussion of the effects of load network Q on the output spectral purity and efficiency is based on Class-D PA, the conclusion is generally applicable to other classes of power amplifiers.

4.6 Summary

In Chapter 3 and Chapter 4 , we discussed several key representatives of transconductor-based and switch-mode power amplifiers. Their key characteristics are summarized in Table 4-1.

Year	Type	Max Eff.	Power Capacity	Remarks
Class-A	Gm-based	50%	0.125	Most Linear, most inefficiency
Reduced Conduction angle PAs	Gm-based	50%-100%	0.134 - 0.0	Linear, but higher efficiency translates to lower power capacity
Class-F1	Gm-based	88.3%	0.141	Linear, but difficult to control drain voltage waveform
Class-F2	Switch-mode	100%	0.159	Nonlinear, good efficiency and power capacity
Class-D	Switch-mode	100%	0.159	Nonlinear, good efficiency and power capacity

Table 4-1 Comparison among main forms of gm-based and switch-mode PAs

Transconductor-based power-amplifiers are relatively linear but they suffer from poor efficiency and low power capacity. If constant envelope modulation is used in the system, then switch-mode PAs generally results in better power efficiency and power capacity.

Although switch-mode PAs can theoretically achieve 100% efficiency, the actual achievable values are limited by capacitor charging/discharging loss and circuit

parasitics. In Chapter 6 , we will revisit issues in the discussion of the switch-mode Class E power amplifier and examine how these loss-factors can be further alleviated.

Chapter 5 - Power Amplifier Implementation Technologies

5.1 Introduction

One basic goal in PA design is to fully utilize the capability of the underlying semiconductor technology [Berg99]. Currently mainstream power amplifiers are predominantly manufactured in three prevailing technologies, namely GaAs MESFET, GaAs HBT and SiGe HBT [Jos01] [Chan02]. In this chapter, we will study the basic device characteristics of these technologies and compare them with standard CMOS in the context of PA realization. Furthermore, we will also discuss the key challenges involve in CMOS PA realization and in particular on the mechanism of oxide breakdown.

5.2 GaAs MESFET

Metal Semiconductor Field Effect Transistor (MESFET) is one f the predominant choices of GaAs technology [Goli03]. It accounts for a good portion of the RF power amplifiers used in wireless applications, especially in cellular and satellite systems. MESFET was first fabricated by Mead in 1966 and its capability in microwave frequency was quickly demonstrated by Hooper and Lehrer in 1967.

Compared with CMOS, GaAs MESFET holds a significant speed advantage. Electron mobility in GaAs (8500 cm^2/V sec) is about 6X higher than in silicon (1450 cm^2/V sec); and saturation velocity of electron in GaAs (0.13um/psec) is about 40% higher than in silicon (0.09um/psec). In addition, the semi-insulating GaAs substrate greatly reduces eddy current and capacitive substrate loss commonly found in CMOS substrate. As a result, very high quality inductors and capacitor are routinely available in GaAs process, which significantly enhance the overall RF performance. From PA perspective, GaAs MESFET exhibits a high breakdown voltage (up to 20V) and the availability of low loss gold interconnects. MESFET PAs operating up to 40GHz with decent power efficiency are commercially available.

The major drawbacks of GaAs are manufacturing cost and incompatibility with VLSI technology. The fact that most GaAs ICs are still fabricated on 6 inch wafers (compared with 12 inch for CMOS) has been a bottleneck for aggressive cost reduction.

GaAs MESFET Device Structure

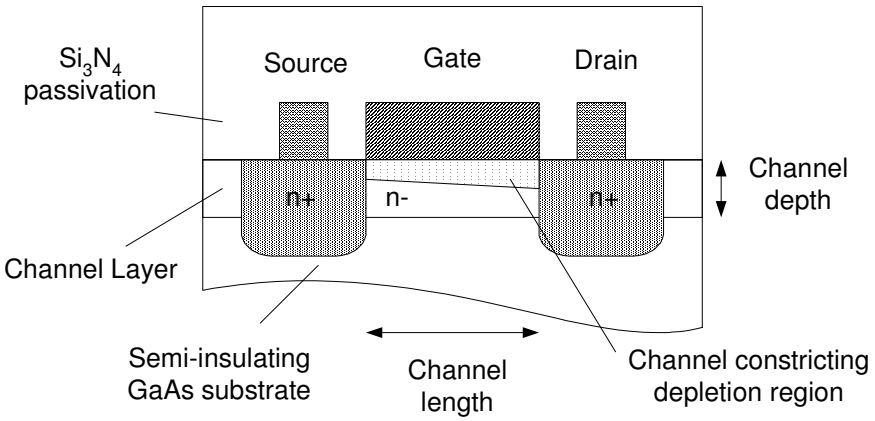


Figure 5-1 GaAs MESFET device structure

Figure 5-1 shows the cross section of a typical GaAs MESFET. The n-GaAs channel layer is formed either by epitaxial growth or by directly implanting an n-type dopant (typically silicon) into the substrate. Direct ion implantation is of lower quality but it is also lower cost due to the elimination of the expensive MBE/MOCVD process step. The channel thickness is usually below 0.3 um and it is a function of the required power handling capability – power devices normally prefer deeper channel for high current conduction. The ratio between the change length and thickness over generations is kept fairly constant around 4, resulting from the trade-off among speed, short-channel effect and parasitic. The gate material for an n-channel MESFET is usually aluminum or

tungsten silicide. The source and drain are formed by ion implantation with concentration around 10^{18} cm^{-3}. The source/drain junction capacitance is typically kept low by the semi-insulating substrate. Source/drain implantation can be self-aligned with the gate (Figure 5-1), resulting in a low source and drain resistance. The self-alignment however has a tendency to lower the drain-gate break-down voltage which ultimately affects the reliability of the PA. One way to mitigate this effect is to extend the drain away from the gate as shown in Figure 5-2.

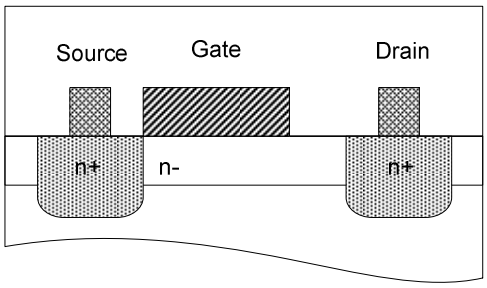


Figure 5-2 GaAs MESFET with extended drain to increase breakdown voltage

GaAs MESFET Device Characteristics

Unlike CMOS, MESFET is a majority carrier device where current conduction is formed by majority carriers within channel. In an n-channel device, the source, drain and channel are all doped n-type; the channel is thus conductive in its native state and channel current

will flow as soon as a voltage is applied across the drain and the source. To control the channel current, the channel cross section is restricted by a depletion layer beneath the gate, whose depth is control by the gate–source voltage. This depletion layer is formed by the Schottky barrier between the gate and the channel. This barrier electrically isolates the gate from the channel, resulting in desirable high gate impedance. When a voltage is applied to the gate, the depletion layer depth and the channel cross section are modified to control the current flow, and thus resulting in the basic characteristic of the transistor.

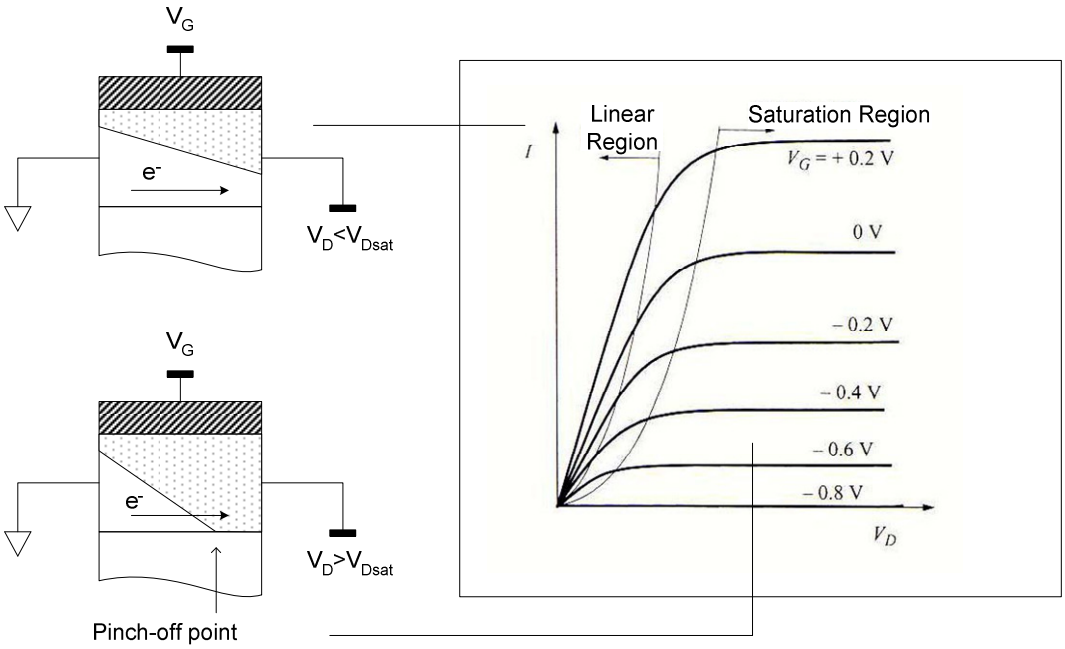


Figure 5-3 I-V Characteristic of GaAs MESFET

The IV characteristic of GaAs MESFET is shown in Figure 5-3 where we assume the source is at ground potential. If the gate voltage (V_G) is below the threshold voltage (V_T), the device is cut-off because the depletion layer reaches the entire depth of the channel and no current would flow regardless of the drain voltage. When V_G rises above V_T, the channel becomes conductive. If the drain voltage is low, the depletion depth is relatively uniform along the channel length, and the channel behaves much like a resistor where the channel current rises proportionally with V_D. In this case the device is in the linear region. As V_D continue to increase, the build-up in the drain-channel potential will increase the depletion depth around the drain. This reduces the local channel cross section area and raises the incremental channel resistance. As a result the channel will continue to increase with V_D but at a progressively lower rate.

Further increasing V_D will eventually drive the MESFET into the saturation region in which the current becomes independent of V_D. For long channel devices (approx. >1um), when V_D approaches V_G-V_T, the depletion depth around the drain reaches the full channel depth, i.e. the channel is pinched off. Beyond this point further increase in V_D will only push the pinch-of point towards the source, but the channel voltage at the pinch-off point remains unchanged. Since the electric field within the drift region remains fairly constant, change in V_D does not affect the channel current significantly except to a small degree due to channel length modulation. The only way to effectively modulate the channel current now is by varying V_G. The saturation region is where most amplifiers are biased

at, since the device behaves much like an ideal voltage controlled current source. The IV characteristic of the MESFET in the saturation region is similar to the square law model of a MOS transistor. It can be approximated by:

$$Isat = \frac{\mu_n}{2} \cdot \left(\frac{\varepsilon}{a}\right) \cdot \left(\frac{W}{L}\right) \cdot \left(V_{GS} - V_T\right)^2 \tag{5.1}$$

where μ_n is the electron mobility, ε is the permittivity of GaAs, α is the channel depth and W and L are the channel width and length respectively. Similar to MOS, MESFET is also subject to short channel effects. In a short channel device, the electric field build-up along the channel may reach the critical field before pinch-off occurs. In this case, velocity saturation occurs and channel current will cease to increase with V_D before V_D reaches V_G-V_T. As a result, both the saturation current and the gm would be degraded.

The GaAs MESFET shown in Figure 5-3 is a depletion-mode device (i.e. $V_T < 0$). Enhancement-mode MESFET can also be built by adjusting the channel doping and reducing the channel layer thickness. However, they suffer from limited gate voltage swing (to avoid forward biasing the Schottky diode) and limited current driving capability. As a result, depletion-mode MESFET is usually used for PA. However, this implies the need for either dual polarity supply, or negative voltage generation, which gives GaAs MESFET a slight disadvantage when compared with GaAs HBT and SiGe devices.

5.3 GaAs HBT

The structure of a GaAs Heterojunction Bipolar Transistor (HBT) is similar to Silicon bipolar, except a wide band-gap material is used in the emitter. This allows a reversal in the base/emitter doping concentration and results in significant enhancement in high frequency performance. Similar to GaAs MESFET, GaAs HBT also benefit from the semi-insulting substrate, low parasitic and high-Q passive components.

One advantage HBT has over MESFET and MOSFET is that it conducts current vertically through a relatively large emitter cross section area. The current carrying capacity of HBT is therefore substantially higher than the FETs where the current capacity is limited by the channel thickness. In addition, HBT devices also feature higher gain due to its exponential IV characteristics. These characteristics have made GaAs HBT quite popular in power amplifier realization.

GaAs HBT Device Structure

Figure 5-4 depicts the structure of a GaAs HBT based on a graded AlGaAs/GaAs emitter-base junction. The starting material is semi-insulating GaAs substrate; on top of which a number of layers are epitaxially grown. These layers include the n+ sub-collector layer (to reduce collector resistance), the n-collector layer, the p-base, and the n-emitter. All of these layers are built with GaAs except for the emitter where a graded deposition of $Al_xGa_{1-x}As$ is used. The fraction x increases from 0 at the emitter-base interface to a

maximum of about 0.3 towards the emitter. Since the lattice constant of AlAs is close to that of GaAs, it is possible to grow $Al_xGa_{1-x}As$ on GaAs with any fraction x without lattice mismatch which can cause dislocations and device failure. Ti/Al is typically used for the metal contact.

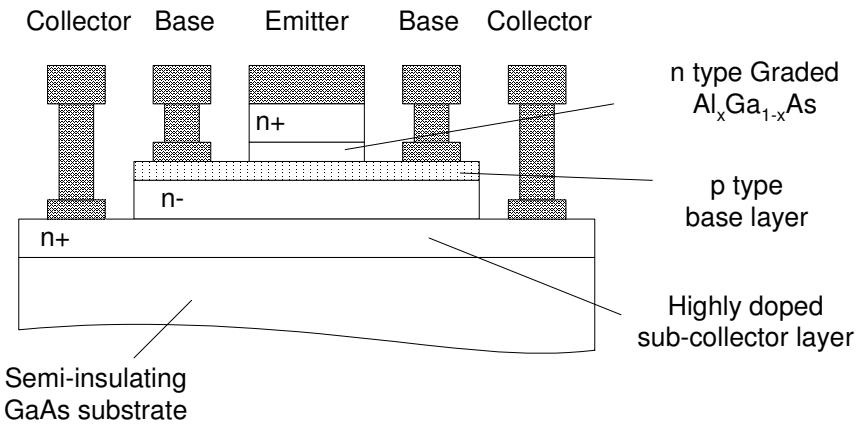


Figure 5-4 AlGaAs/GaAs HBT device structure

GaAs HBT Device Characteristic

As shown in Figure 5-5 (a), the current conduction of a typical npn BJT has two main components: The first component is the current conducted by the electrons in the emitter crossing the energy barrier formed by the forward biased B-E junction. These electrons diffuse across the thin base, then they are swept by the electric field across the reverse biased B-C junction and finally emerge from collector as the collector current I_C. While

the electrons are in the base, a small portion of them are lost to the recombination with holes; and this loss accounts for the finite emitter efficiency of the BJT. Increasing V_{BE} lowers the energy barrier at the B-E junction and exponentially increases the electron emission from the emitter, hence results in the high transconductance of BJT.

The second component is the bases current conducted by the holes in the base crossing the energy barrier across the forward bias B-E junction and diffuse into the emitter. This current component is known as the base back-injection current $I_{B,back\text{-}inject}$. Although other components such as the recombination current also contribute, $I_{B,back\text{-}inject}$ is typically accounts for most of the total base current. It therefore is a dominant factor for the transistor current gain β. Since increasing V_{BE} lowers the energy barrier for both the emitter current and $I_{B,back\text{-}inject}$, β is to the first order independent of V_{BE}.

If we ignore the small base recombination current, and assume that the base-emitter is a homojunction (i.e. the base and emitter are made of material with the same bandgap voltage, such as in Silicon BJT), an expression can be derived for the ratio between the collector current and $I_{B,back\text{-}inject}$:

$$\frac{I_C}{I_{B,back-inject}} = \frac{X_E}{X_B} \cdot \frac{D_{n,B}}{D_{p,E}} \cdot \frac{N_E}{N_B} \tag{5.2}$$

where X_E and X_B are the emitter and base thickness respectively; $D_{n,B}$ and $D_{p,E}$ are the diffusion constant of electrons in the base and holes in the emitter respectively; and N_E

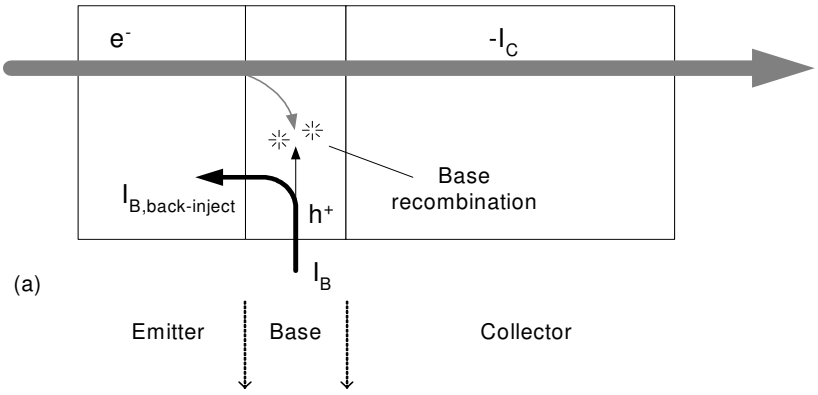
and N_B are the doping concentration in the emitter and the base respectively. Given that the thickness and diffusion constants are in the same order, it is clear from (5.2) that in homojunction BJT, the doping level in the emitter has to be much higher than that in the base in order to achieve a reasonable β. However, both low base doping and high emitter doping degrade the high frequency performance of the transistor. Low base doping increases base resistance; and high emitter doping increases the base-emitter junction capacitance. This therefore leads to a fundamental trade-off between current gain and high frequency performance in homojunction BJT.

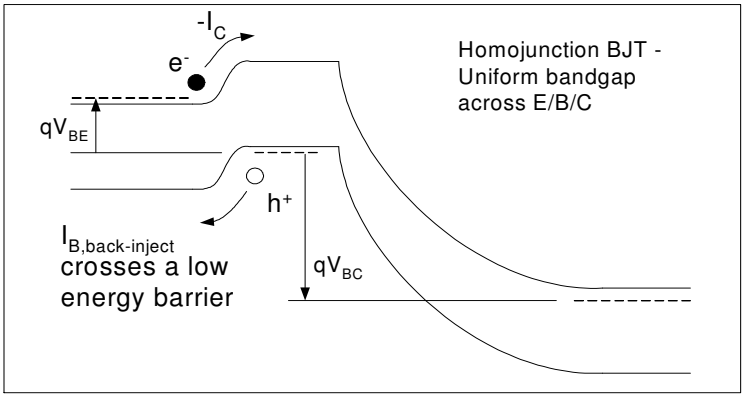
GaAs HBT, being a heterojunction bipolar device, introduces another degree of freedom in the above mentioned trade-off, and results in high β without scarifying high frequency performance. As illustrated Figure 5-3 (b) (c), by introducing a wide bandgap material in the emitter, the energy barrier for back-injection base current is increased without affecting the collector current. In a GaAs HBT, the ratio between I_C and $I_{B,back\text{-}inject}$ becomes [Goli03]

$$\frac{I_C}{I_{B,back-inject}} = \frac{X_E}{X_B} \cdot \frac{D_{n,B}}{D_{p,E}} \cdot \frac{N_E}{N_B} \cdot \exp(\frac{\Delta E_g}{kT}) \tag{5.3}$$

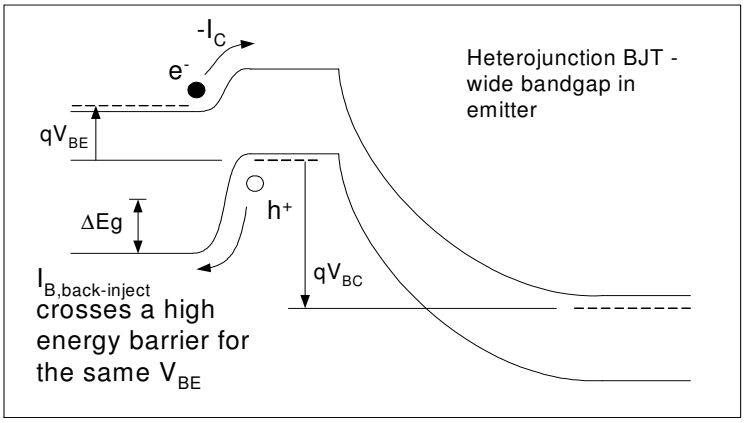
where ΔE_g is the difference between the bandgap of the emitter and the collector. Because of the exponential dependence on ΔE_g, fairly high β can be achieved even with low emitter and high base doping concentrations. Compared with homojunction BJT, the

GaAs HBT thus exhibits low R_B and C_{je}, and marked improvement in high frequency performance.





(b)



(c)

Figure 5-5 BJT current components and energy bandgap

Breakdown Mechanism in Bipolar Devices

The breakdown mechanism is a major consideration for PA design. In bipolar transistor, breakdown is typically induced by avalanche breakdown at the reversed biased B-C junction. The collector-emitter breakdown voltage in bipolar is characterized by two parameters, BV_{VBO} and BV_{CEO}, and distinction lies in the base termination.

In the common base configuration (Figure 5-6a), the base is grounded and a controlled current I_E is sunk from the emitter. Regardless of the value of I_E, when V_{CE} ramps up to a certain voltage (BV_{CBO}), the B-C junction will start to breakdown and cause sudden increase in collector-base current. Breakdown under this configuration is induced by the following mechanism: Electrons emitted from the emitter diffuse across the thin base and eventually enter the depletion region at the B-C junction. Within the depletion region, the electrons are accelerated by the electric field due to the reverse bias V_{CB} voltage. At high $V_{CB,}$ bias, some electrons gather enough energy so that when they collide with the lattice they induce impact ionization and generate new electron-hole pairs. The free electrons generated by impact ionization are accelerated by the same strong field, and may cause further impact ionization. There is therefore a multiplying effect where a great number of electrons and holes are generated in the C-B space charge region. The electrons and holes accelerate towards the collector and the base respectively, causing a large breakdown current from the collector to the base.

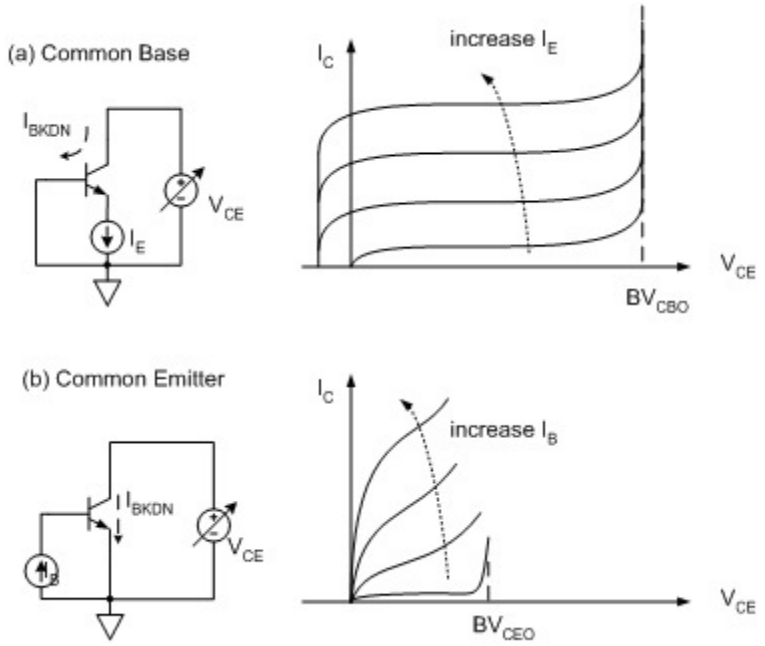


Figure 5-6 BJT breakdown characterization (a) common base, (b) common emitter

The breakdown voltage BV_{CBO} can be expressed as

$$BV_{CBO} = \frac{\varepsilon \cdot E_{br}^{\ 2}}{2q \cdot N_c} \tag{5.4}$$

where ε is the permitivity of the semiconductor, E_{br} is the breakdown field, q is the electron charge and N_C is the collector doping concentration. The values of ε and E_{br} in GaAs are higher than that of Si ($\varepsilon_{GaAs} = 12.9 \cdot \varepsilon_o$, $\varepsilon_{Si} = 11.7 \cdot \varepsilon_o$, $E_{br,GaAs} = 40$V/um and $E_{br,Si}$ ~ 30V/μm), resulting in a somewhat higher breakdown limit in GaAs devices. BV_{CBO} is a

strong function of the collector doping, but relatively insensitive to base doping. Doping concentration is usually much higher in the base than in the collector, so that the depletion region exists mainly on the collector side. In reality, breakdown voltage is one of the main factors in determining the collector doping level.

In the common emitter configuration (Figure 5-6b), a controlled current I_B is sourced into the base terminal as the voltage across collector and emitter is increased. In this case, the onset of the breakdown is also caused by avalanche breakdown in the B-C junction. The key distinction is that, even at low level of impact ionization, the holes generated are eventually trapped in base because of the high impedance base termination. The accumulation of these excess holes eventually raises the base potential and causes a rapid rise in the emitter current. The onset of the common emitter breakdown thus happens at a much lower voltage than in the common base configuration ($BV_{CBO}>>BV_{CEO}$). Besides, the breakdown current flows from the collector to the emitter instead of the base. The two breakdown voltages are related by

$$BV_{CEO} = \frac{BV_{CBO}}{\beta^{1/n}} \tag{5.5}$$

where β is the bipolar current gain and n is a material constant (approximately equals to 3 in silicon).

BV_{CEO} should be treated as a worse case limit. In typical RF circuits it is very rare to have truly high base impedance termination. In practice, even moderate base impedance in the order of a few hundred Ohms can substantially increase the breakdown limit above BV_{CEO}.

5.4 SiGe HBT

The first successful implementation of SiGe HBT was reported by IBM [Patt90] in 1987. Early effort in building SiGe HBT was hindered by the lattice mismatch between Si and SiGe. Unlike in the AlGaAs/GaAs HBT where the two constituting materials have very similar lattice constant, the lattice constant of SiGe is about 4% larger than Si. This mismatch results in a compression strain on the SiGe layer that is deposited onto the Si substrate; and imposes a stringent limit on the thickness of SiGe that can be grown without causing dislocation in the SiGe alloy and device failure [Meye96]. It was after the investigation by Matthews and Blakeslee [Matt74] on the issue of critical thickness did reliable fabrication of SiGe become possible.

SiGe HBT Device Structure

Similar to GaAs HBT, SiGe HBT uses heterojunction structures to optimize high-frequency characteristic without scarifying the current gain. The detail construction of the two devices however differs.

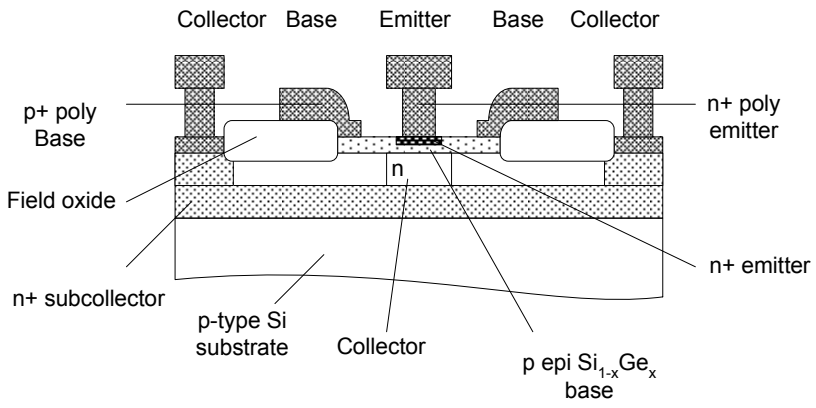


Figure 5-7 SiGe HBT device structure

Instead of using a wide bandgap emitter material, SiGe HBT uses narrow bandgap base material (i.e. $Si_{1-x}Ge_x$ alloy). As shown in Figure 5-7, this results in a double heterojunction structure, one at the B-E junction, the other at the B-C junction. In addition, the germanium concentration in the base is typically graded to create a build-in electric field in the base. The performance of SiGe HBT thus benefits from the two aspects:

Firstly, delta bandgap energy (ΔE_g) at the B-E junction enhances the current gain by a factor $exp(\Delta E_g/kT)$, as shown in (5.3). This thus allows base doping to be increased and the emitter doping to be reduced without scarifying the current gain β. Changing the base and emitter doping in these directions reduces the base charging time and B-E junction capacitance and thus helps to enhance the high frequency characteristic.

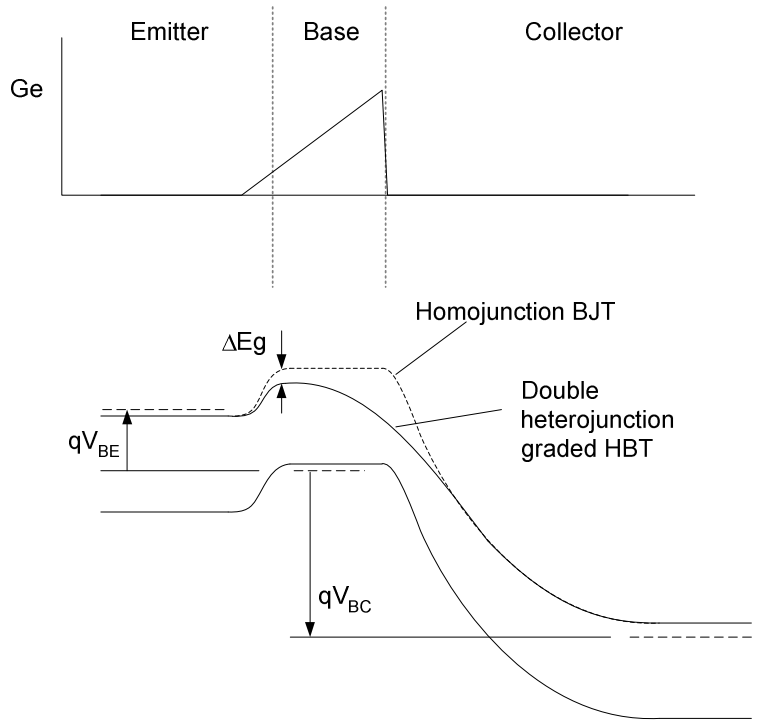


Figure 5-8 Band diagram of a double heterojunction Si HBT

Secondly, since the bandgap voltage of the SiGe alloy is a function of its germanium concentration, the graded germanium concentration in the base helps to create a built-in electric field within the base (Figure 5-8). Normally, once electrons cross the B-E junction barrier, they will diffuse across the base until eventually swept by the filed in the reverse bias B-C junction. The base diffusion is typically a speed-limiting factor. With the built-in field, electron transport in the base is assisted by both diffusion and drift. This

leads to shorter base transit time and enhanced high frequency characteristics of the device [Goli03].

SiGe HBT Device Characteristics

The principle advantage of SiGe HBT is its outstanding high frequency performance which is a few times better than conventional Si bipolar. Si HBT device with f_T = 130GHz and f_{max} = 160GHz has been reported [Schu95] [Oda97]. Most importantly, Si HBT fabrication is compatible with standard silicon process and the cost of SiGe BiCMOS process is similar to standard Si BiCMOS. The process described in [King00], for example, can realize SiGe HBT devices with four additional masks to a baseline 0.25um CMOS flow.

SiGe HBT is attractive to PA designers because of its high gain and high current carrying capacity resulting from a large cross-section, vertical current conduction BJT structure. The high thermal conductivity of silicon (approximately 3 times higher than GaAs) also facilitates heat dissipation from the PA devices. This is a significant advantage because much of the high current limits of HBT, such as current gain dropping, stems from thermal effect.

Compared with GaAs devices, one major drawback of SiGe HBT is the lack of a semi-insulating substrate. Even with high-resistivity Si substrate, the electrical conductivity is

a few orders of magnitude higher than intrinsic GaAs. The conductive Si substrate results in lossy passive components.

Another limitation of Si HBT is its moderate breakdown voltage. Foundries typically offer options to trade off device speed with high voltage operation, mainly by engineering the collector doping level with additional mask steps. Typical BV_{CEO} for a 'high speed' HBT device is in the range of 2-3V, whereas a 'high voltage' variant may have BV_{CEO} increased up to 6-7V at the expense of degraded f_T. Breakdown limit of SiGe HBT is therefore somewhat better than CMOS, but much lower when compared with GaAs MESFET.

5.5 CMOS

Standard silicon CMOS is by far the most cost effective semiconductor process and performance of the deep submicron CMOS has reached or even surpassed GaAs in many aspects. However, CMOS still presents a number of important technological challenges in term of power amplifier realization.

The lack of high quality factor passive components, the lossy substrates, and the package parasitic typically found in standard CMOS processes do hinder the realization of efficient power amplifiers. Yet, the most prominent challenge is perhaps the low breakdown voltage of CMOS device [Wolt85]. This limits the on-chip voltage swing and results in high-current circuits that are much more susceptible to parasitic losses; and the

combination of a high current level and low gain of CMOS transistors also results in large devices whose gate and drain capacitors because very difficult to drive at high frequency. In this section we will focus on the breakdown mechanism in silicon CMOS technology.

CMOS Oxide Breakdown

When MOS transistor is used to implement power amplifiers, the device is subject to two possible breakdown mechanisms – junction breakdown across the drain-substrate reverse biased junction, and oxide breakdown across the oxide in the drain-gate overlap area. In modern thin-gate sub-micron process, oxide breakdown is typically the limiting factor.

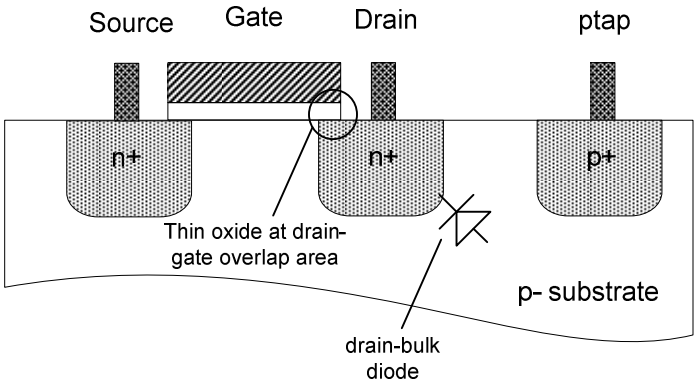


Figure 5-9 potential breakdown locations in a CMOS device

It is generally agreed that the breakdown process occurs in two phases – a gradual, field-induced defect accumulation, followed by thermal runaway and permanent damage of the oxide. In the reciprocal electric-field model [Lee88], under sufficiently high field, electron at the cathode can be injected into the oxide through the process of Fowler-Nordheim tunneling, some of the electron collide with the oxide lattice and cause impact ionization in which multiple electron-hole pairs are generated. Some of the slow moving holes are eventually trapped within the oxide and near the oxide-cathode interface.

Since the holes are electrically charged, the newly generated traps can disturb the local electric field in their vicinity. In the regions where the local field is intensified, the electron tunneling process is enhanced which encourage the built up of more traps and defects. Over time, the defect density reaches a critical level where sufficiently high local field and current results in Joule heating which in turn make the dielectric matrix more conductive and pass more current. Once this thermal runaway process is started, it quickly results in a short across the oxide and permanently destroys the associated device. Figure 5-10 shows the damage resulted from oxide after breakdown [Ohri98].

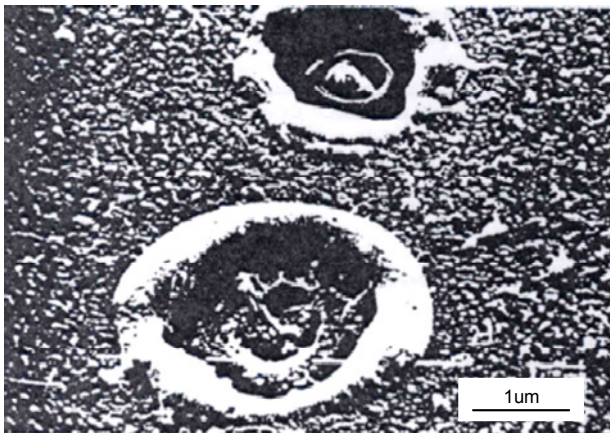


Figure 5-10 Damage caused by oxide breakdown, tox-40nm [Ohri98]

Oxide breakdown can be characterized by the Mean Time to Failure (*MTTF*) parameter which is the anticipated time for the defect density to reach the critical level under the applied field. *MTTF* can be expressed as

$$MTTF = t_{OR} \cdot \exp(\frac{G_R \cdot t_{ox}}{V_{ox}}) \tag{5.6}$$

where V_{ox} and t_{ox} are the voltage across the oxide and the oxide thickness, t_{OR} and G_R are temperature dependent constants:

$$t_{OR} = 5.4 \times 10^{-7} \cdot \exp(\frac{-0.28eV}{kT}) \cdot \sec \tag{5.7}$$

$$G_R = \left(12 + \frac{0.58}{kT}\right) \cdot V / nm \tag{5.8}$$

(5.6) shows that *MTTF* depends exponentially on the electric field (V_{ox}/t_{ox}). When the applied field within the oxide reaches a sufficiently high value, oxide breakdown will happen almost instantaneously. The intrinsic breakdown field is the maximum field that is sustainable by a defect-free dielectric. The intrinsic breakdown field for SiO_2 is about 1.1V/nm. In practice, however, the rated maximum gate voltage for a typical CMOS process is always lower than that corresponds to the intrinsic breakdown field.

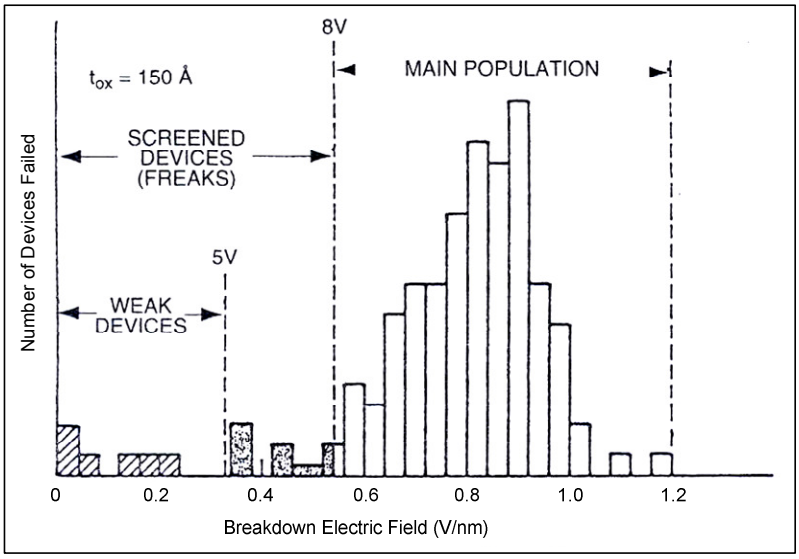


Figure 5-11 Experimental Oxide Breakdown Data [Sabr90]

Figure 5-11 shows the breakdown voltage measurement on a large number of MOS capacitor structures with t_{ox}=150nm [Sabr90]. From these data, it can be seen that the main population of the samples starts to break down at as low as 0.55V/nm.

One reason for device to breakdown before reaching the critical field is the existence of built-in defects in the oxide. These defects are in the form of impurities, broken or dangling bonds, and stacking faults. Although defect density in CMOS has been significantly reduced over generation, it is still impossible to guarantee zero-defect during the fabrication process. In addition, the presence of defects, charge bodies and surface roughness at the Si/SiO_2 interface also reduce the local oxide thickness and increase the effective field for a given gate voltage. A common industrial guideline for maximum gate voltage in a typical CMOS process is 70% above the rated V_{DD}. For example, for a 0.35μm process, the maximum gate voltage should be kept below 5.6V, which corresponds to an oxide field of about 0.6V/nm.

AC Breakdown Limit

The industry standard for oxide breakdown characterization is to test with static voltage under elevated temperature. However, for power amplifiers, the active device experiences large voltage across their gate oxide only for a portion of the operation cycle. It is thus important to know if the device will exhibit a different breakdown limit under high frequency stressing.

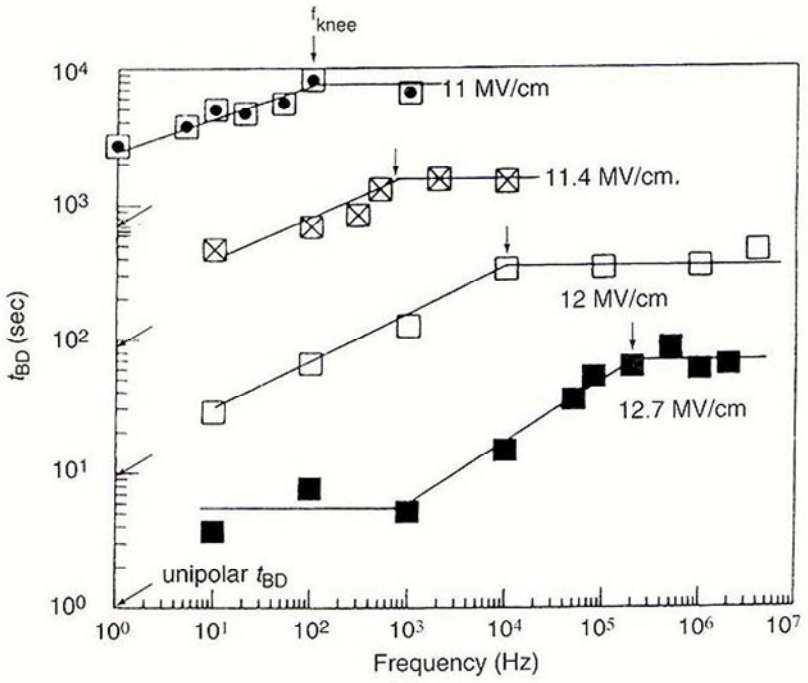


Figure 5-12 Time to breakdown vs. frequency [Hu98]

Indeed, it was reported that about ten times increase in gate oxide lifetime was possible if the device is subjected to AC instead of DC stressing (Figure 5-12) [Hu98]. The improvement in breakdown limit at high frequency is attributed to the fact that, when the oxide field is alternating rapidly, it becomes more difficult for the slow moving holes to drift deep inside the oxide bulk and form harmful electron traps that would eventually lead to breakdown. Instead, these holes tend to aggregate close to the interface to form the more benign interface traps.

Unfortunately AC oxide breakdown limit currently is not part of the standard characterization in production CMOS. As a result, PA designers often fall back to the more conservative DC limit. However, the difference between AC and DC stressing does provide insight on the relative reliability of PAs with different voltage waveforms. For example, between two PAs with the same peak instantaneous voltage across the gate, we would expect the one with a lower equivalent duty cycle to be more reliable.

Chapter 6 - Analysis of Class-E PA

6.1 Introduction

Our research in CMOS PA realization has led to the development a 1-W 1.9GHz power amplifier using a standard 0.35μm CMOS process. The design exploits the switching class-E operation to achieve high efficiency and is targeting at wireless systems using constant envelope modulation schemes. The PA operates with a low supply voltage to avoid breakdown, and is compatible with modern submicron, low-voltage CMOS technologies. A fully differential topology is used to extend the on-chip voltage swing and to minimize effect of substrate coupling. In addition, the demanding input driving

requirement of the large transistors is significantly reduced by using the injection-locking technique. In addition, we also propose a compact balun to interface the differential PA output with single ended off-chip components.

Design approach of the Class-E CMOS PA will be discussed in Chapters Chapter 6 Chapter 7 and Chapter 8 We begin with this chapter on the analytical background of Class-E operation and exploit its characteristics for PA realization. We will then continue in Chapter 7 to present the design of the proposed CMOS PA, and in particular to discuss the considerations involved in the differential topology and the injection locking technique. The theory behind the compact microstrip balun will then be examined in Chapter 8 .

6.2 Hard-switching vs. Soft-switching

The discussion in Chapter 4 on Classe-F2 and Class-D PAs highlights the principle advantages of switch-mode power amplifiers which include high efficiency and high power capacity. However, both of these switch-mode PAs suffer from the same drawbacks that they are sensitive to the presence of parasitic drain capacitance (C_{drain}). In fact introduction of C_{drain} can cause substantial power loss due to charging and discharging of the capacitor. In the case of Class-F2 PA, it also disrupts the harmonic termination of the load network.

The capacitor charging and discharging loss in Classe-F2 and Class-D PAs can be attributed to their hard-switching nature. By this we refer to the fact that in these PAs, the switch is turned on when the drain voltage is still high. For example, in the Class-F2 PA of Figure 4-2, the device is turned on when the drain voltage is at $2V_{DD}$. Once the switch is closed, C_{drain} is discharged through the device's on-resistance and the energy initially stored on the drain capacitor ($C_{drain}V_{DD}^2$) is dissipated at the device as heat.

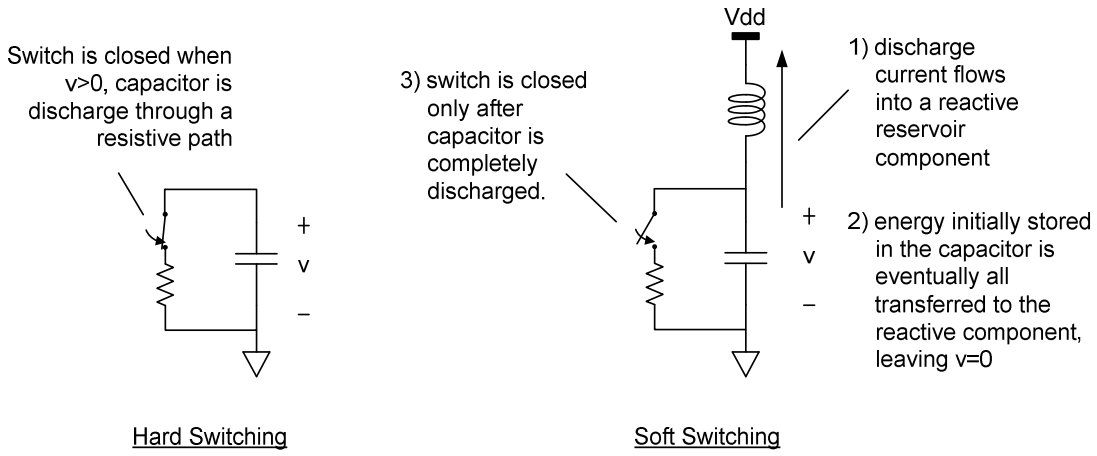


Figure 6-1 Illustration of the difference between hard- and soft-switching

The charging and discharging loss would be eliminated if C_{drain} is pre-discharged losslessly before the switch is turned on. One way to achieve this is to design the load network in a way that it transfers the energy stored in C_{drain} to a 'reservoir' reactive

component (e.g. an inductor) immediately before the switch is closed. Such configuration is known as zero-voltage, or soft-switching. The difference between hard- and soft-switching is illustrated in Figure 6-1. Note that the reservoir component acts as a temporary depository for the energy stored at C_{drain}. Once the switch is closed, the circuit topology is altered and the stored energy at the reservoir component can be retrieved and returned back to the load network. Soft-switching thus eliminates the power loss associated with the charging and discharging of C_{drain}.

6.3 Class-E operation

The concept of soft switching was first applied to power amplifier by Sokal et al in 1975 in their invention of the Class-E PA [Soka75]. Class-E PA operation is defined by a set of conditions on the drain voltage and current waveforms [Raab77]. We will illustrate these conditions with the conceptual model shown in Figure 6-2.

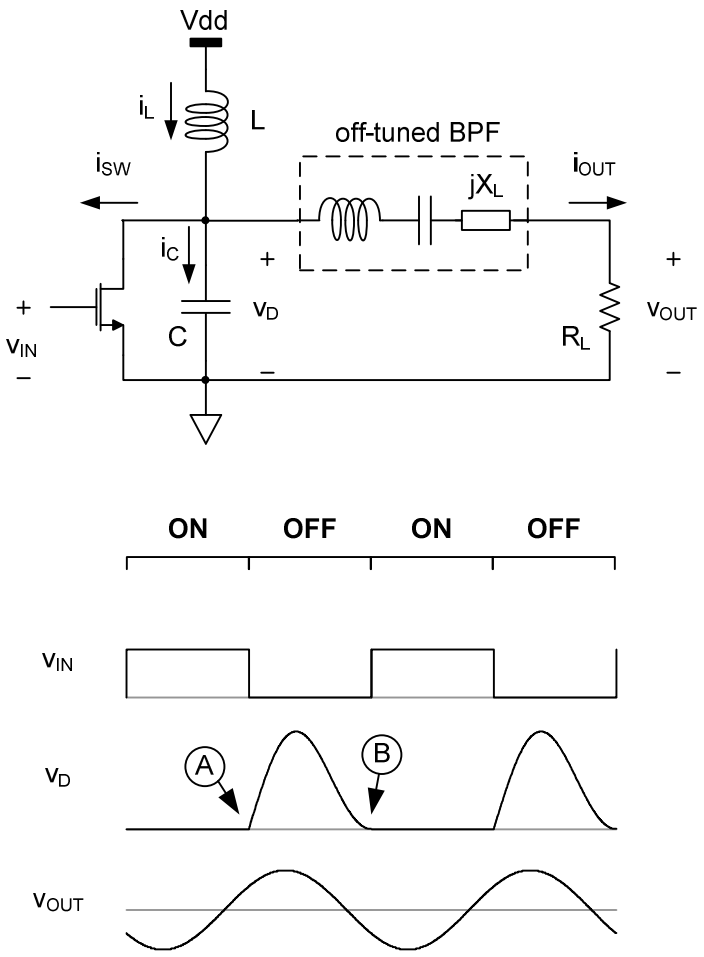


Figure 6-2 A conceptual model of a Class-E PA

The transistor in Figure 6-2 acts as an ON/OFF switch that is controlled by an input signal. The load network comprises a band-pass filter which selectively passes current at the fundamental frequency ω_o to the load resistance R_L. The BPF is slightly off-tuned from the operating frequency so that there is a non-zero phase shift between the

fundamental component of the drain voltage v_D and the output current i_{OUT}. This phase shift is represented by an equivalent series reactance jX_L which can be either capacitive or inductive depending on amount of phase shift. A finite inductor L connects the device drain to the supply. In contrast to a Class-D or Class-F PA where device drain capacitance could impair the circuit performance, the capacitor C (which also includes parasitic capacitance of the device) in the Class-E PA actually serves an essential role in maintaining the Class-E operation.

In operation, the switch is toggled between ON- and OFF- states with approximately 50% duty cycle. During the time when the switch is ON, the drain node is clamped to ground and therefore the supply voltage V_{DD} is applied directly across the inductor L. This results in a linearly ramping inductor current. Part of this ramping current is used to support the sinusoidal output current, while the rest is sunk to ground through the switch. Note that the capacitor current is zero during this time since the voltage across C is constant.

Since neither the drain inductor nor the output BPF can support discontinuity in their current waveforms, the current through the switch is completely transferred to capacitor C when the switch is turned OFF. This causes the drain voltage v_D to rise (marker-A in Figure 6-2). The subsequent voltage excursion of v_D is determined by the transient response of the load network.

The definition of Class-E operation is that the load network is designed in a way such that the voltage v_D will return back to zero with a zero slope immediately before the next turn-ON instance of the switch (marker-B in Figure 6-2). Figure 6-2 shows the resulting steady state waveform of the drain voltage v_D. The role of the BPF is to extract the fundamental component of the drain waveform and pass it to the output; thus creating a sinusoidal output signal that is synchronized in phase and frequency with the input signal. If the input signal is either phase or frequency modulated, the embedded information will be transferred to the output with amplified power.

6.4 Class-E PA characteristics

Since there is ideally no overlap between the device drain voltage and drain current, a Class-E PA can theoretically operate with 100% efficiency. In practice, the achievable efficiency is limited by several secondary loss mechanisms. Nonetheless, when compared with other switch-mode PAs, a Class-E PA exhibits characteristics that alleviate the impact of these secondary effects [Raab78] [Raab78b]:

Effects of Parasitic Capacitance

Since the switch on-resistance is the dominant cause of efficiency degradation in switch-mode PAs, it is generally desirable to maximize the device size in these amplifiers. However, in Class-D and Class-F2 PAs, if we keep increasing the device size eventually

efficiency will be degraded due to losses associated with the increased drain capacitance. These losses are caused by capacitor charging and discharging, as well as mistuning of the load network. On the other hand, a Class E PA can typically tolerate much larger drain capacitance because the amplifier is soft-switching. As a result, larger device size can be used in a Class-E PA which helps to optimize the overall efficiency.

Switch Transition Loss

In high frequency operations, the switching transition time will eventually become an appreciable fraction of the operation period [Kazi83]. During the switching transitions, the device voltage and current can be simultaneously non-zero, which leads to power loss in a switch-mode PA. One example is during the OFF/ON transition in a Class-F2 operation - before the device is turned on, its drain voltage is at $2V_{DD}$. Since it takes time to bring the drain voltage down to ground, the device will operate in the saturation region during the initial moments of the transition period, which results in a direct current path from the supply, through the RFC and the device to ground. As the transition period becomes a large portion on the operation cycle, the associated loss also increases accordingly. The Class-E operation however naturally alleviates losses during both OFF→ON and ON→OFF transitions.

During OFF→ON transition, loss will occur if there is a surge of current through the switch while the drain voltage is still high. In a Class-E PA, zero drain voltage during this

transition is guaranteed by definition. In addition, current serge is also inherently avoided because of the zero dv_D/dt boundary condition during this transition. This ensures that the drain capacitor current is zero before and after the transition, and therefore none of the branches that are connected to the switch can support any current discontinuity during the transition.

During the ON→OFF transition, it takes finite time to remove the channel charge within the transistor and completely shut off the switch. Loss can therefore be induced if there is a suddenly surge in the drain voltage v_D. Since the ramping rate of v_D in a Class-E PA is limited by a large drain capacitance, the drain voltage surge is avoided which results in low loss during this transition.

Dependency on Device Characteristic

The soft-switching characteristic of a Class-E PA has an important consequence: since the drain voltage is low before and after the OFF/ON transition, it allows the device to enter the triode region directly when it turns on. This is in contrast to a hard-switching PA where the transistor may momentarily operate in the saturation region before the drain is pulled low. As a result, a Class-E PA can be very accurately modeled as a switched linear circuit, where its topology toggles between two linear L-R-C networks. In such a switched linear circuit, the voltage and circuit waveforms are not directly related to the current driving capability of the transistor. Instead, they are determined largely by

how the reactive components are charged and discharged in the ON and OFF phrases of the operation cycle. As long as the voltage drop across the device during the ON-phase is small compared with V_{DD}, the exact value of the on-resistance R_{ON} does not significantly affect the operation of the amplifier. Although the overall PA efficiency will still rely on R_{ON}, many characteristics of a Class-E PA, including output power, power gain, and the voltage and current waveforms are rather insensitive to the detailed transistor characteristics. This results in designs that are robust against transistor modeling inaccuracy, as well as process and temperature variation.

Tolerance to Frequency Variation

In wireless applications, the operation frequency may vary slightly as the system switches from one end of the transmission band to the other (PSC1900, for example, has a 60MHz TX-band, which corresponds to ±1.6% variation). Since the drain voltage waveform of a Class-E PA returns to zero with a zero slope at the nominal turn-on instant, small variation in the operation frequency does not significantly affect the turn-on voltage. This results in consistent efficiency over a relatively wide bandwidth.

6.5 Theoretical Analysis of Class-E PA operation

The analysis for a Class-E generally involves solving the time-domain waveforms against the boundary conditions at transitions between the ON and OFF phases. The classical

derivation assumes that an RF choke is used in place of the drain inductor L shown in Figure 6-2. This results in a unique solution of component values that would conform to the Class-E operation criteria [Char89].

In an integrated RF IC environment, the used of true RF choke can occupy large silicon area and introduce significant parasitic resistance that degrades power efficiency. Finite drain inductance is therefore preferable. In this section, we will thus accommodate finite drain inductance. This approach expands the design space and produces a family of Class-E solutions that is parameterized by the resonant frequency of the drain inductor and capacitance; and we will discuss the trade-off among the possible solutions. In order to simplify the derivation, we will assume an ideal switch with zero on-resistance. Although this simplification prevents us from producing an analytical expression for the efficiency, it will preserve all essential characteristics of the circuit's operation.

Output Current

If the Q of the output band-pass filter is sufficiently high, we can assume the output current to take the form of a sinusoid at the operating frequency, and with a phase ϕ and an amplitude I_o:

$$i_{OUT} = I_o \cos(\omega_o t - \phi) \tag{6.1}$$

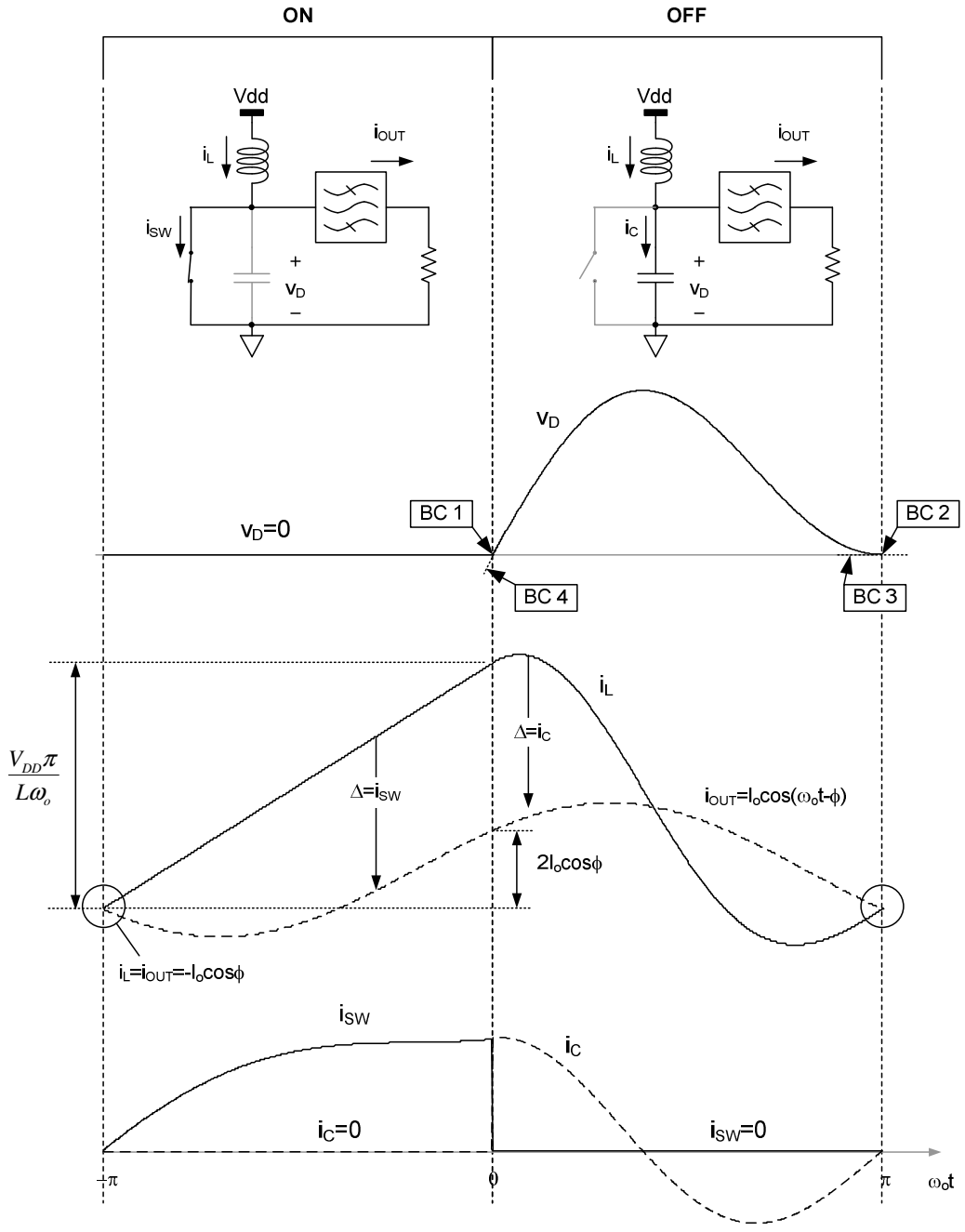


Figure 6-3 Circuit configuration and characteristic waveforms of a Class-E PA

Boundary Conditions

Depicted in Figure 6-3 are the circuit topologies of a Class-E PA during the ON-phase ($-\pi/\omega_o \le t < 0$) and OFF-phase ($0 \le t < -\pi/\omega_o$). During the OFF-phase where the switch is open, the drain voltage v_D is governed by the following differential equation:

$$LC\frac{d^2 v_D}{dt^2} + v_D = V_{DD} + \omega_o L \cdot I_o \sin(\omega_o t - \phi) \tag{6.2}$$

In addition, in order to satisfy the Class-E operation criteria, equation (6.2) must conform to the boundary conditions listed in (6.3).

$$\begin{aligned} &\text{BC1:} \quad v_D(0) = 0 \\ &\text{BC2:} \quad v_D(\pi/\omega_o) = 0 \\ &\text{BC3:} \quad \frac{dv_D}{dt}(\pi/\omega_o) = 0 \\ &\text{BC4:} \quad \frac{dv_D}{dt}(0) = \frac{1}{C}\left(\frac{V_{DD}\pi}{L\omega_o} - 2I_o\cos\phi\right) \end{aligned} \tag{6.3}$$

While BC1-3 comes directly from the assumption of zero switch on-resistance and the definition of Class-E operation, BC4 desires some explanation. First we notice that at the end of the OFF-phase ($t \to \pi/\omega_o$), both the capacitor C and the switch draw no current because $dv_D/dt = 0$, and the switch is open. As a result, the inductor current i_L at $t = \pm\pi/\omega_o$ must equal the output current i_{OUT}. From(6.1), we conclude that

$$i_L(\pi/\omega_o) = i_L(-\pi/\omega_o) = -I_o\cos\phi \tag{6.4}$$

The first equality in (6.4) is due to the periodicity and continuity of the inductor current. We now have the initial inductor current during the ON-phase ($i_L(-\pi/\omega_o)$). After the switch is turned on, v_D is clamped to zero and the supply voltage V_{DD} is applied directly across the inductor L. i_L therefore ramps up linearly from its initial value with a slope that equals to V_{DD}/L. At the end of the ON-phase, its value would reach

$$i_L(0) = \frac{V_{DD}\pi}{L\omega_o} - I_o\cos\phi \tag{6.5}$$

The capacitor current i_C within the entire ON-phase is zero since $dv_D/dt = 0$, the switch current therefore equals to the difference between i_L and i_{OUT}. At the turn-off instant at t=0:

$$i_{SW}(0^-) = i_L(0) - i_{OUT}(0) = \frac{V_{DD}\pi}{L\omega_o} - 2I_o\cos\phi = i_C(0^+) \tag{6.6}$$

Note that we use $i_{SW}(0^-)$ to indicate the switch current at the moment prior to the turn-off instant. This distinction is necessary because, unlike the inductor current and the output BPF current, the switch current can be a discontinuous function of time. In fact, at the turn-off instinct i_{SW} would abruptly jumps to zero as the switch is opened. Since neither the inductor nor the output BPF can support current discontinuity, the current that used to sink through the switch ($i_{SW}(0^-)$) must be entirely transferred into the drain capacitor; and

this corresponds to the last equality in equation (6.6). The initial v_D slope during the OFF-phase is determined by $i_C(0^+)$, which results in the expression of BC4 in (6.3).

Voltage and Current Normalization

Equation (6.2) can be solved directly against the boundary conditions, but the analysis can be simplified with proper normalization. In particular, we would normalize the voltage and current in the system by V_{DD} and $V_{DD}\omega_o C$ respectively. The resulting normalized voltage and current (indicated by the ^ in their symbols) will become dimensionless and can be expressed as:

$$\begin{aligned} \hat{v}_D &= v_D / V_{DD} \\ \hat{I}_o &= I_o / (V_{DDo}\omega_o C) \end{aligned} \tag{6.7}$$

The normalization factors are chosen based on the following observations: Firstly, since the supply voltage V_{DD} is the sole voltage reference in the switched-linear circuit, all voltage and current within the circuit are proportional to $V_{DD.}$ Secondly, we would like to normalize current to the impedance level of the circuit, and one good representation of the impedance level is the drain capacitor admittance $\omega_o C$. In most case where the drain capacitor comprise primarily the inherent capacitance of the switching device, the normalize current would therefore give a good indication of the current density of the switch.

Besides normalization, we also introduce two new variables to replace direct reference to t, L and C.

$$\begin{aligned} \theta &= \omega_o t \\ \varepsilon &= \omega_o \sqrt{LC} \end{aligned} \tag{6.8}$$

Applying (6.7) and (6.8) to the differential equation(6.2):

$$\varepsilon^2 \frac{d^2 \hat{v}_D}{d\theta^2} + \hat{v}_D = 1 + \varepsilon^2 \hat{I}_o \sin(\theta - \phi) \tag{6.9}$$

Note that (6.9) represents a family of differential equations that are parameterized by the design variable ε which is the ratio of the operation frequency and the resonant frequency of the drain inductor and capacitor. Equation (6.9) can be solved against the boundary conditions for any given value of ε to produce a Class-E solution. Applying the same normalization to the boundary conditions in (6.3):

$$\begin{aligned} &\text{BC1: } \hat{v}_D(0) = 0 \\ &\text{BC2: } \hat{v}_D(\pi) = 0 \\ &\text{BC3: } \frac{d\hat{v}_D}{dt}(\pi) = 0 \\ &\text{BC4: } \frac{d\hat{v}_D}{dt}(0) = \frac{\pi}{\varepsilon^2} - 2\widehat{I_o} \cos\phi \end{aligned} \tag{6.10}$$

Family of Class-E Solutions

The general solution to (6.9) is in the form of:

$$\hat{v}_D(\theta) = A\cos(\theta/\varepsilon) + B\sin(\theta/\varepsilon) + \frac{\hat{I}_o \sin(\theta-\phi)}{1/\varepsilon^2 - 1} + 1 \tag{6.11}$$

Treating ε as a design parameter, the four unknowns, A, B, $\hat{I}_o$ and ϕ, can be determined by solving (6.11) against (6.10). With some manipulation, we can show that:

$$\begin{aligned} A &= -\left[\frac{\varepsilon\pi}{2} - (1-\varepsilon^2)\cdot\tan(\frac{\pi}{2\varepsilon})\right]\cdot\tan(\frac{\pi}{2\varepsilon}) - 1 \\ B &= \frac{\varepsilon\pi}{2} - (2-\varepsilon^2)\cdot\tan(\frac{\pi}{2\varepsilon}) \\ \hat{I}_o \sin\phi &= \frac{1-\varepsilon^2}{\varepsilon^2}(A+1) \\ \hat{I}_o \cos\phi &= \frac{1-\varepsilon^2}{\varepsilon^2}\left[\frac{\pi}{2} + \varepsilon\cdot\tan(\frac{\pi}{2\varepsilon})\right] \end{aligned} \tag{6.12}$$

Hence we have derived a family of solutions (parameterized by ε) of the normalized drain voltage that would satisfy the Class-E operation criteria. Once $\hat{v}_D(\theta)$ is obtained, the normalized output, inductor, switch currents can be easily derived. Figure 6-4 depicts the family of Class-E solutions corresponding to different values of ε. The minimum ε value shown corresponds to the smallest ε that results in non-negative switch current.

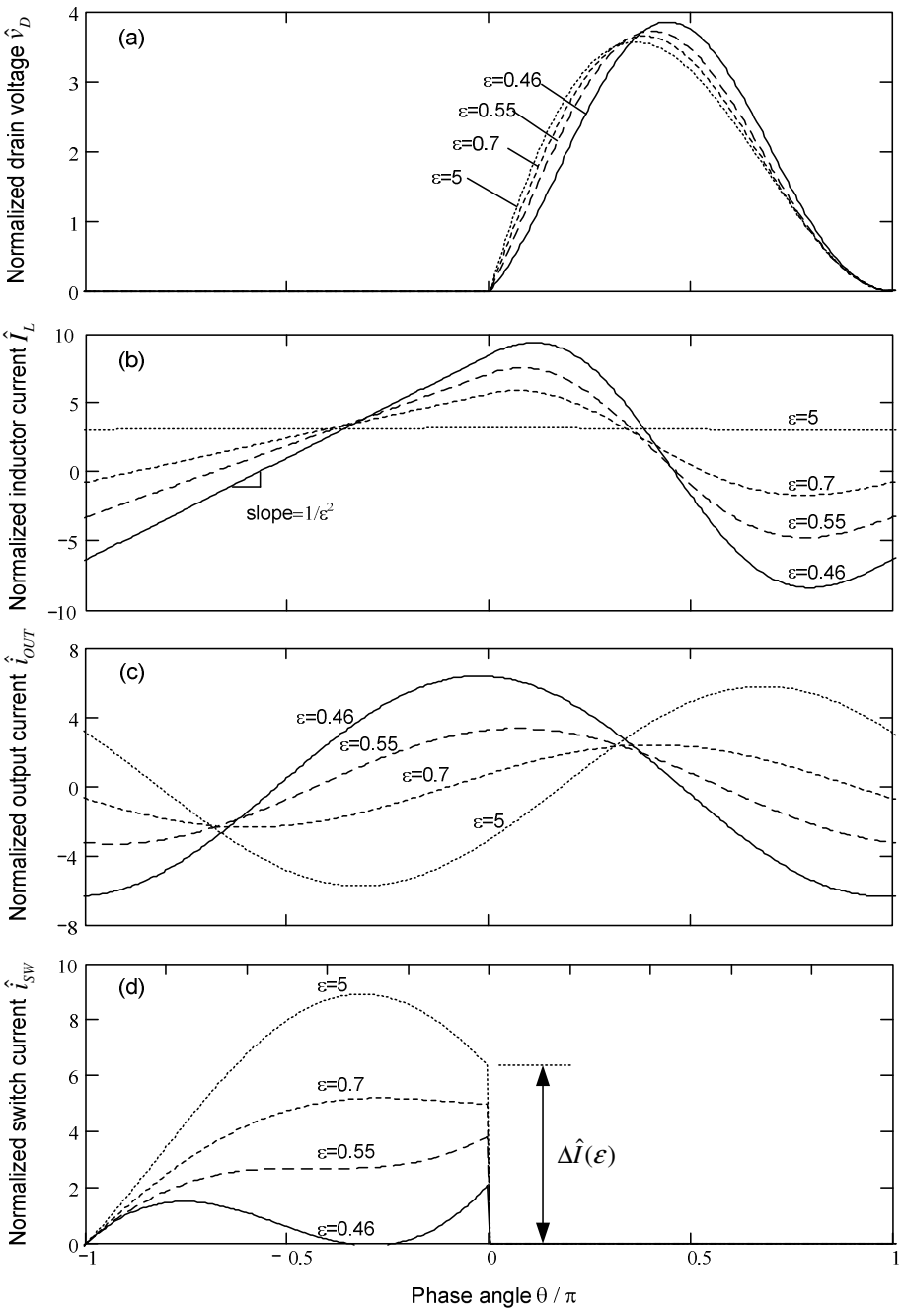


Figure 6-4 Family of Class-E solutions corresponding to different values of parameter ε

As shown in Figure 6-4(a), the shape of the normalized drain voltage is relatively insensitive to ε. This is because Class-E operation tightly restricts the initial value, the final value, and the final slope of the drain voltage during the OFF-phase. In addition, the area under the $\hat{v}_D(\theta)$ curves is invariant because the DC short provide by the inductor L clamps the average drain voltage to V_{DD} (corresponding to average $\hat{v}_D$=1).

The normalized inductor current $\hat{i}_L$ is shown in Figure 6-4(b). During the ON-phase, it increases linearly at the rate of $1/\varepsilon^2$. As we increase parameter ε, the solution would eventually approach the classical case where the drain inductance L is infinite. We can show that in this limit, $\hat{i}_L$ would asymptotically converge to a constant value of π, which corresponds to a supply power of $\pi V_{DD}^2 \omega_o C$.

Load Impedance Termination

Figure 6-4(c) indicates that the amplitude and phase of the normalized output current $\hat{I}_{out}$ is sensitive to ε. The fundamental component of the normalized drain voltage ($\hat{v}_{D1}$) can be derived by taking Fourier series of $\hat{v}_D(\theta)$:

$$\begin{aligned}\hat{v}_{D1}(\theta) = &\left[\frac{\varepsilon^2}{1-\varepsilon^2} + (\frac{\pi}{4} - \frac{2}{\pi})\varepsilon\tan(\frac{\pi}{2\varepsilon}) - (\frac{1-\varepsilon^2}{2})\tan^2(\frac{\pi}{2\varepsilon})\right]\cos\theta \\ &+\left[\frac{\pi}{4} + \frac{2}{\pi(1-\varepsilon^2)} + \frac{\varepsilon}{2}\tan(\frac{\pi}{2\varepsilon})\right]\sin\theta\end{aligned} \tag{6.13}$$

Amplitudes of $\hat{v}_{D1}$ and $\hat{I}_{out}$, and their phase difference are plotted against ε in Figure 6-5. The normalized load network impedance ($\hat{Z}_1$) at the fundamental frequency can be derived from $\hat{v}_{D1}$ and $\hat{i}_{OUT}$ according to (6.14)

$$\begin{aligned} \mathrm{Re}(\hat{Z}_1) &= \left|\frac{\hat{v}_{D1}}{\hat{I}_o}\right| \cos(\angle\hat{v}_{D1} - \phi) \\ \mathrm{Im}(\hat{Z}_1) &= \left|\frac{\hat{v}_{D1}}{\hat{I}_o}\right| \sin(\angle\hat{v}_{D1} - \phi) \end{aligned} \tag{6.14}$$

In Figure 6-6, we plot he trajectory of $\hat{Z}_1$ on the complex plane as a function of ε. $\hat{Z}_1$ is inductive at small values of ε, but becomes purely real value at ε = 0.71, and turns capacitive at larger values of ε. As ε approaches infinity, $\hat{Z}_1$ converges to a constant value. The asymptotic limits of $\hat{v}_{D1}$, $\hat{i}_{OUT}$ and $\hat{Z}_1$ as ε approaches infinity are:

$$\begin{aligned} \lim_{\varepsilon\to\infty} \hat{v}_{D1} &= (\frac{\pi^2}{4} - 2)\cos\theta + \frac{\pi}{2}\sin\theta \\ \lim_{\varepsilon\to\infty} \hat{i}_{OUT} &= -\pi\cos\theta + \frac{\pi^2}{2}\sin\theta \\ \lim_{\varepsilon\to\infty} \hat{Z}_1 &= \frac{8}{\pi(\pi^2+4)} - j\frac{\pi^2-4}{2(\pi^2+4)} \end{aligned} \tag{6.15}$$

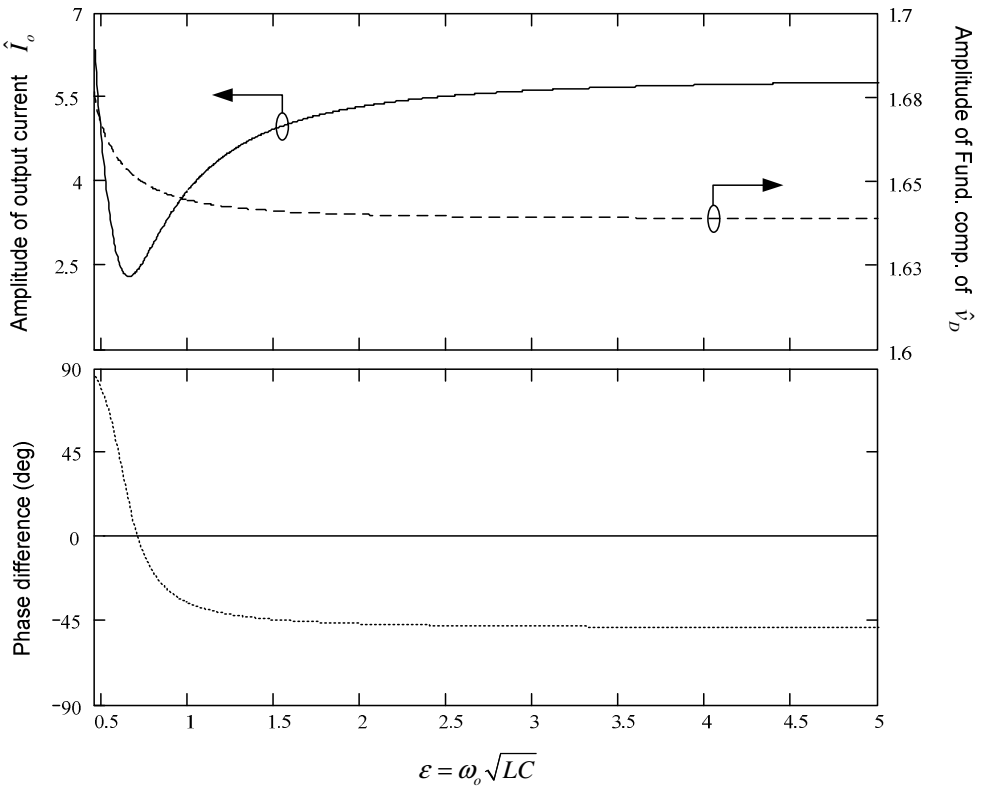


Figure 6-5 Fundamental component of normalized drain voltage and output current vs. ε

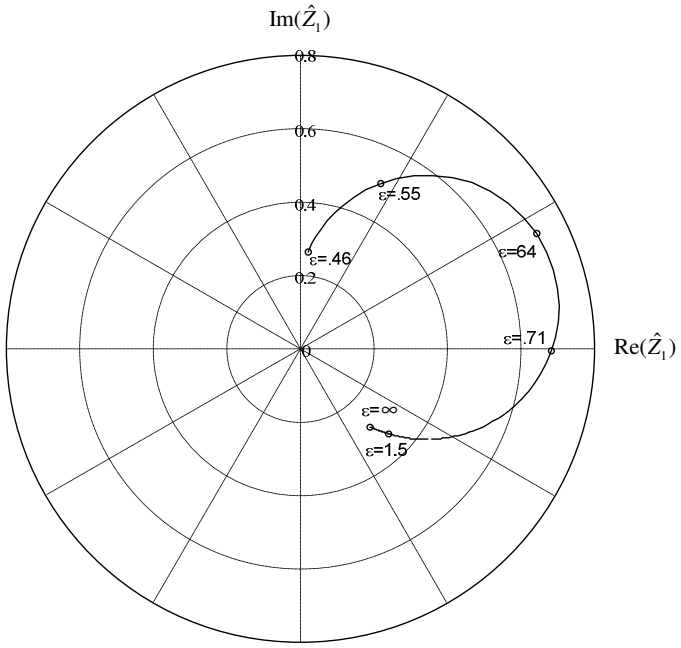


Figure 6-6 Trajectory of normalized load network impedance

Finally, the load component values of R_L and X_L can be determined by de-normalizing $\hat{Z}_1$ as shown in (6.16).

$$\begin{aligned} R_L &= \mathrm{Re}(\hat{Z}_1)/\omega_o C \\ X_L &= \mathrm{Im}(\hat{Z}_1)/\omega_o C \end{aligned} \tag{6.16}$$

Output Power

In [Moln84], Molnar proved that the output power is related to the current discontinuity $\Delta\hat{I}(\varepsilon)$ as shown in Figure 6-4. He proved that P_{OUT} can be expressed as a function of the time derivatives of the switch current and voltage:

$$P_{OUT} = -\frac{T}{4\pi^2}\int_0^T \frac{dv_D}{dt}\cdot\frac{di_{SW}}{dt}\,dt \tag{6.17}$$

where $T=2\pi/\omega_o$ is the period of operation. Since $dv_D/dt=0$ during the ON-phase and $di_{SW}/dt = 0$ during the OFF-phase, the only time instant where the integral in (6.17) can return a non-zero value is during the turn-off ($t=0$) instant where i_{SW} is discontinuous. Evaluating di_{SW}/dt at $t=0$:

$$\begin{aligned}\left.\frac{di_{SW}}{dt}\right|_{t=0} &= \lim_{\xi\to 0}\frac{i_{SW}(\xi/2)-i_{SW}(-\xi/2)}{\xi}\\ &= -\Delta I\,\delta(t)\end{aligned} \tag{6.18}$$

where $\delta(t)$ is the Dirac delta function and ΔI is the (un-normalized) magnitude of the current discontinuity at $t=0$. dv_D/dt at $t=0$ can be computed as:

$$\begin{aligned}\left.\frac{dv_D}{dt}\right|_{t=0} &= \lim_{\xi\to\infty}\frac{v_D(\xi/2)-v_D(-\xi/2)}{\xi}\\ &= \frac{1}{2}\frac{dv_D}{dt}(0^+) = \frac{\Delta I}{2C}\end{aligned} \tag{6.19}$$

Substituting (6.18) and (6.19) into (6.17), and applying the usual normalization, the normalized output power becomes

$$\hat{P}_{OUT} = \frac{\Delta \hat{I}^2}{4\pi} \tag{6.20}$$

where $\Delta \hat{I}$ is the normalized version of the current step ΔI (see Figure 6-4(d)). (6.20) therefore implies that discrete current jump at the turn-off instant is a necessary condition for the Class-E PA to deliver non-zero output power. Another implication is that in practice when the device on-resistance is finite, the drain voltage at the turn-on instant can be close to, but never truly equals zero.

Applying (6.6) and (6.12) to (6.20), $\hat{P}_{OUT}$ can be expressed in terms of ε:

$$\hat{P}_{OUT} = \frac{1}{4\pi}\left[\pi + 2(\varepsilon - \frac{1}{\varepsilon})\tan(\frac{\pi}{2\varepsilon})\right]^2 \tag{6.21}$$

The normalized output power is plotted against ε in Figure 6-7, which shows that $\hat{P}_{OUT}$ increases monotonically with ε, and approaches π as ε goes to infinity (i.e. the same asymptotic value of the DC power from supply since efficiency is 100%).

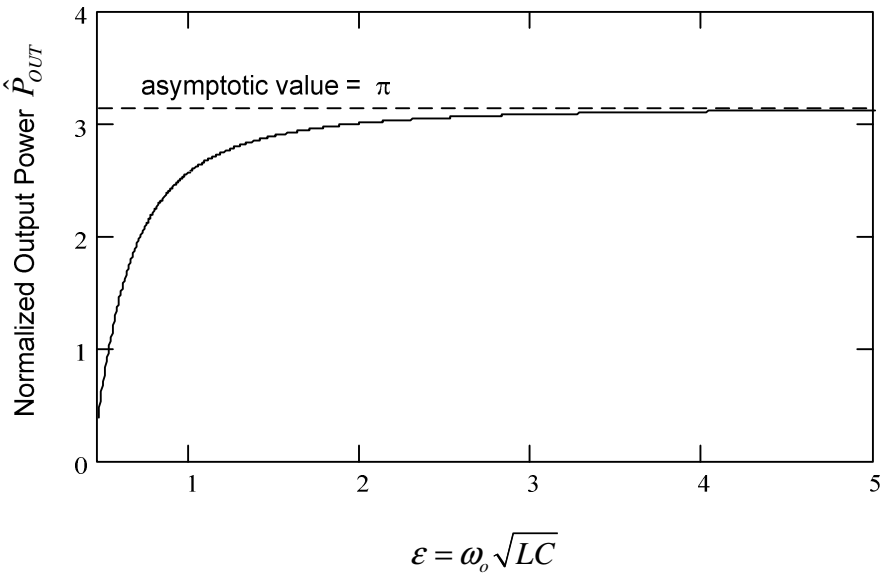


Figure 6-7 Normalized output power vs. ε

Power Capacity

Referring back to Figure 6-4(a), the normalized drain voltage reaches a value greater than 3.5 during the OFF-phase. This corresponds to the peak drain voltage reaching $3.5V_{DD}$. Such peak v_D value is significantly higher than that of a Class-D or F2 PA and is a direct consequence of the Class-E waveform criteria - since the average value of v_D is fixed at V_{DD}, imposing a zero v_D and zero slop at the turn-on instant thus push the peak value above $2V_{DD}$. We therefore expect a Class-E PA to exhibit somewhat inferior power capacity when compared with the other switch-mode PAs previously introduced. Figure

6-8 shows the peak $\hat{v}_D$ and the peak $\hat{i}_{SW}$ versus parameter ε. Finally, the power capacity can be evaluated as

$$PC = \frac{\hat{P}_{OUT}}{\hat{v}_{D,pk}\hat{i}_{SW,pk}} \tag{6.22}$$

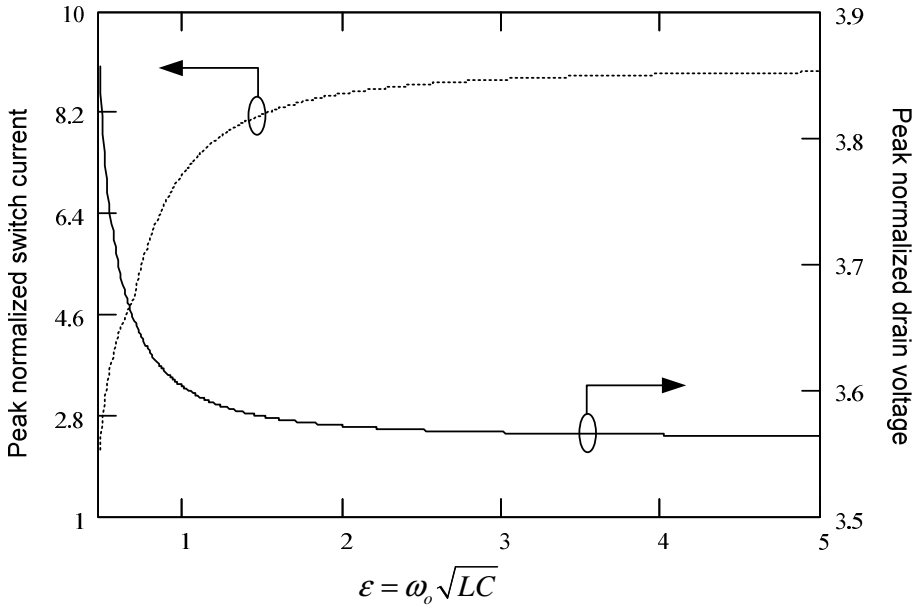


Figure 6-8 Peak Normalized drain voltage and switch current

Power capacity of a Class-E PA is plotted against ε in Figure 6-9. Notice that for ε > 1.5, the power capacity is approximately constant. The asymptotic value as ε goes to infinity

is 0.102. At $\varepsilon = 0.68$, the power capacity peaks slightly at 0.107. However, the power capacity drops sharply for smaller values of ε. As ε is reduced to below 0.68, the switch current peaks at the turn-off instant, where the peak switch current ($= \Delta I$) drops at a slower rate than the output power ($\propto \Delta I^2$ from (6.20)), resulting in a fast roll-off in the power capacity.

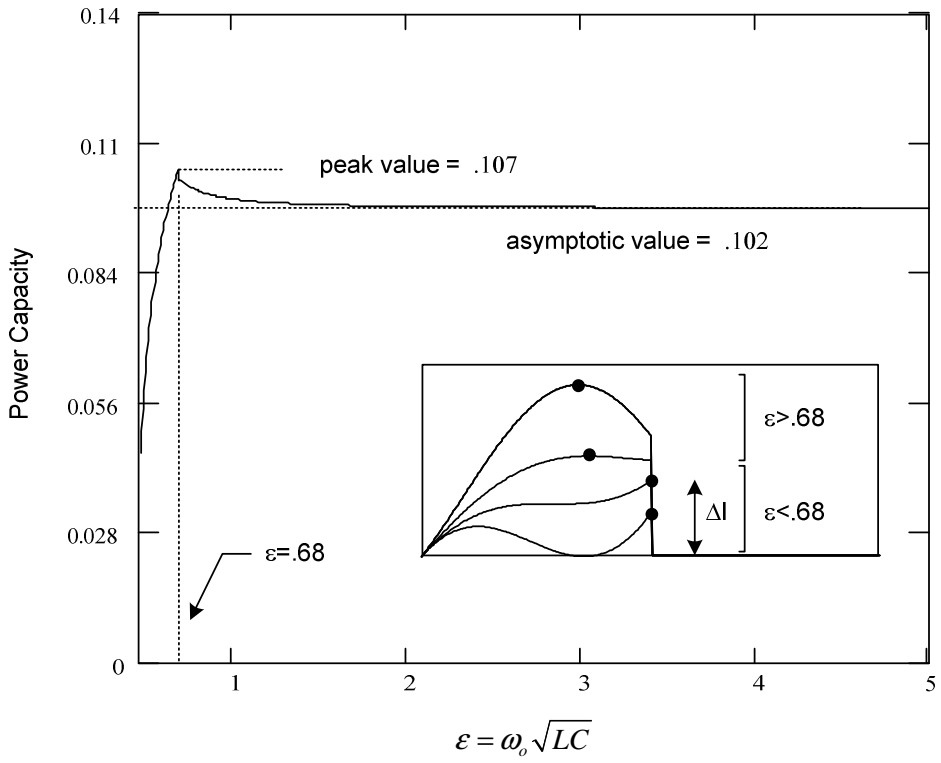


Figure 6-9 Power capacity as a function of ε

6.6 Considerations on Power Control

In a Class-E PA, the input signal provides only timing information to synchronize the circuit. It is therefore impossible to control the transmitted output power by varying the input signal amplitude. On the other hand, the switched linear nature of Class-E PA allows an effective way to control output power through a variable supply, especially when provided by a DC-DC converter. Since the voltage of every node within the switched linear circuit is proportional to V_{DD}, all power terms are proportional to V_{DD}^2. As a result, the output power can be controlled by the supply voltage, and in a way that maintains a constant efficiency over a wide range of output power. This is illustrated in Figure 6-10, in which we assume, without loss of generality, that the only loss is from the finite switch on-resistance R_{ON}. Since both the loss and the output power scale with V_{DD}^2, their ratio, and the overall efficiency, is unaffected as the output power is adjusted through the variable supply

$$P_{OUT}, P_{LOSS} \propto V_{DD}{}^2 \Rightarrow \text{efficiency} = \frac{P_{OUT}}{P_{OUT} + P_{LOSS}} = \text{constant} \tag{6.23}$$

This is in contrary to conventional power amplifiers using fixed supply, where efficiency typically peaks only at the maximum output power. Because the PA in reality may spend most of its time operating at medium and low power levels [Sevi97], constant efficiency of the Class-E PA can potentially result in substantial power saving.

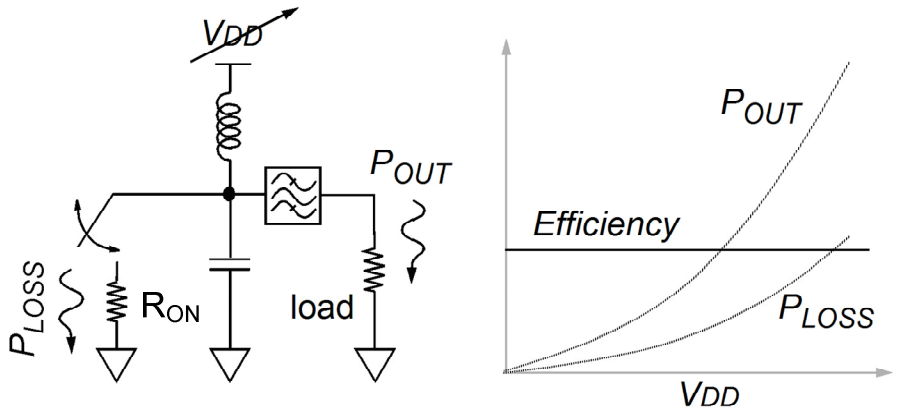


Figure 6-10 Constant efficiency over supply voltage

The switched linear nature of Class-E PA has also been exploited to extend its usage to non-constant envelope modulated signals. It is mainly achieved with a linearization scheme known as Envelope Elimination and Restoration (EER) [Kahn52] [Su98]. In an EER transmitter, a complex modulated signal is separated into a purely phase-modulated (PM) RF signal and an amplitude-modulating (AM) baseband signal. While the PM signal is amplified by a Class-E PA as usual, the AM signal modulates the PA supply. As a result, the complex modulated signal is reconstructed at the PA output with amplified power. The main challenge for EER is to achieve a wide bandwidth on the supply modulation path, but for relatively narrowband signals EER can be an effective way to improve the overall PA efficiency.

6.7 Summary

The Class-E PA design approach offers many desirable characteristics in efficiency enhancement. However, traditional Class-E PA design also poses challenges for CMOS realization. In particular, the low CMOS transistor breakdown voltage is of concern because the maximum drain voltage of a Class-E PA can reach more than three times the supply voltage. In many CMOS processes, the transistor oxide break-down voltage is adjusted to approximately 1.7 x V_{DD}. To achieve reliable operation, we thus need to lower the PA supply to below the nominal rating. Yet, delivering a given power with reduced supply voltage implies lowering the impedance level of the amplifier and using larger switches. The input capacitance of the resulting devices may reach a value so large that it is no longer possible to tune out with on-chip or bondwire inductors.

In addition, the large gate-to-drain capacitance of a transistor may also induce coupling within the PA and causes stability issues. From a noise coupling perspective, the singe-ended circuit in Figure 6-2 injects large amount of current to ground (and the substrate) once per cycle. This generates an unwanted noise component at the operation frequency and this is particularly undesirable in an integrated environment. In order to adapt the Class-E PA into an efficient realization in CMOS, the classical design will have to be modified and be enhanced to overcome the afore-mentioned challenges. We will discuss the potential techniques in Chapter 7

Chapter 7 - Design of a 1-W 1.9GHz CMOS Class-E PA

7.1 Introduction

In this chapter, we present the design considerations of a CMOS Class-E PA that delivers 1-W of output power at 1.9GHz. The target application is a GSM type of cellular system which utilities a constant envelope modulation scheme such as Gaussian Minimum Shift Keying (GMSK). For these systems, PA efficiency is a key performance criterion because power amplifiers typically dominate the overall current consumption during active transmits.

We have chosen Class-E operation as the basis of the design because of its inherent high efficiency. In particular, the distinctive soft-switching characteristic of a Class-E PA is effective in alleviating the charging and discharging energy loss in the heavily capacitive CMOS PA. Nonetheless, the classical single-end Class-E PA has drawbacks regarding input driving requirement, low impedance level, sensitivity to package parasitic and potential noise-coupling. These factors ultimately degrade the achievable performance in an integrated CMOS environment. It is the focus of this chapter to circumvent these drawbacks by incorporating a fully differential topology and by using injection locking technique.

7.2 Circuit Schematic

Figure 7-1 illustrates the overall schematic of the fully differential Class-E CMOS PA. The amplifier is formed by two cascaded gain stages. At the output stage, the drain inductor (L3, L4) and the output matching network are configured such that the voltages at the two drain nodes follow the desirable Class-E characteristics – i.e. they return to ground with a zero slope at the moment the switch is turned on. Within each gain stage, two cross-coupled assisting devices (M5-M8) are inserted. Because of the differential topology, each of the assisting devices is switched synchronously with the devices that they are connected in parallel with. The gate capacitance of these assisting devices served as part of the total drain capacitors that are required to tune the amplifier into Class-E operation. At the same time, their presence reduces the size of the input devices (M1-M4)

in each stage and significantly relaxes the input driving requirement. Inserting cross-coupled pairs effectively turn each gain stage into an injection locking oscillator and we will discuss the injection locking mechanism in greater details later in this chapter.

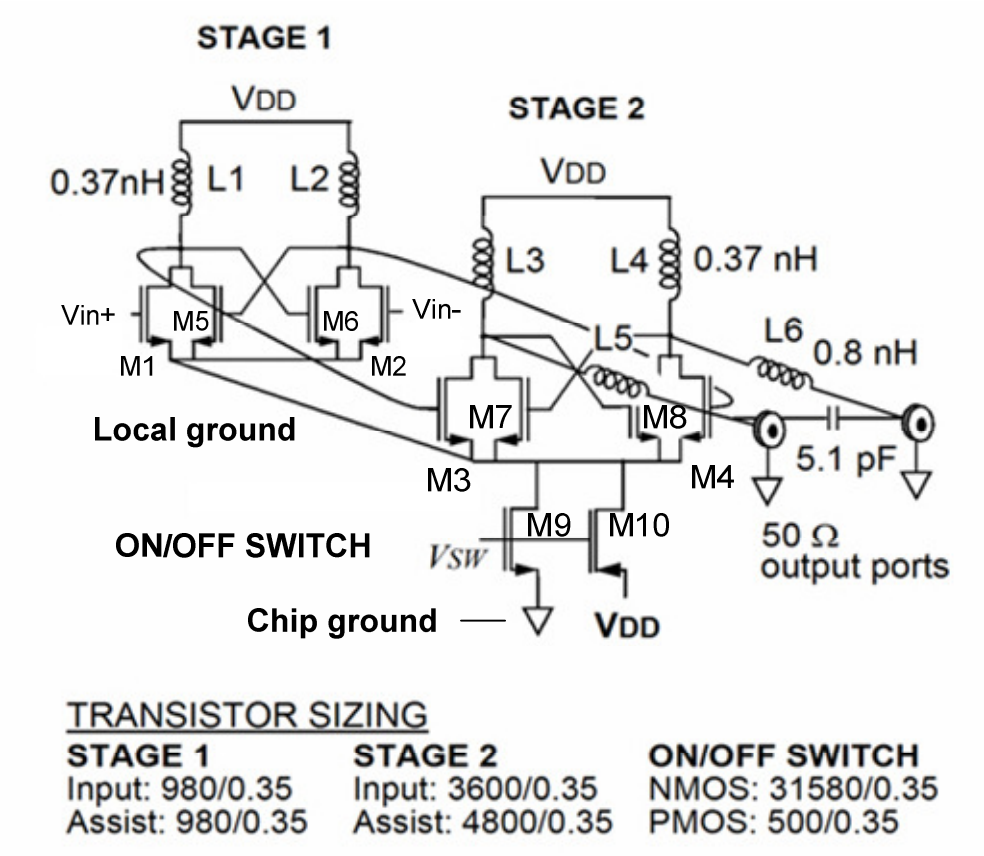


Figure 7-1 Schematic of the 1-W 1.9GHZ CMOS Differential Class-E power amplifier

Power-down Mechanism

A power-down mechanism is necessary to avoid free oscillation when the PA is not driven. To serve this need, we added an ON/OFF switch to the local ground node. During power up, the ON-switch (M9) is turned on to provide a low-loss current path between the local ground and the chip ground. Since the on-resistance of M9 would degrade the power efficiency, a very large size device is needed. Fortunately, it is straight forward in this case because the amplifier can tolerate large capacitance along the ground path with little impact on the overall performance. During power down, M9 is turned off and a small PMOS transistor M10 pulls the local ground node to V_{DD}, effectively shutting down the entire amplifier.

Floorplan and Layout Considerations

Figure 7-2 (a) depicts the overall floorplan of the power amplifier. Placement of devices is optimized for the simplest internal routing. In our CMOS process, five layers of interconnect metal with different sheet resistance (M1: 200mΩ/sq, M2-4: 50 mΩ/sq, M5: 35 mΩ/sq) are available. Since parasitic resistance of on-chip interconnect is a significant cause of efficiency degradation, special attention is paid to minimize interconnect loss. This is mainly done by placing bonding pads in proximity of the active devices, and by strapping multiple metal layers in parallel along high current paths wherever possible.

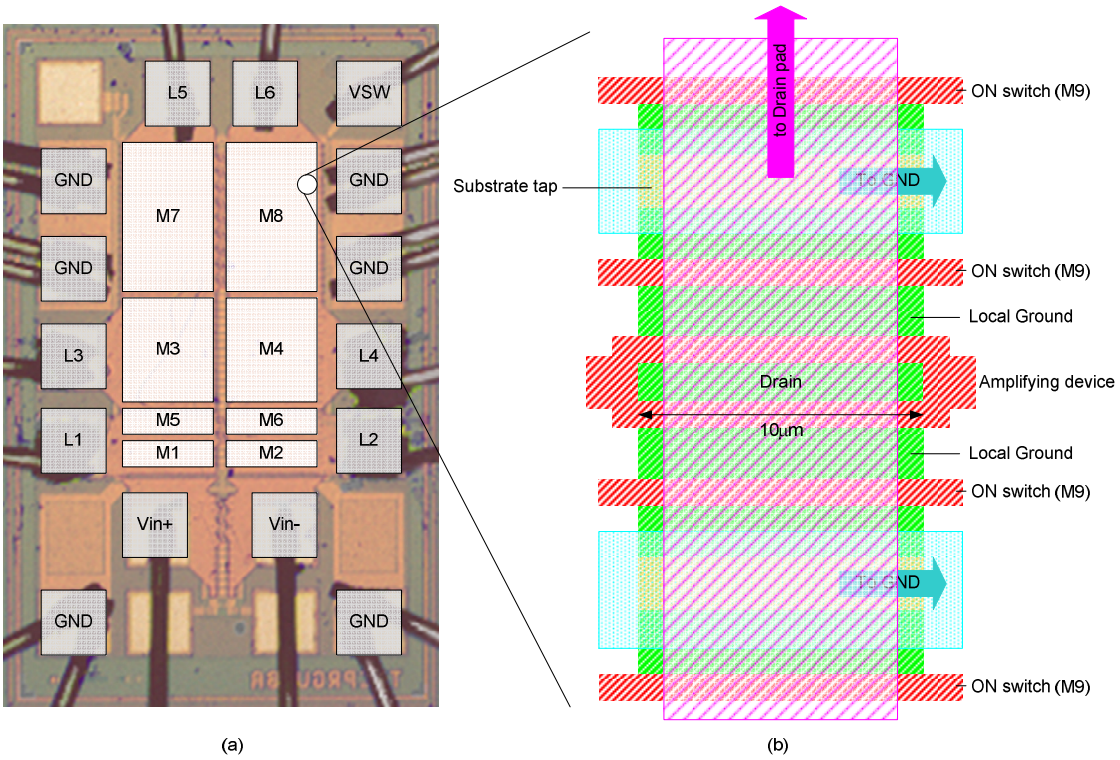


Figure 7-2 (a) overall floorplan (b) layout of unit cell

Each of the amplifying devices (M1-M8) in Figure 7-1 is formed by an array of unit cells. Figure 7-2 (b) illustrates the unit cell layout. Since the highest current path is along the drain and source connections, the unit cell layout is arranged to have wide metal connection on these paths. The drain capacitance of the amplifying devices is minimized by reducing the diffusion area to minimum allowed by the design rules. A row of stacked vias bring the drain node of each amplifying device up to a big slab of top metal which eventually leads to the corresponding bond pads and bondwire inductors.

The ON-switch device (M9) is distributed among the amplifying devices to ensure a small total on-resistance and even current distribution. The ON-switch device connects the source (local ground) of all amplifying devices to the shared chip ground which also contain substrate taps. Finally, the chip ground is brought to the GND bonding pads through wide and stacked lower-metal interconnect running horizontally.

7.3 Inductor realization

High quality-factor (Q) inductors are essential in achieving good power efficiency. This is especially true at the output stage of the power amplifier where large amount of current circulates in the inductors; and even a small amount of parasitic resistance can cause significant efficiency loss. The Q-factor of monolithic spiral inductors is limited by the resistive loss of the thin metal interconnects [Nguy90]. To achieve the maximum inductor Q, we have chosen to realize all inductors with aluminum bondwires. The relatively large cross sectional area of bondwires significantly reduces conduction loss and experimental results show that Q-factor of 40 is achievable. One main challenge we faced was accurate prediction of the inductance values. This requires detail knowledge of the bonding profile, the ground return path, and the presence of adjacent conductors. To address this problem, we first rely on Greenhouse equations to provide a first order estimation of the bondwire inductance [Gree74]:

$$L_{self} = \frac{l}{5}\left[\ln(\frac{2l}{r}) - \frac{3}{4} + \frac{r}{l}\right]\frac{nH}{mm} \tag{7.1}$$

$$M_u = 0.2l\left[\ln\left[\frac{1}{gmd} + \sqrt{1+(\frac{1}{gmd})^2}\right] - \sqrt{1+(\frac{gmd}{l})^2} + \frac{gmd}{l}\right]\frac{nH}{mm} \tag{7.2}$$

where r and l are the radius and length of two coupled cylindrical conductors, and gmd is the geometric mean distance in between. We then validated the results with EM simulations, and finally the test board was designed to have large degree of freedom in realizing different bonding configurations so that the optimal values can be achieved experimentally with manual bonding.

Resistive loss of bondwire are generally frequency dependent because current tends to crowd towards the surface of the conductor at high frequency due to skin effect. To the first order, we can approximate the AC resistance to be

$$R_{ac} = \frac{\rho \cdot l}{\pi\left[r^2 - (r-\sigma)^2\right]} \tag{7.3}$$

where ρ and r are the resistivity and radius of the cylindrical conductor respectively; and σ is the skin depth at the frequency of operation f_o:

$$\sigma = \sqrt{\frac{\rho}{\mu \cdot \pi \cdot f_o}} \tag{7.4}$$

where μ is the permeability of the conductor (=$4\pi 10^{-7}$H/m for non-magnetic material).

7.4 Differential Topology

Traditionally, discrete and hybrid PAs are mostly implemented in single-ended topology. Such design minimizes the transistor count and simplifies the interface with single-ended filters and antenna. For integrated, low-voltage CMOS PAs, the use of differential topology however presented a number of important advantages.

For the same amount of delivered output power, a differential PA presents a higher impedance load network and therefore is less susceptible to efficiency degradation due to parasitic losses. Also, since the differential configuration confines the fundamental and odd harmonics of the ground and supply current on chip, it minimizes the circuit's sensitivity to package inductance. In places where the fundamental and odd harmonics do cross the chip boundary, e.g. at the differential output, effect of package inductance can still be alleviated by exploiting mutual inductance between anti-phased current paths.

From a stability point of view, multiple gain stages are often used in a low-voltage integrated PA because of the limited realizable gain per stage; and differential PAs are significantly less vulnerable to oscillation caused by coupling among gain stages. From a substrate noise injection perspective, a differential PA reduces the noise injection amplitude and also shifts the injection frequency to double the operation frequency. In an

integrated IC environment, this helps to reduce potential interfere to sensitive blocks such as the Low Noise Amplifier and the Frequency synthesizer.

Effects on Load Network Impedance

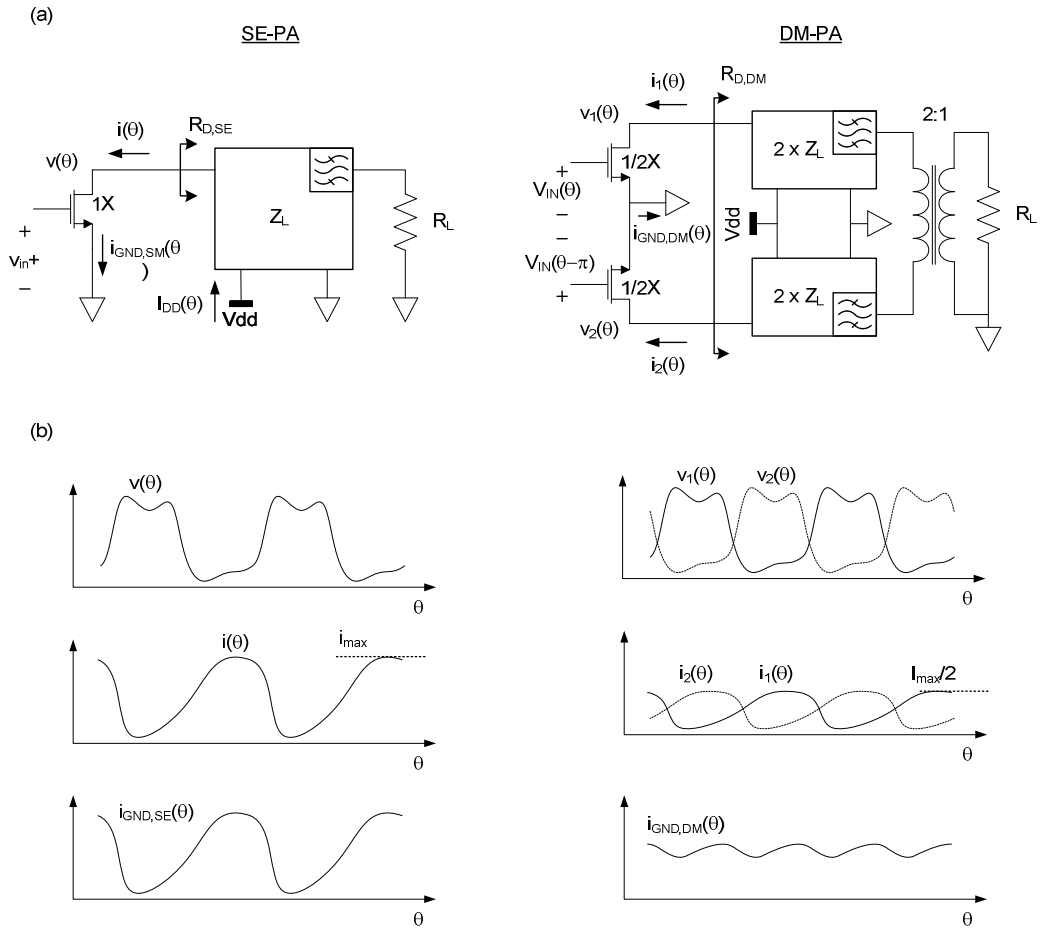


Figure 7-3 Comparison between a single-ended and a differential PA

To examine the effect of differential topology on load impedance, let us compare the single-ended PA (SE-PA) and the differential mode PA (DM-PA) shown in Figure 7-3(a). The total active device area is kept the same; but in the case of the DM-PA, the total circuit is split into two halves and driven differentially. Through the ideal transformer, the differential output of the DM-PA is converted into a single-ended signal that is delivered to the load resistance R_L.

The is a subtle difference between the output harmonic termination of the two PAs – whereas the SE-PA output is terminated by the same impedance (R_L) at all frequency, the DM-PA output faces high impedance termination at even order harmonics as a result of the common-mode response of the transformer. We will however ignore this minor difference because in most cases the output branch of the PA would have a band-pass frequency response; and the amount of high-order harmonics that will eventually be delivered to R_L is limited.

Comparing the DM-PA half circuit and the SE-PA in Figure 7-3(a), we notice that the two are essentially the same except for the impedance level. Since they are driven by the same input signal $v_{IN}(\theta)$, the voltage waveforms in the corresponding nodes of the two circuits are identical. However, because of the difference in impedance level, the current on the corresponding branches of DM-PA half circuit will be only half of that in the SE-PA.

The voltage and current between the DM-PA half circuits are 180^o phase shifted. Let v, v_1 and v_2 be the voltage at the drain node of the SE-PA, and in each of the DM-PA half circuits respectively. Similarly, i, i_1 and i_2 denote the corresponding drain currents. We have

$$\begin{aligned} v_1(\theta) &= v(\theta) \qquad v_2(\theta) = v(\theta - \pi) \\ i_1(\theta) &= i(\theta)/2 \qquad i_2(\theta) = i(\theta - \pi)/2 \end{aligned} \tag{7.5}$$

The total supply current i_{DD} of the DM-PA can be related to that of the SE-PA by

$$i_{DD,DM}(\theta) = \left[i_{DD,SE}(\theta) + i_{DD,SE}(\theta - \pi) \right] / 2 \tag{7.6}$$

Given (7.5), (7.6), and the fact that the device is the only lossy component within the circuit, we can show that the DM-PA and the SE-PA deliver the same amount of output power to the load:

$$\begin{aligned} P_{OUT,DM} &= P_{DC} - P_{Device} \\ &= \frac{V_{DD}}{2\pi} \int_{-\pi}^{\pi} i_{DD,DM}(\theta) d\theta - \frac{1}{2\pi} \int_{-\pi}^{\pi} v_1(\theta) i_1(\theta) + v_2(\theta) i_2(\theta) d\theta \\ &= \frac{V_{DD}}{2\pi} \int_{-\pi}^{\pi} i_{DD,SE}(\theta) d\theta - \frac{1}{2\pi} \int_{-\pi}^{\pi} v(\theta) i(\theta) d\theta \\ &= P_{OUT,SE} \end{aligned} \tag{7.7}$$

Let $Z_{D,DM}$ be the load impedance (at the operating frequency) as seen across the drains of DM-PA; and $Z_{D,SE}$ be the load impedance between drain of the SE-PA and ground, we can show that

$$\begin{aligned} Z_{D,DM} &= -\frac{\int_{-\pi}^{\pi}\left[v_1(\theta)-v_2(\theta)\right]e^{-j\theta}d\theta}{\int_{-\pi}^{\pi}\left[\frac{i_1(\theta)-i_2(\theta)}{2}\right]e^{-j\theta}d\theta} \\ &= \frac{\int_{-\pi}^{\pi}\left[v(\theta)-v(\theta-\pi)\right]e^{-j\theta}d\theta}{\int_{-\pi}^{\pi}\left[\frac{i(\theta)-i(\theta-\pi)}{4}\right]e^{-j\theta}d\theta} \\ &= 4Z_{D,SE} \end{aligned} \tag{7.8}$$

In other words, for the same delivered output power, the impedance level of the load network as referred to the drain is four times higher in a DM-PA than in a corresponding SE-PA. This conclusion is fairly general and is independent of whether the PA is gm-based or switch-mode. The only assumption made is that the output signal has no strong even order harmonics.

In a PA that delivers high output power, high load network impedance level is beneficiary because it reduces loss due to parasitic resistance inherent to the load network. One example of such parasitic is the contact resistance of the output wirebond pads. The DM-PA is subjected to double the amount of contact resistance at its output

due to the differential configuration. However, since the DM-PA has 4 times higher load impedance than the corresponding SE-PA, the loss due to the contact resistance will be half of that in the SE-PA.

Effects on Package Parasitic at Ground Connection

A main drawback of single-ended PAs is its sensitivity to the parasitic ground inductance. In a typical IC package, the ground node of the PA is wire-bonded either directly to the die attach area or to package pins similar to the illustration in Figure 7-4. Depending on the size of the package, the bondwires and pin leads can present significant parasitic inductance. Parallel bondwires are often used to reduce the total inductance, but since they conduct in-phase current, the effect of parallel bonding is undermined by the mutual inductance among the closely packed bondwires.

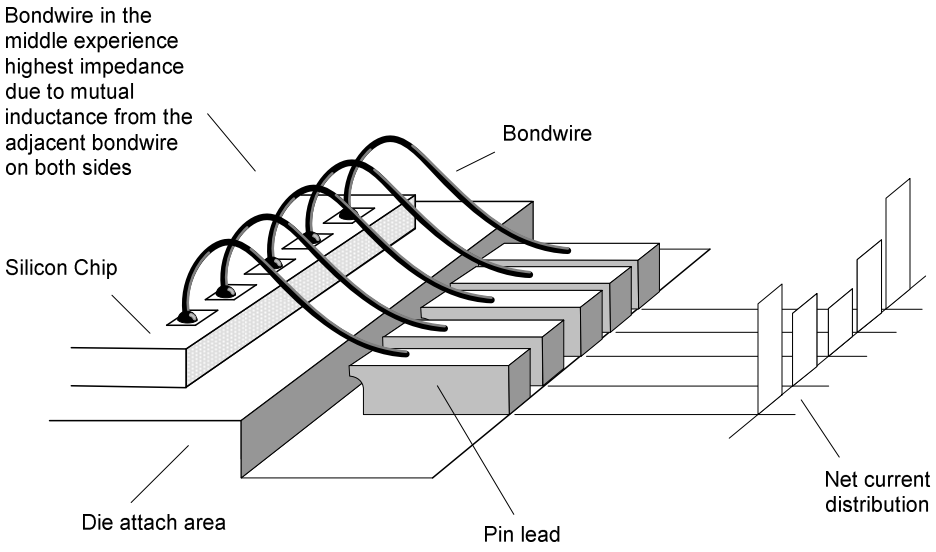


Figure 7-4 illustration of a wirebond package construction

Referring to Figure 7-6(a), in a common-source configured single-ended PA, the presence of parasitic elements on the ground path acts as source degeneration and degrades the device's gain and current driving capability. Assuming a square law model for the transistor, the drain current can be expressed as a function of the input voltage v_{IN} and the source voltage v_X:

$$i_D = K(v_{IN} - v_X - v_T)^2 \tag{7.9}$$

where v_T is the device threshold voltage, and v_X can be determined by

$$v_X = i_D R_{BW} + L\frac{di_D}{dt} \tag{7.10}$$

To gain a perspective on the severity of the ground parasitic effect, let us consider an ideal Class-E PA delivering 1-W of output power at 2GHz from a supply voltage of 2V. To simplify the analysis, we assume the classical form in which the parameter ε is infinite (i.e. a RF choke is used in place of the drain inductor). From Figure 6-7:

$$P_{OUT}\big|_{\varepsilon\to\infty} = \pi \cdot \frac{{V_{DD}}^2}{\omega_o C} \tag{7.11}$$

From (7.11) and (6.15), we can calculate the initial ramp rate of the switch current at the turn-on instant.

$$\frac{di_{SW}}{dt} = -\frac{di_{OUT}}{dt} = \frac{\pi}{2}\frac{P_{OUT}\omega_o}{V_{DD}} = 9.9A/n\sec \tag{7.12}$$

Therefore, even with a very optimistic assumption of L_{BW} = 0.1nH, voltage drop across the parasitic ground path will approach 1V, which will seriously reduce the device overdrive and impact the on-resistance and current driving capability.

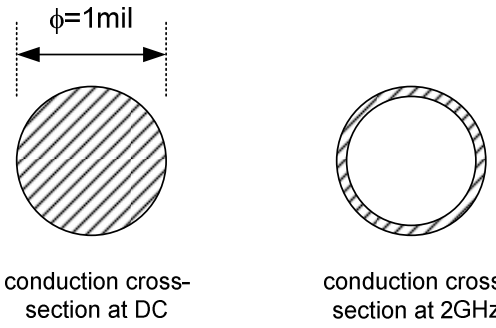


Figure 7-5 Dependency of bondwire conduction cross-section on frequency

Bondwires are typically made of either gold or aluminum with diameters in the order of 1mil. At low frequency, the high conductivity and large cross section area of the bondwire results in low resistive loss. However, if large amount of high frequency components are present in the bondwire current, such as in the ground connection of a single-ended PA, the resistive loss of the bondwire can increase significantly due to skin effect. For example, the DC resistance of a 1.5mm long, 1mil (25.4μm) diameter aluminum bondwire is about 80mΩ, but at 2GHz, the skin-depth will be reduced to 1.75μm (Figure 7-5). This results in a approximately 4X increase in the parasitic resistance (i.e., 320mΩ). In a high power, low-voltage switch-mode PA, switch on-resistance in the order of 100mΩ is sometime required, and the AC resistance the ground bondwire can therefore cause significant degradation in efficiency.

Alleviating Ground Connection Loss

In a differential PA where the common source nodes are couple on-chip as shown in Figure 7-6(b), all fundamental and odd order harmonics are confined within the chip boundary. As a result, only the DC and even order harmonics will appear on the ground current through the package parasitic. This effectively eliminates the degeneration effects of the ground parasitic (refer to the plot of i_{GND} of the SE-PA and the DM-PA in Figure 7-3(b)). In addition, since the AC current of the ground bondwire is reduced due to the removal of the fundamental component, power loss due to AC resistance of the bondwire is also alleviated.

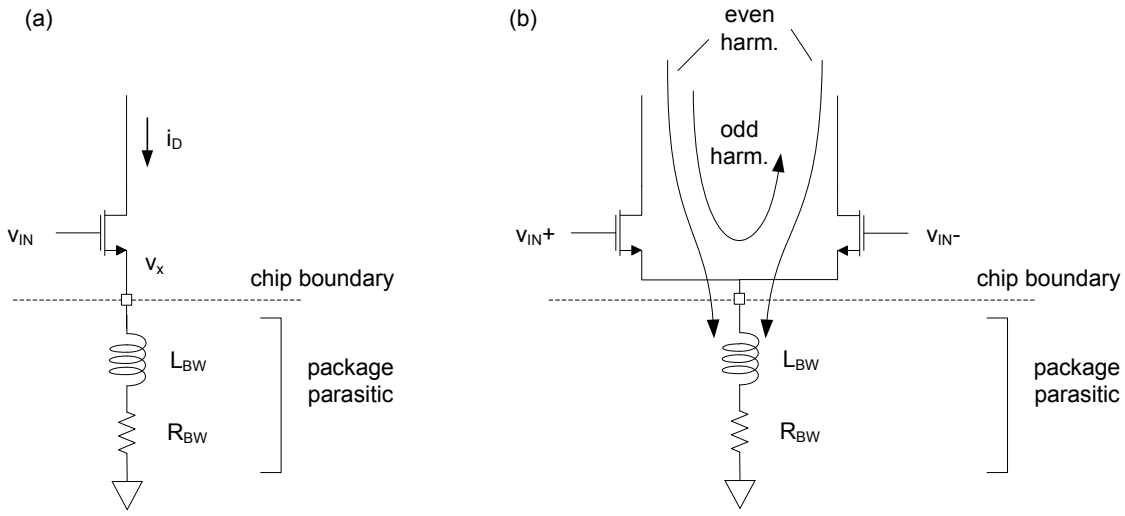


Figure 7-6 Comparison of ground current be a single-ended and a differential PA

Reducing the AC ground current going off-chip in a differential PA also reduces the voltage swing on the common source node. Since the common source is usually tied to the local substrate taps, this helps to reduce noise injection into the silicon substrate. Besides the reduced amplitude, the main frequency component of the injected noise is also shifted from the fundamental frequency in a single-ended PA to the second harmonic frequency in a differential PA. In a highly integrated transceiver, this can potentially help to avoid crosstalk between the power amplifier and other sensible components on chips.

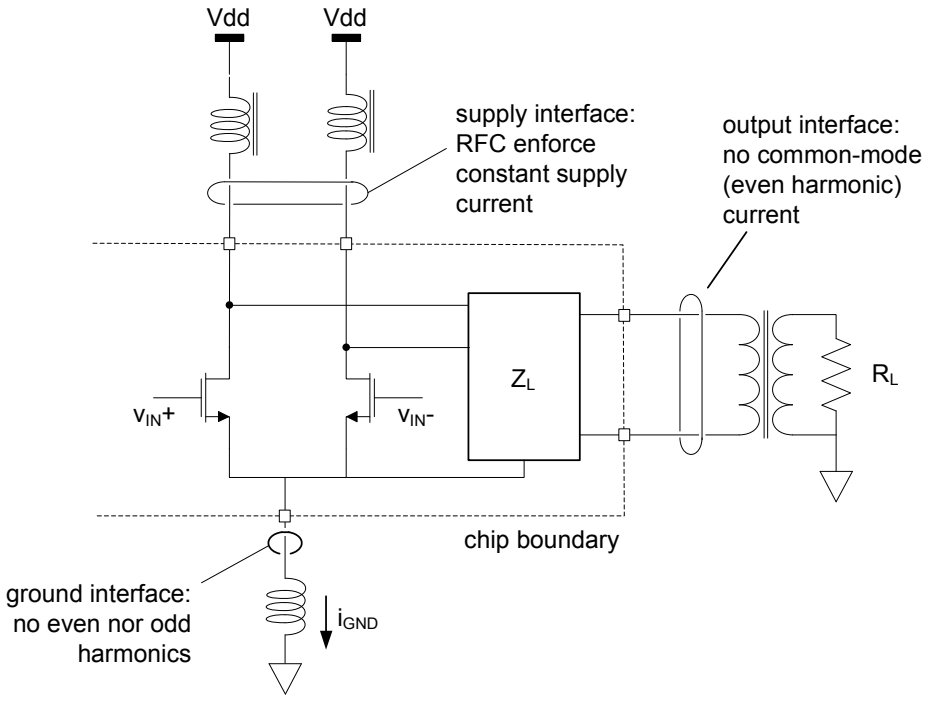


Figure 7-7 A differential PA with RFC in supply feed

In addition, in differential power amplifiers where the drain inductors are realized with RF chokes (RFC), the ground bounce will be further reduced. Consider the differential PA shown in Figure 7-7 where the supply current is fed from a pair of RFCs. In this case the supply current only contains DC components because of the large inductance of RFCs. The output current ideally only contains the fundamental component because of the band-pass filter within the output network. Since the sum of all current crossing the chip boundary must be zero at all harmonics due to KCL, we therefore conclude that the ground current must contain only DC component and thus generates no ground bounce.

Reducing Output Parasitic Inductance

In practical power amplifier design, it is often desirable to minimize the parasitic inductance of the output bondwires and package leads. This is because such parasitic inductance can limits the realizable impedance transformation ratio of the off-chip matching network and hence limits the deliverable output power. Although a series capacitor can in principle be used to tune out any excess inductance, this often leads to narrow-band tuning and impacts the overall bandwidth.

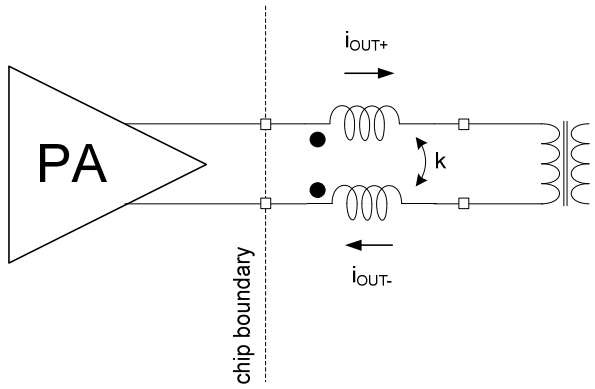


Figure 7-8 Effect of mutual inductance at the output of a differential PA

In a single-ended PA, the only effective way to minimize the output parasitic inductance is to reduce the physical length of the bondwires and to adopt more advanced (and costly) packaging technologies such as flip-chip. On the other hand, in a differential PA, effective output inductance can be reduced by exploiting the mutual inductance between adjacent bondwires. As illustrated in Figure 7-8, if the differential output bondwires are place next to each other, the anti-phase current will reduce the effective output inductance by a factor of (1-*k*), where *k* is the coupling factor between the bondwires.

7.5 Injection-locking Technique

Compared with a single-ended configuration, a differential topology reduces the power amplifier's driving requirement by half. This is often a significant factor because PA

devices usually have very large peripheries due to the amount of current they need to handle. The input capacitor involved is often so large that it cannot be effectively tuned out with on-chip inductors; and large amount of current has to be spent at the input just to drive the PA devices. For PAs that needs to handle high levels of output power such as those in cellular handsets, even a simple differential configuration alone may not adequately reduce the input drive requirement. To further reduce the input driving requirement and enhance the overall efficiency, we therefore propose an injection locking topology that is particularly well-suited for amplification of constant envelope modulated signals.

Injection locking refers to the condition in which an otherwise self-oscillating circuit is forced to run at the same frequency as an input signal, resulting in a substantial reduction in the input driving requirement. This is realized in each stage of the amplifier by a pair of cross-coupled assisting devices, as shown in Figure 7-9. The two input voltages are out of phase, as are the two output voltages. The load impedance at the output nodes is designed such that *Vo2* and *Vin1* run in phase to control the composite switch. As far as each half circuit is concerned, the operation is similar to a simple single-ended PA. There are however, differences in two aspects: First, the current originally circulating at each tuned load is now utilized to assist switching of the opposing half circuit. Second, the capacitance at each input can now be significantly reduced without increasing the overall composite switch on-resistance. Note that a sizable capacitor at the drain of each device

is inherently needed to realize the desirable Class-E operation. By replacing these passive drain capacitors with the gate capacitors of the assisting devices, we have reused the AC current circulating the matching and load networks to simultaneously reduce the device on-resistance and the input driving requirements.

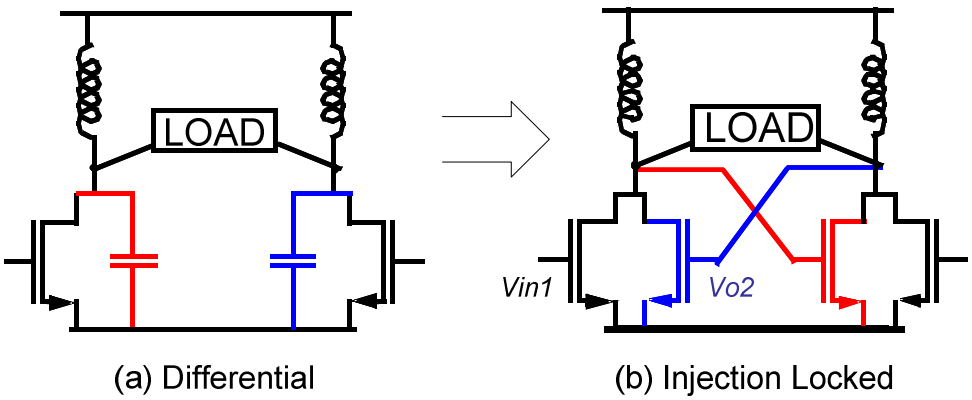


Figure 7-9 Illustration of the injection locking concept

By introducing the assisting devices, we have added a positive feedback loop within the differential power amplifier and effectively turned the PA into an oscillator. In fact, if left un-driven, the injection-locking amplifier will self-oscillates at its natural oscillating frequency. To avoid self-oscillation, we need to guarantee two conditions through out the operation.

Firstly, we need to ensure that when the PA is powered up, an input will always be available and is strong enough to synchronous the PA such that the output faithfully follows the instantaneous frequency and phase variation of the input. Secondly, we also need to provide a power-down mode that eliminates self-oscillation and DC current consumption when the amplifier is not driven.

Injection Locking Range

Because of its tuned nature, an injection-locked amplifier can operate only within a certain frequency range around its naturally oscillating frequency. Synchronization of oscillators by the influence of an external signal has been an object of extensive studies. One classical analysis on such phenomena was proposed by Alder [Adle46], and we will explore the injection-locking behavior of the Class-E PA based on his approach. We may partition the core of the injection-locked Class-E PA into three portions as shown in Figure 7-10: 1) a network with impedance ($H(j\omega)$) formed by the matching and the load networks; 2) a negative gm stage (-gm) formed by the cross-coupled assisting devices, and 3) a current i_j that is injected into the PA for synchronization purpose.

During operation, a differential voltage drives the input device pair (see Figure 7-9 (b)) to produce current i_j. This current is combined with the current i_g of the –gm stage into the load current i_l which is injected into the load $H(j\omega)$ to produce the drain voltage v. The drain voltage v in turns becomes the input of the –gm stage. If the input successfully

locks the PA into a synchronous operation, the amplitude and phase of i_l, i_g and v will eventually reach a constant value and follows that of i_j.

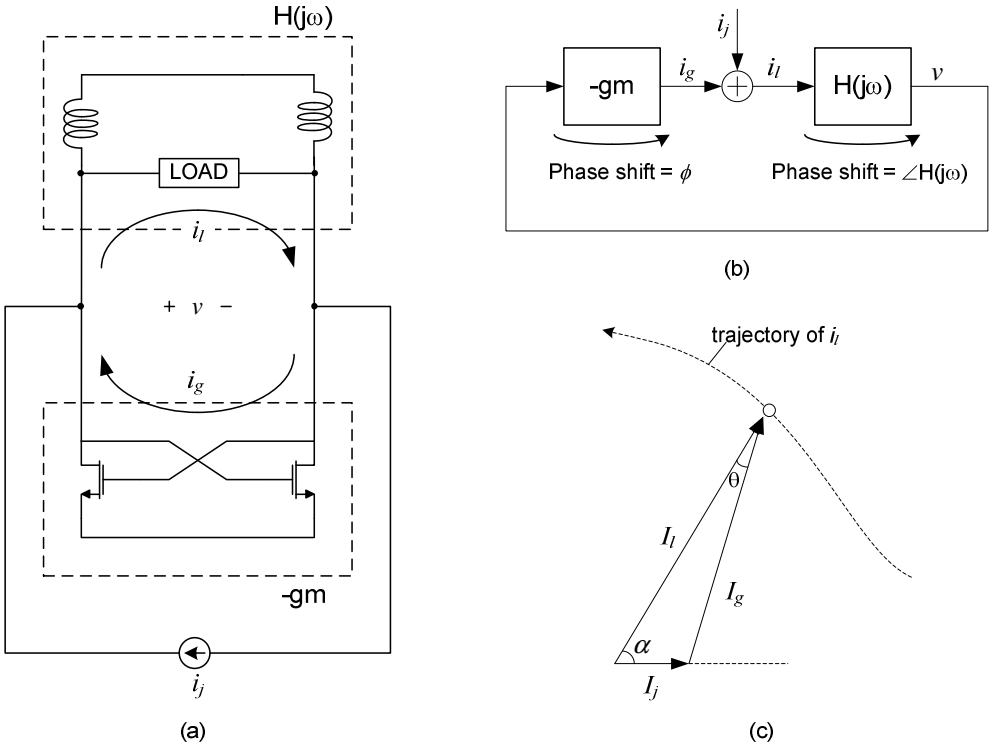


Figure 7-10 (a) basic injection-locked Class-E PA structure (b) signal flow diagram, and (c) vector trajectory of the load current

Figure 7-10(b) shows the signal-flow diagram of the injection-locked Class-E PA. First let us consider the case where no external current is injected in to the PA core (i.e. i_j=0).

In this case, the PA core would self oscillate at a natural oscillating frequency ω_o if the loop gain formed by –gm and $H(j\omega_o)$ is greater then unity. For the oscillation to sustain, the total phase shift around the loop has to be zero, i.e.

$$\phi + \angle H(j\omega_o) = 0 \tag{7.13}$$

where we have assumed ϕ, the phase shift between the output current and input voltage of the –gm block, is constant within our frequency of interest.

Let us now inject a periodic current i_j into the PA core at a frequency ω_I that is close to the natural oscillating frequency ω_o. Figure 7-10(c) shows the resulting phase relationship among i_g, i_l and i_j. Here the phase plane is referenced to ω_I so that vector i_j appears stationary whereas i_g and i_l may revolve around the origin as a function of time. If injection locking is successful, the whole circuit will eventually settle into a periodic operation at frequency ω_I, and all three vectors will appear static. If locking failed, i_g and i_l will keep revolving on the phase and never settle into static values.

It is our primary concern to understand the conditions that determines the frequency range within which injection locking can be guaranteed. As shown in Figure 7-10(c), injection of i_j into the PA core will cause a phase shift θ between i_g and i_l. Base on simple trigonometric identity, we identify that

$$\frac{\sin\theta}{I_j} = \frac{\sin\alpha}{I_g} \tag{7.14}$$

where α is the phase angle of the load current i_l with respect to the injection current i_j. If we assume that the injected current is small compared with the nominal output current of the –gm block, we can approximate

$$\theta \approx \sin\theta = \frac{I_j}{I_g}\sin\alpha \tag{7.15}$$

Regardless of where locking is successful; the total phase shift around the loop will once again become zero when the circuit enters steady state. Therefore,

$$\theta + \phi + \angle H(j(\omega_1 + \frac{d\alpha}{dt})) = 0 \tag{7.16}$$

where $\omega_l + d\alpha/dt$ is the instantaneous frequency of the load current i_l. Substituting (7.13) and (7.15) into (7.16), we have

$$\angle H(j(\omega_1 + \frac{d\alpha}{dt})) - \angle H(j\omega_o) = -\frac{I_j}{I_g}\sin\alpha \tag{7.17}$$

If ω_l is close to ω_o, LHS of (7.17) can be further simplified to

$$\angle H(j(\omega_1 + \frac{d\alpha}{dt})) - \angle H(j\omega_o) \approx \frac{d}{d\omega}\angle H(j\omega)\Big|_{\omega_o} (\omega_1 + \frac{d\alpha}{dt} - \omega_o) \tag{7.18}$$

where $d\angle H(j\omega)/d\omega$ as illustrated in Figure 7-11 is the rate of change of the output impedance phase and is it is proportional to the quality factor Q of the output network.

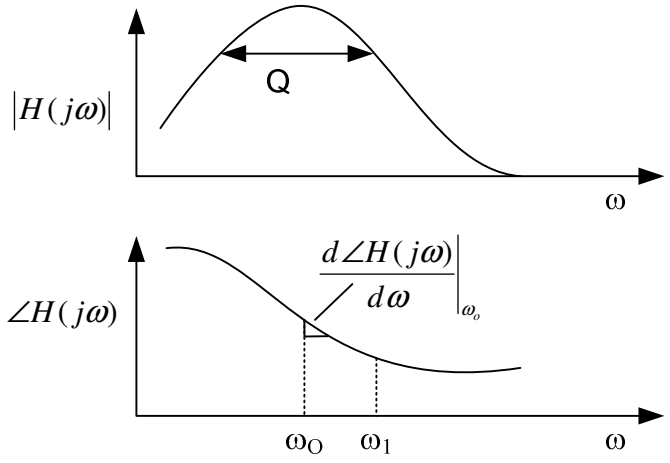


Figure 7-11 Output network impedance vs. frequency

Equation (7.17) can now be expressed as

$$\frac{d\alpha}{dt} \approx -\frac{I_j}{I_g}\sin\alpha\left(\frac{d}{d\omega}\angle H(j\omega)\Big|_{\omega_o}\right)^{-1} + (\omega_o - \omega_1) \tag{7.19}$$

Equation (7.19) describes the dynamics of the of the injection locked PA core. If the circuit is locked and synchronized with the injected current, then $d\alpha/dt$ will eventually settle to zero, and

$$\begin{aligned} |\omega_1 - \omega_o| &= \frac{I_j}{I_g}\left|\sin\alpha\left(\frac{d}{d\omega}\angle H(j\omega)\Big|_{\omega_o}\right)^{-1}\right| \\ &\le \frac{I_j}{I_g}\left(\left|\frac{d}{d\omega}\angle H(j\omega)\Big|_{\omega_o}\right|\right)^{-1} \end{aligned} \tag{7.20}$$

Equation (7.20) defines the maximum possible difference between the injection frequency ω_1 and the natural oscillation frequency ω_o if injection locking is to be established. It therefore defines the locking range of the injection locking PA. Equation (7.20) suggests that injection locking can occur only within a finite frequency range around the natural oscillation frequency of the circuit. The extent of the locking range is directly proportional to the amplitude of the injected current I_j and inversely proportional to Q of the output network at the natural oscillation frequency. For the injection locking PA to operate properly, we thus need to guarantee that the locking range is wide enough to cover the entire frequency band of interest.

Chapter 8 - A Compact Differential to Single-ended Converter

8.1 Introduction

Utilizing a differential topology in integrated CMOS PAs alleviates the problem of low breakdown voltage. It also mitigates the issues of substrate noise and package interface. Yet RF components (e.g. filters, switches, and antenna) that are driven by power amplifiers are often available only in single-ended forms. To take full advantage of a differential PA, we thus have to ensure that the differential PA output signal can be readily converted into a single-ended form with minimal loss and cost. A balun (stands for BALanced-Unbalanced) is a network that performs such conversion.

Since the differential-to-single-ended conversion occurs at the PA output, the balun's insertion loss is critical to the overall power efficiency. Unfortunately, low loss baluns often come with large physical sizes – conventional, distributed, baluns achieve low cost and low loss, but their dimensions are typically in the order of the signal wavelength, which results in unacceptably large footprints at low GHz frequencies.

In this chapter, we focus on the solution to reduce balun form-factor without sacrificing insertion loss and cost. We will begin in section 8.2 with illustrations of conventional balun structures. It is followed by the proposal and the analysis of a compact balun design in sections 8.3 8.4 . The proposed structure incorporates both distributed and lumped elements in achieving a miniaturized size at low GHz frequencies. Finally, practical layout, simulation and measurement considerations will be addressed in section 8.5 .

8.2 Conventional Baluns

Conventional microwave baluns are often purely distributive circuits that are built entirely with planar elements such as transmission lines, open and short stubs. These baluns exhibit low insertion loss and are relatively compact at frequencies above 10GHz.

One example of purely distributive baluns is the ring hybrid (a.k.a. rat-race hybrid) [Poza93] as illustrated in Figure 8-1. It is formed by a circular pattern of transmission line whose total length is 1.5 times the signal wavelength (λ) at the desired operating frequency. The characteristic impedance of the transmission line is $\sqrt{2}$ times the port

reference impedance Z_o (typically 50Ω). The entire ring hybrid balun can be implemented with microstrip line laid out on a printed circuit board (PCB.) Three ports are attached to specific locations around the ring: port 1 is single-ended, while ports 2 and 3 form the differential ports.

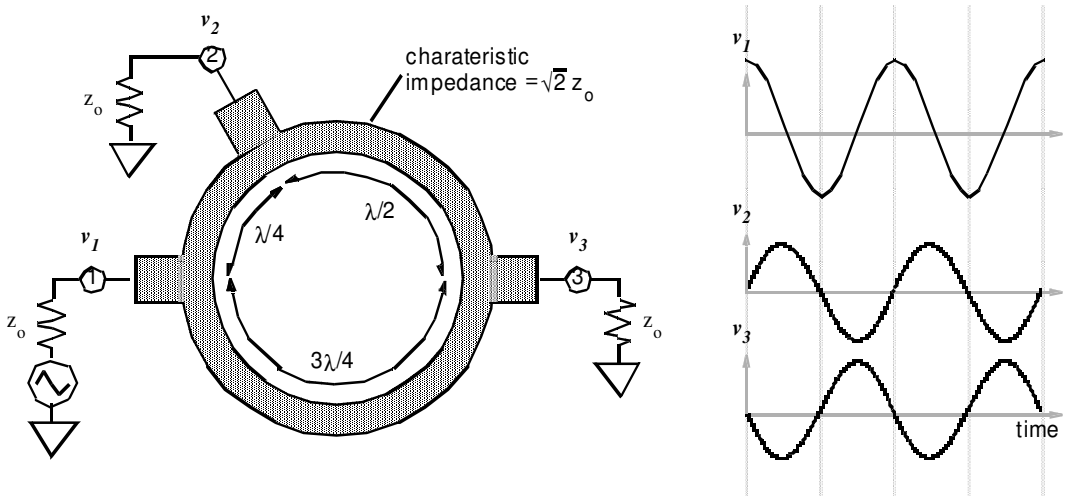


Figure 8-1 The ring hybrid balun

Operation of the ring hybrid balun can be visualized by considering a signal that is launched into port 1 at the operating frequency. This incident signal is split into a clockwise traveling wave towards port 2, and a counter-clockwise traveling wave towards port 3. Since port 2 is located at a quarter wavelength down from port 1, the signal arrives at port 2 is delayed by 90^o relative to port 1. Likewise, the signal at port 3 is delayed by

270 ^{o} relative to port 1. As a result, the differential port signals are 180 ^{o} out of phase, and this relationship is reinforced by the half wavelength (corresponding to 180 ^{o} phase shift) section between the differential ports.

With all ports matched to *Zo*, incident power from the single-ended signal will be evenly delivered into the differential ports. Since this network is reciprocal (as in all the baluns discussed hereafter, which involve only passive elements and no ferrite material), the conversion is also reversible. In other word, when the differential ports are driven by two balanced signals, the differential signal power will be combined at the single-ended port. The ability to convert between single-ended and differential signals constitutes the desired function of a balun.

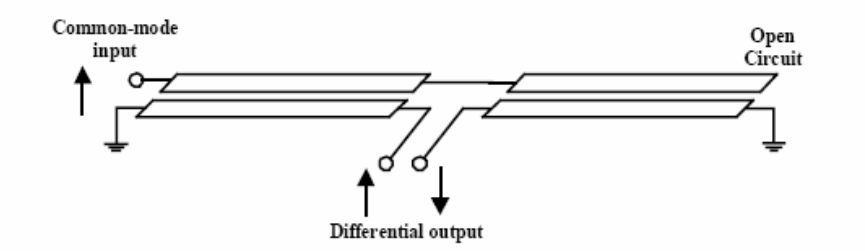


Figure 8-2 A Marchand balun realized with microstrip lines

Another example of a classical balun structure is the Marchand balun invented by Nathan Marchand in 1944 [Marc44]. The original implementation was built with coaxial

transmission lines, and Figure 8-2 depicts a simplified planer version realizable with coupled microstrip lines [Delv02].

The Operation of Marchand balun is based on quarter wavelength coupled microstrip lines (CML). Referring to Figure 8-2, an incident singled-end signal at the common-mode input port is divided with a 90^{o} phase difference at the through and coupled ports of the $\lambda/4$ CML on the left hand side. The through signal is reflected at the open termination on the right hand side CML, while the coupled signal is reflected at the short termination on the left. The two signals eventually combine to form a pair of differential signals at the output ports.

The Marchand balun requires two pairs of CML with total length of one half wavelength. Although 'folded' variations were proposed by Nguyen [Nguy93] in attempt to reduce the balun footprint, but the overall size at low GHz frequencies (in the order of cm^2 for typical PCB material) is still much too large for most portable handheld devices.

A common characteristic among distributive baluns is their wide bandwidth — typically a good fraction of the carrier frequency. However, such bandwidth is often not necessary in typical wireless applications. For example, none of the prevailing cellular, cordless or wireless LAN standards in the 1GHz or 2.4GHz ISM requires bandwidth more than a few percents of the carrier frequency. In fact, narrower bandwidth baluns are often desirable because it helps to attenuate out-of-band spurious emission.

Our focus in the remaining parts of this chapter is to develop an alternative approach to reduce the bandwidth of the traditional \baluns in exchange for a smaller physical size. The result is a compact balun that combines distributive transmission lines with discrete components, and is well suited for operation at low GHz frequencies.

8.3 A Compact Hybrid Balun

In 1996 Kumar et al. [Kuma96] proposed a lumped-distributed balun structure shown in Figure 8-3. This balun is formed by a pair of short coupled microstrip lines on a printed circuit board. In contrary to the baluns previously described, the length of these coupled lines is only a few percent of the wavelength, which results in a very compact footprint even at low GHz frequencies.

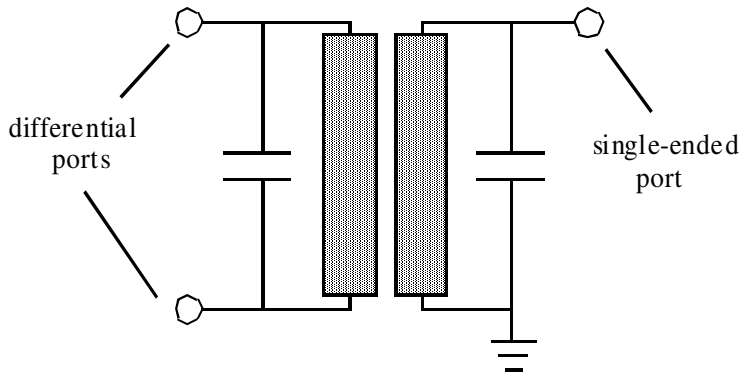


Figure 8-3 Illustration of the Kumar balun [Kuma96]

The short coupled lines does introduce significant impedance mismatch at the I/O ports, and this mismatch is corrected by two discrete tuning capacitors. If the added cost of the capacitors is acceptable, then such approach presents a viable solution in achieving compact balun form factor. The original Kumar balun however suffered from a relative high insertion loss, which is a major limitation if the balun were to be used at the output of a power amplifier. This has motivated us to work on a more rigorous analysis on the fundamental working principle of the balun and to exploit ways to improve its insertion loss.

8.4 Theoretical Background

A good understanding on the operation of the compact hybrid balun can be developed from the basic characteristic of coupled microstrip lines (CMLs), and one convenient way to analyze CMLs is to make use of the admittance matrix. In this section, we will first define the admittance matrix of a 2-port network formed by a single (uncoupled) transmission line, and then extend the result to handle mutually coupling transmission lines.

For a given transmission line dimension and operation frequency, our analysis shows that the resulting admittance matrix of the CMLs corresponds to an equivalent lumped circuit network formed by purely reactive elements. At the core of the equivalent circuit is an L-C structure that is capable of splitting an incoming signal into out-of-phase output

signals, and thus performs the basic function of a balun. This balun characteristic is however, normally hampered by an excess of inherent inductance at the I/O ports. In order to exploit the desired balun characteristics, discrete capacitors are thus added to tune out the excess inductance. The derivation of the balun characteristics suggested that certain characteristics in the original Kumar realization introduce non-idealities and may negatively impact the balun's performance. We thus propose a new structure to minimize these effects at the end of the chapter.

Admittance Matrix of a Single Microstrip Line

Figure 8-4 shows a 2-port network formed by a single transmission line. The admittance matrix of this network is defined by:

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{8.1}$$

where i_k is the current flowing into, and v_k is the voltage at port k. Due to symmetry of the structure, the resulting network is reciprocal. In other word, $y_{11} = y_{22}$, and $y_{21} = y_{21}$. This leaves only two independent admittance parameters, y_{11} and y_{21}, to be determined.

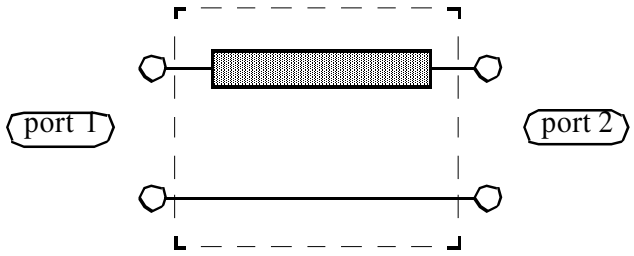


Figure 8-4 a 2-port network form by a single microstrip line

y_{11} and y_{21} can be calculated by terminating port 2 by a short and excite port 1 as shown in Figure 8-5. Here, e_{ik} and e_{rk} are phasers of the incident (traveling into a port) and reflected (traveling out of a port) voltage waves at port k.

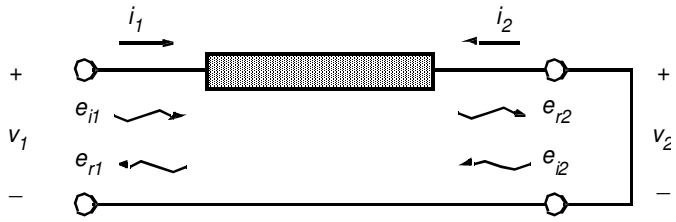


Figure 8-5 incident and reflected waves in a shorted microstrip line

Since signals propagate down the transmission line at finite speed, when the incident wave at port 1 arrives at port 2, it is phase-shifted, or delayed by an amount that equals to

$$\theta = \frac{2\pi L}{\lambda} \tag{8.2}$$

where L is the physical length of the transmission line and λ is the signal wavelength at the operation frequency. θ is also known as the electrical length of the transmission line. The short termination at port 2 presents a reflection coefficient of -1. Therefore,

$$e_{i2} = -e_{r2} = -e_{i1} \cdot e^{-j\theta} \tag{8.3}$$

when e_{i2} returns back to port 1, it is delayed by another θ, therefore

$$e_{r1} = e_{i2} \cdot e^{-j\theta} = -e_{i1} \cdot e^{-j2\theta} \tag{8.4}$$

The terminal voltages and currents at the two ports can now be calculated as

$$\begin{aligned} v_1 &= e_{i1} + e_{r1} = e_{i1}(1 - e^{-j2\theta}) \\ v_2 &= e_{i2} + e_{r2} = 0 \\ i_1 &= \frac{e_{i1} - e_{r1}}{z_o} = \frac{e_{i1}}{z_o}(1 + e^{-j2\theta}) \\ i_2 &= \frac{e_{i2} - e_{r2}}{z_o} = \frac{-2 \cdot e_{i1}}{z_o} \cdot e^{-j\theta} \end{aligned} \tag{8.5}$$

where Z_o is the characteristic impedance of the transmission line. We also define $y_o = 1/Z_o$ be the characteristic admittance. Substituting $v_2=0$ into (8.1) and evaluate the admittance parameters in terms of the terminal voltage and current results in

$$y_{11} = \frac{i_1}{v_1} = -jy_o \cot\theta$$
$$y_{21} = \frac{i_2}{v_1} = jy_o \csc\theta \tag{8.6}$$

In summary, the 2-port network that corresponds to a single transmission line can be fully characterized in terms of its admittance matrix, whose entries depend only on the characteristic admittance y_o and electrical length θ of the line.

Admittance Matrix of Coupled Microstrip Lines

If two microstrip lines are placed in close proximity such that they electrically and magnetically couple with each other as shown in Figure 8-6, a total of four basic parameters are required to completely characterize the system. These basic parameters are the odd (even) mode characteristic admittance, y_{oo} (y_{oe}), and the odd (even) mode electrical length, θ_o (θ_e). The four parameters correspond to the two distinct modes of propagate along the transmission line pair. The characteristic admittance, y_{oo} and y_{oe}, depend on the cross sectional dimension of the microstrip lines and permittivity of the substrate material, while the electrical length, θ_o and θ_e, depend on the physical length of lines relative to the signal wavelength. Detailed formulation of these parameters can be found in [Fook90], but it is helpful for us to first understand the relationship between y_{oo} and y_{oe}, as well as between θ_o and θ_e.

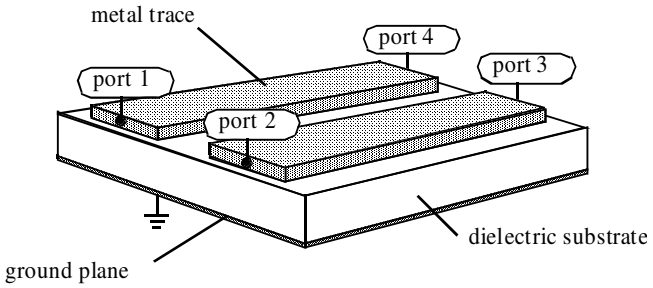


Figure 8-6 construction of coupled microstrip lines

Figure 8-7 shows that when the microstrip lines are driven in odd mode, they experience higher capacitance due to the additional fringing field in between the two metal strips. This results in a higher odd mode characteristic admittance than in even mode

$$y_{oo} \geq y_{oe} \tag{8.7}$$

The equality condition in (8.7) corresponds to infinite separation between the strips (i.e. they are uncoupled) in which case y_{oo} and y_{oe} would become identical and simply equal to the characteristic admittance of a stand-alone microstrip.

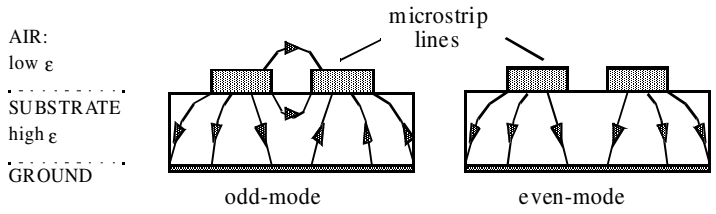


Figure 8-7 Field patterns of odd and even mode propagation

The microstrip line we consider is said to be in a heterogeneous dielectric medium because the conducting metal strips are located between the substrate on the bottom and air on top, and the two mediums have different dielectric constants. As shown in Figure 8-7, when driven in odd mode, a large portion of the field goes through air which has a lower dielectric constant. The effective dielectric constant in odd mode is thus lower than that in even mode. Since electrical length is related to the effective dielectric constant by

$$\theta = \frac{\varpi L}{c}\sqrt{\varepsilon_{eff}} \tag{8.8}$$

where ω is the signal frequency in radian/s, L is the physical line length and c is the speed of light in vacuum. This implies that the odd mode electrical length of the CMLs is lower than in even mode:

$$\theta_o < \theta_e \tag{8.9}$$

Equation (8.9) applies only to transmission lines with heterogeneous dielectric medium. If the transmission lines were enclosed within dielectric of uniform permitivity (e.g. stripline), then there would be not difference between θ_o and θ_e.

Next we will derive the 4-port admittance matrix of the CMLs in terms of the four basic parameters: y_{oo}, y_{oe}, θ_o, and θ_e. The admittance matrix defines the relationship among the terminal current and voltage as follow:

$$\begin{bmatrix} i_1 \\ i_2 \\ i_3 \\ i_4 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & y_{13} & y_{14} \\ y_{21} & y_{22} & y_{23} & y_{24} \\ y_{31} & y_{32} & y_{33} & y_{34} \\ y_{41} & y_{42} & y_{43} & y_{44} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} \tag{8.10}$$

where i_k is the current flowing into port k of the network shown in Figure 8-6; and v_k is the voltage measured at the same port. Base on the symmetry and reciprocity of the network, we can reduce the matrix into four independent entries:

$$\begin{aligned} y_{11} &= y_{22} = y_{33} = y_{44} \\ y_{21} &= y_{12} = y_{43} = y_{34} \\ y_{31} &= y_{13} = y_{42} = y_{24} \\ y_{41} &= y_{14} = y_{32} = y_{23} \end{aligned} \tag{8.11}$$

This leaves only four unknown parameters, y_{11}, y_{21}, y_{31} and y_{41}, within the matrix. Based on the definition in (8.10), these four unknowns can be calculated by driving port 1 with an input signal while leaving all other ports grounded ($v2=v3=v4=0$) as shown in

configuration (a) of Figure 8-8. In this case, the admittance parameter $y_{k1}=i_k/v_1$ is simply the ratio between the terminal current at port k and the voltage at port 1.

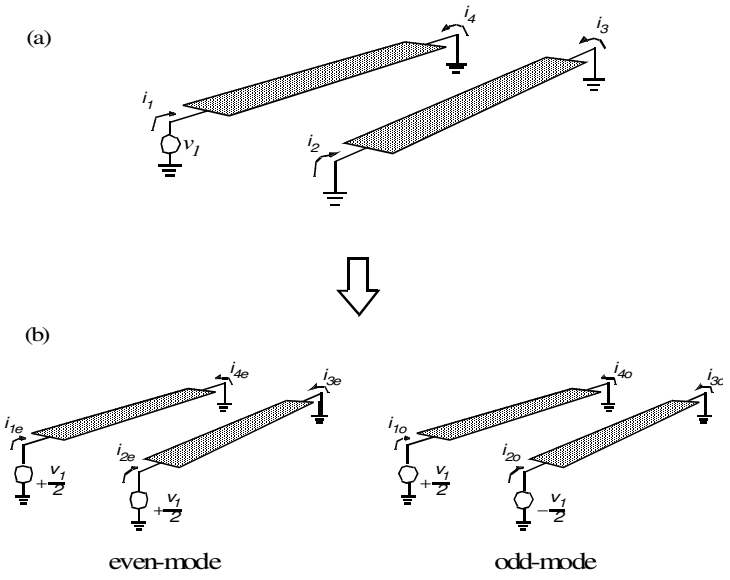


Figure 8-8 Decomposition of single-port excitation into even and odd mode

Since the network is linear, we can decompose the single-port excitation configuration in Figure 8-8(a) into a superposition of even and odd mode excitation as shown in Figure 8-8(b). In each mode, the pair of coupled microstrip lines is equivalent to a single uncoupled line with characteristic admittance and phase length of the corresponding mode. In other word, the relationship between the terminal current and voltage derived in

(8.6) still applies after replacing y_o and θ by the characteristic admittance and phase length of the corresponding mode. In even mode, (8.6) becomes

$$\begin{aligned} i_{1e} &= i_{2e} = -j \cdot y_{oe} \cot\theta_e \cdot \frac{v_1}{2} \\ i_{3e} &= i_{4e} = \quad j \cdot y_{oe} \csc\theta_e \cdot \frac{v_1}{2} \end{aligned} \tag{8.12}$$

while in odd mode it becomes:

$$\begin{aligned} i_{1o} &= -i_{2o} = -j \cdot y_{oo} \cot\theta_o \cdot \frac{v_1}{2} \\ i_{3o} &= -i_{4o} = \quad j \cdot y_{oo} \csc\theta_o \cdot \frac{v_1}{2} \end{aligned} \tag{8.13}$$

Note that the sign change at the first equal sign of (8.13) is due to the differential drive in odd mode. With (8.12) and (8.13), the total terminal current at each port can be obtained by combining the corresponding even and odd mode current; and the corresponding admittance matrix parameters can be derived as:

$$\begin{aligned} y_{11} &= \frac{i_1}{v_1} = \frac{i_{1o} + i_{1e}}{v_1} = -\frac{j}{2} \cdot (y_{oo} \cdot \cot\theta_o + y_{oe} \cdot \cot\theta_e) \\ y_{21} &= \frac{i_2}{v_1} = \frac{i_{2o} + i_{2e}}{v_1} = \quad \frac{j}{2} \cdot (y_{oo} \cdot \cot\theta_o - y_{oe} \cdot \cot\theta_e) \\ y_{31} &= \frac{i_3}{v_1} = \frac{i_{3o} + i_{3e}}{v_1} = -\frac{j}{2} \cdot (y_{oo} \cdot \csc\theta_o - y_{oe} \cdot \csc\theta_e) \\ y_{41} &= \frac{i_4}{v_1} = \frac{i_{4o} + i_{4e}}{v_1} = \quad \frac{j}{2} \cdot (y_{oo} \cdot \csc\theta_o + y_{oe} \cdot \csc\theta_e) \end{aligned} \tag{8.14}$$

Note from (8.14) that all admittance parameters are purely imaginary, which is consistent with the fact that the coupled microstrip structure is ideally lossless and hence all the admittance parameters should be purely reactive. Each admittance parameter is either capacitive (if imaginary part is positive), or inductive (if imaginary part is negative). Based on (8.7) and (8.14), if we further assume that the coupled microstrip lines are sufficiently short such that both even and odd mode electrical length are less than $\pi/2$ (i.e. quarter wavelength), it can be shown that two of the admittance parameters (y_{11} and y_{31}) are inductive, while the other two (y_{21} and y_{41}) are capacitive:

$$\begin{array}{l} y_{oo} \geq y_{oe} \\ \theta_o < \theta_e < \pi/2 \end{array} \quad \Rightarrow \quad \begin{cases} \mathrm{Im}(y_{11}) < 0 \\ \mathrm{Im}(y_{21}) > 0 \\ \mathrm{Im}(y_{31}) < 0 \\ \mathrm{Im}(y_{41}) > 0 \end{cases} \tag{8.15}$$

A Lumped Circuit Model of Coupled Microstrip Lines

At any given operating frequency where $\theta_o < \theta_e < \pi/2$, the admittance matrix can be represented by an equivalent lumped reactive network as shown in Figure 8-9, where each component represents an inductor or a capacitor in accordance with (8.15).

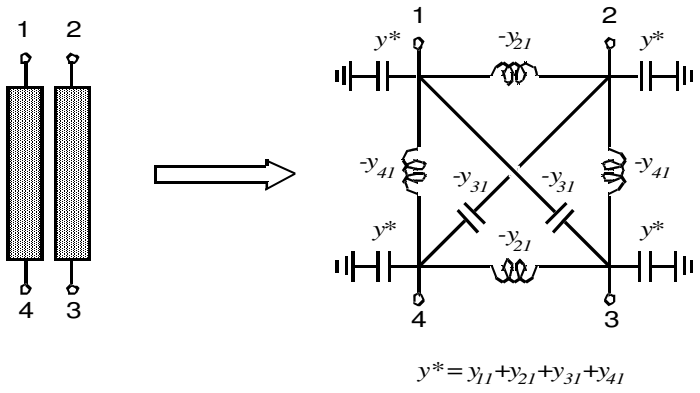


Figure 8-9 Equivalent lumped L-C network of a pair of sub-quarter length CMLs

At first sight the presence of inductance $-y_{21}$ between ports 1 and 2 and between ports 3 and 4 may seem unnatural as physically there is no DC path between the two pair of ports. The reason for this paradox is because we have evaluated the equivalent L-C network at the operation frequency, and all component values are functions of frequency as well (through the dependency on θo and θe). In particular, one can verify that if ports 1 and 4 are shorted together to form one node, and likewise ports 2 and 3 are shorted to form a second node, then the conductance between the two nodes is

$$-2(y_{21}+y_{31}) = j[y_{oo}(\csc\theta_o - \cot\theta_o) - y_{oe}(\csc\theta_e - \cot\theta_e)] \tag{8.16}$$

and it does converge to zero as frequency approaches zero (i.e. both θo and $\theta e \rightarrow 0$), and is therefore consistent with the physical structure.

Constructing the Compact Hybrid Balun

The coupled microstrip structure in Figure 8-9 forms the basic building block of the compact hybrid balun. To convert it into the actual balun, we ground port 4 and insert a tuning capacitor with conductance y_{41} across ports 2 and 3 to tune out the inherent inductance of the coupled microstrips. The resulting structure is shown inside the dashed box in Figure 8-10.

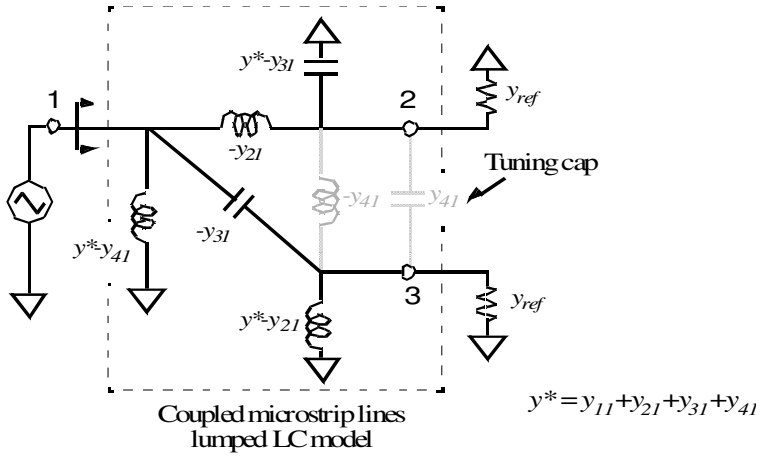


Figure 8-10 Equivalent lumped L-C model of the hybrid balun

The L-C section between port 1 and port 2, together with the C-L section between port 1 and port 3, form the core of the balun. They allows an incident signal at port 1 to be split into the other two ports with oppose phrase shift.

To verify that the structure is indeed capable of converting a single-ended input signal at port 1 into differential output signals at ports 2 and 3, we terminate ports 2 and 3 by a resistive load with admittance y_{ref} (typically 1/50Ω), and drive port 1 with an ideal voltage source as shown in Figure 8-10. It can be shown that the terminal voltage at ports 2 and 3 are

$$v_2 = \frac{-y_{21}}{y_{ref} + y_{11} + y_{41}} \cdot v_1$$
$$v_3 = \frac{-y_{31}}{y_{ref} + y_{11} + y_{41}} \cdot v_1 \tag{8.17}$$

respectively, where v_1 is the terminal voltage at port 1. The ratio between the two output voltages is

$$\frac{v_2}{v_1} = \frac{y_{21}}{y_{31}} = -\frac{y_{oo} \cdot \cot\theta_o - y_{oe} \cdot \cot\theta_e}{y_{oo} \cdot \csc\theta_o - y_{oe} \cdot \csc\theta_e} \tag{8.18}$$

Clearly the two output signals are exactly 180^o output of phase and hence form a differential pair. There is a slight amplitude imbalance since $|v_2 / v_3|$ is always less than one. The amplitude mismatch however diminishes as θ_o and θ_e approach zero, or equivalently when the physical length of the transmission lines becomes much smaller than the signal wavelength. This is a desirable characteristic of such balun where a more compact physical size helps to improve the output amplitude imbalance.

Since the realization of the balun function relies on the parallel resonance between the tuning cap and the inherent inductance between port 2 and port 3, perfect resonance occurs only at one frequency and the hybrid balun is inherently narrow band. Such characteristics can be advantageous in narrow-band applications where the balun can serve to attenuate high frequency harmonics from the PA output.

To ensure maximal transfer of power, the input port of the balun needs to be matched to a desirable admittance level. It can be shown that the input admittance at port 1 in Figure 8-10 is

$$y_{in} = \alpha \cdot y_{ref} + [y_{11} - \alpha \cdot (y_{11} + y_{23})] \tag{8.19}$$

where the first term is purely real and the term within the bracket is purely imaginary. α is the impedance transformation factor which can be expressed as

$$\alpha = \frac{y_{12}^2 + y_{13}^2}{(y_{11} + y_{23})^2 - y^2{}_{ref}} \tag{8.20}$$

The impedance transformation factor is a function of the geometric dimension of the microstrip lines, dielectric constant of the substrate, as well as the reference impedance level. One typical case is to match all three ports of the balun to the same reference admittance. In this case, $\alpha = 1$ and the input admittance at port 1 becomes

$$y_{11} = y_{ref} - y_{23} \tag{8.21}$$

The imaginary part ($-y_{23}$) is inductive and can be tuned out by shunting port 1 with a second tuning capacitor (in this particular case with the same capacitance as the first tuning cap across ports 2 and 3), consequently matching the input port to the same reference admittance y_{ref}. If the source admittance driving port 1 is also y_{ref}, then the incident power at port 1 will be absorbed with no reflection and be split into ports 2 and 3 with almost equal amplitude and 180^{o} out of phase, thus realizing the basic function of a balun.

Thus far we have considered utilizing the balun to convert from single-end to differential signals. If a differential signal is applied between port 2 and port 3, reciprocity of the network ensures that the total incident power will be combined and delivered to the single-ended output port 1. This mode of operation in particular is need at the PA output.

Basic network theory states that it is impossible for a reciprocal, lossless 3-port network to be unconditionally (i.e. regardless of incident phase and amplitude) matched at all three ports. In the differential-to-single-end configuration, impedance matching at port 2 and port 3 is achieved only under the condition that these ports are driven differentially. In fact, the amplitudes of the input signal should be deliberately skewed according to (8.18) if perfect matching were to be achieved. In practice, however the discrepancy is generally negligible if the electrical length of the coupled line is sufficiently small.

8.5 Practical Realization

Previous section described the theoretical background of the hybrid balun. In essence, the conversion between single-ended and differential signals is made possible by the inherent L-C and C-L connections among the three ports as shown in Figure 8-10. Discrete capacitors are added across port 2 and port 3, as well as between port 1 and ground to resonate out unwanted inductive components and to ensure good impedance matching. The first physical realization of the hybrid balun was reported by Ojha in 1996 [Ojha96]. This original realization is shown in Figure 8-11(b). In order to accommodate the physical size of discrete capacitors, four uncoupled extensions are added to the terminals of the coupled microstrips. This deviation from the ideal model, however, tends to increase the overall insertion loss of the balun.

To alleviate this effect, we thus propose an asymmetrical and semicircular structure as shown in Figure 8-11(c). This proposed structure accommodates the physical dimension of the discrete capacitors with minimum deviation from the ideal model. The tapered transitions between the coupled microstrips and the 50Ω impedance matched port extensions also helps to reduce unwanted reflections. The relatively complex geometry of this structure however makes it difficult to be analyzed in closed form. Detailed electromagnetic simulations are essential in optimizing the final structure. Figure 8-11 shows a simulation model used in a 3D electromagnetic simulator known as IE3D. This

model takes into consideration the metal trace thickness, physical dimensions of the discrete capacitors, and parasitic inductance of the ground vias etc.

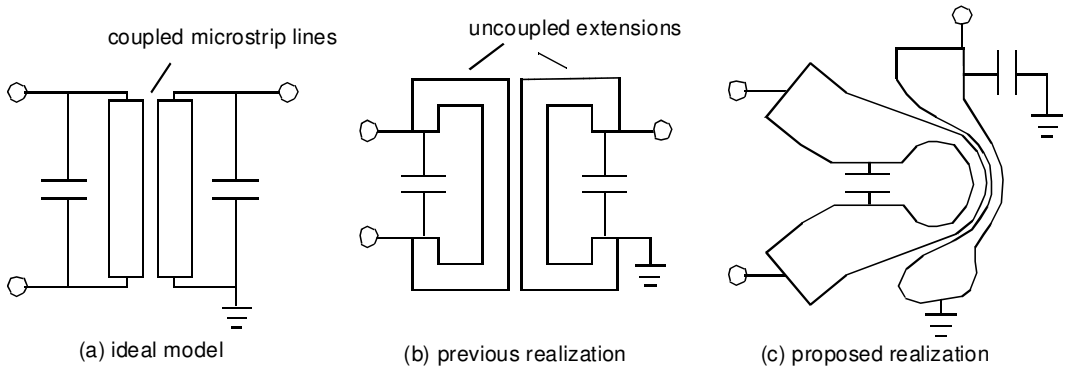


Figure 8-11 Hybrid balun realizations: ideal model (a), original (b) and proposed (C) realization

The EM simulations provide a 7-port (3 ports for the balun, plus 4 ports for the terminals of the discrete capacitors) s-parameter model for the distributed structure. The overall balun response as a function of frequency can then be obtained by co-simulating the s-parameter model with the discrete capacitors in a linear circuit simulator such as Spice.

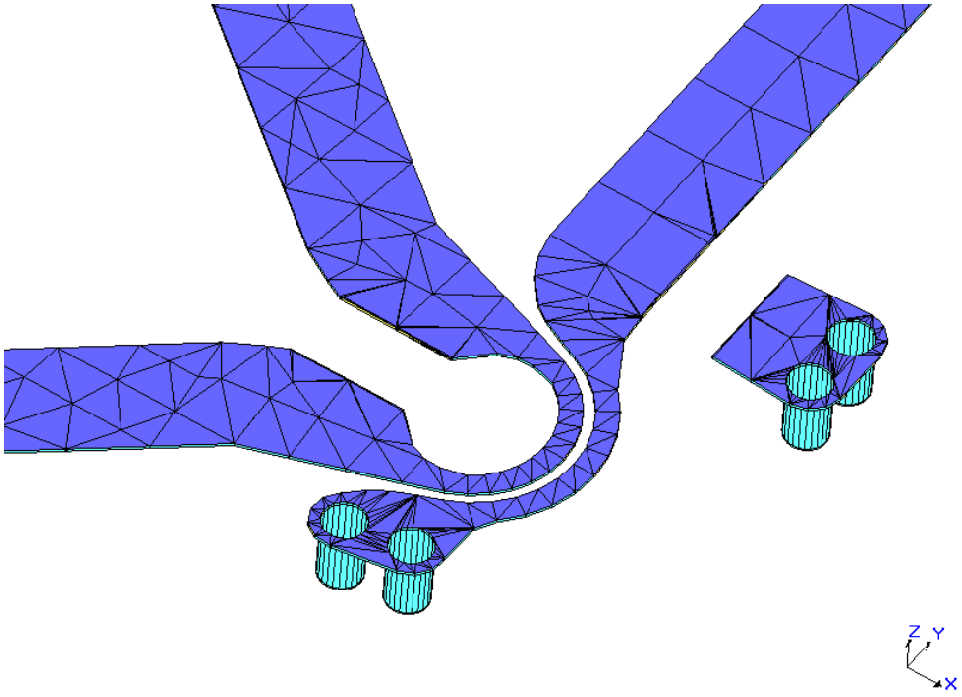


Figure 8-12 3-D EM simulation model of the proposed hybrid balun structure

In this chapter, we have discussed the underlying theory and design considerations of a novel compact balun. We also built a prototype of the balun for evaluating its performance at our frequency of interest. Details on the balun prototype as well as the measurement results will be presented in Chapter 9 .

Chapter 9 - Experimental Prototype and Measured Results

9.1 CMOS Class-E PA Prototype

A prototype of the 1-W 1.9 GHz CMOS power amplifier was fabricated in a 0.35μm, single-poly, five-layer metal CMOS process. The chip area is 1.0mm X 0.6 mm including bonding pads. The prototype device and the test board were assembled using a chip-on-board packaging technique in which the bare die was attached directly to the printed circuit board as shown in Figure 9-1.

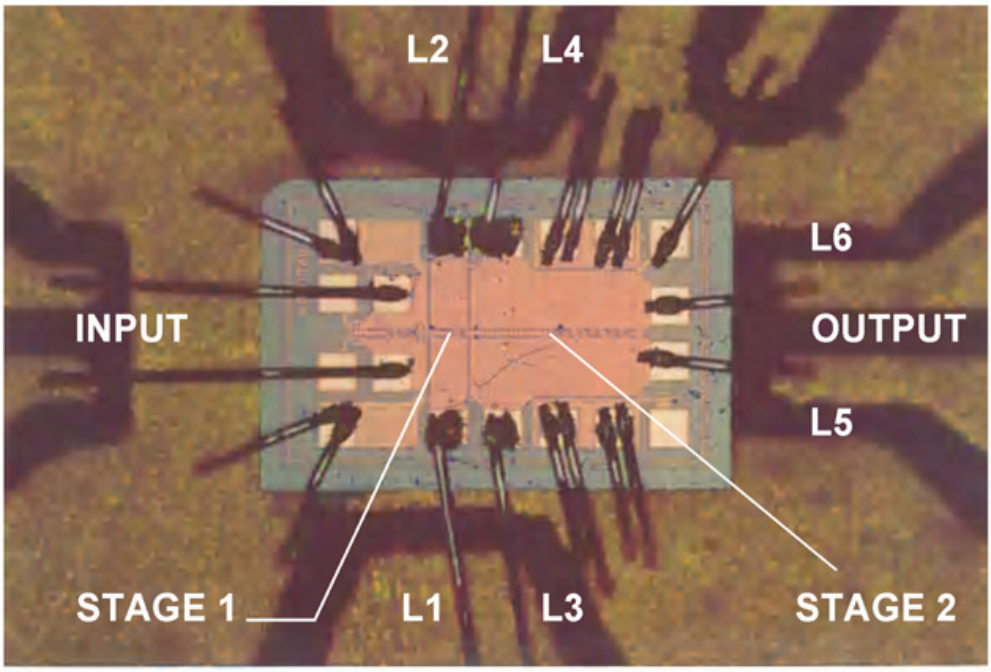


Figure 9-1 Chip-on-board Assembly

The die was manually thinned down to approximately 200µm before being attached to the ground plane using a conductive composite. This was done to reduce the minimum achievable bondwire length and to facilitate heat dissipation. All wire bonding where done manually with aluminum bondwires of 1.25mil diameter. Ground pads were double bonded to reduce parasitic inductance. Various lengths and orientations of the drain (L1-L4) and output inductors were experimented in order to obtain the optimal configuration. Using a dummy chip and on-chip probing, we were able to assess the bondwire inductance by direct s-parameter measurements. These measurements were performed on

pairs of adjacent bondwires to include the effect of mutual inductance and to minimize ground-loop related errors.

Evaluation Test Board

Figure 9-2 shows the test board used for evaluating the CMOS PA prototype. The board is constructed with a standard 4-layer FR4 printed circuit board (PCB). On the left hand side is the supply regulator that regulates an external supply voltage to an internal supply. The internal supply voltage is adjustable through the potentiometer on board. RF input is fed from a SMA connector at the bottom and then converted into a differential signal using a surface mount discrete balun. After that, the differential input is brought on-chip through the input bondwires. The input signal is amplified by the CMOS PA which is mounted directly on the PCB. The differential output signal of the PA is brought off-chip through the output bondwires.

Two ways of measuring the output signals are possible. The first one is to convert the differential output signal back to single-ended by using a surface-mounted discrete balun (same type as used at the input). However, we found that optimally tuning the output matching network with this configuration is very difficult. The sound option is to pick up the differential output signal from the test board using a custom made differential probe. This method allows us to have greater freedom in varying the output matching

termination. It also enables us to interface the PA module with the proposed compact balun.

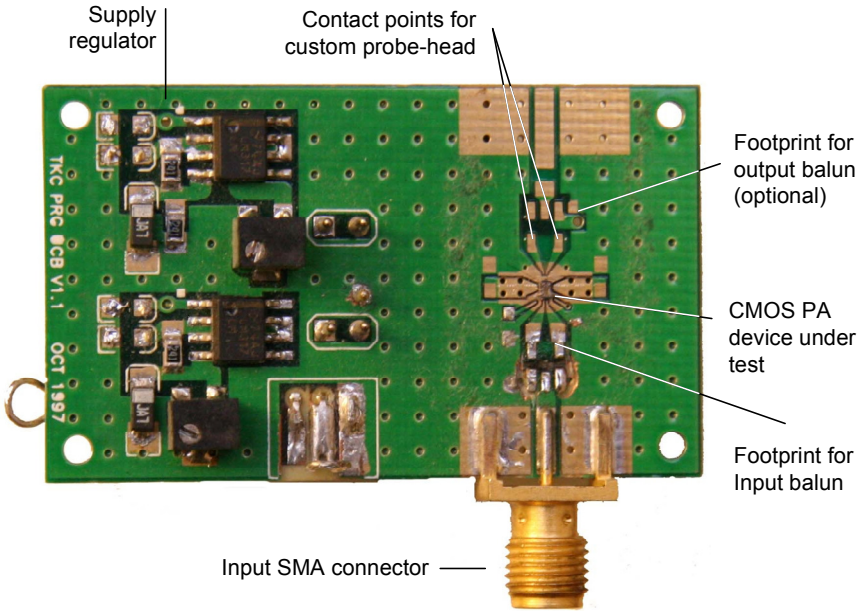


Figure 9-2 Evaluation Test Board

Output Tuning

Due to the parasitic elements presence within the chip, the package and the test fixture, the measured matching network frequency response can deviate significantly from the nominal design values. This is particularly true at the relatively high frequency of operation. In order to optimize the output matching of the power amplifier, we have

improvised a flexible tuning setup as shown in Figure 9-3. Firstly, a custom differential probe is created by modifying two off-the-shelf SMA connectors. A discrete capacitor is soldered directly across the differential input pins of the probe-head. The probe-head is mounted to a manipulator on a probe-station; this allows us to have the one-dimensional freedom of sliding the custom probe-head along the differential output trace on the PCB. With this set up, we can effectively tune the series inductance (by sliding the probe-head along the output trace) and the shunt capacitance (by replacing the discrete capacitor) of the output matching network until optimal performance is achieved.

To obtain the measurement data presented in this chapter, we connect the differential output of the probe-head to an external balun which converts the differential signal into single-ended for equipment interface. In addition to the proposed compact balun, we also took measurement with an off-the-shelf wideband balun and used it as reference.

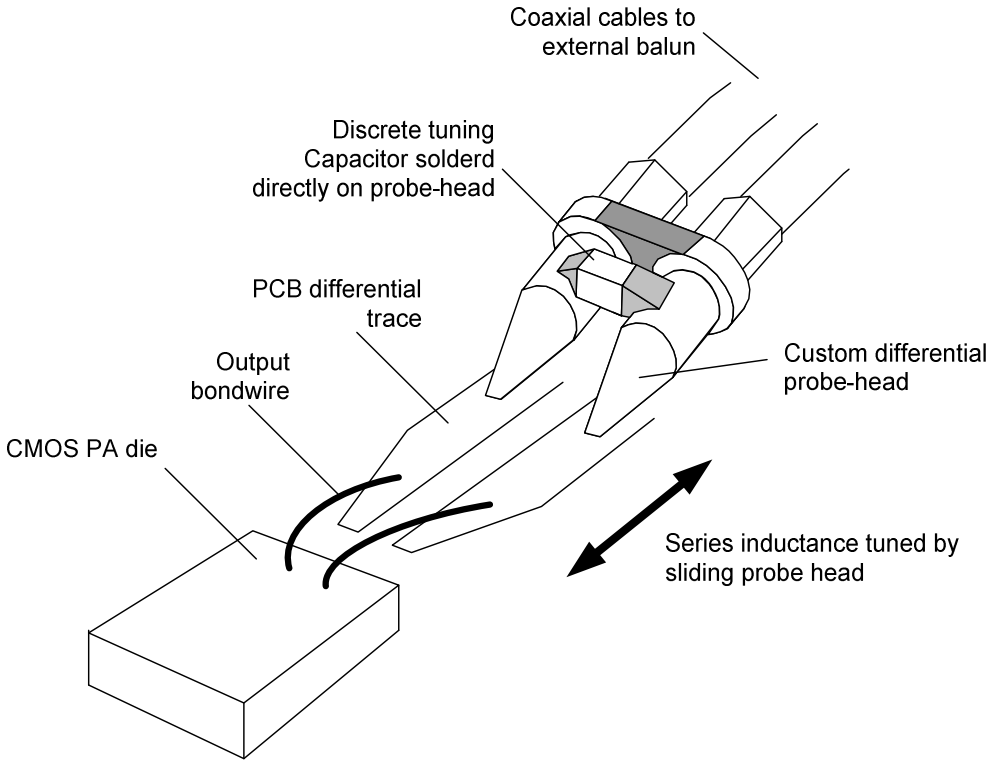


Figure 9-3 Output Tuning Set up

9.2 CMOS PA Measurement results

Measured Output Power, Efficiency and Locking Range

A 10-dBm single-ended input was converted to a differential signal using a commercial surface-mount balun (Murata LDB20C500A1900), and then applied to the PA. RF power was measured at the 50- output ports. Figure 9-4 shows a plot of the output power versus the supply voltage, measured at 1.98GHz. The output power increases from 49mW to

1.0W monotonically as the supply is swept from 0.6 to 2 V and it is approximately proportional to V_{DD}^2. The maximum supply voltage used in this case is determined by the peak drain voltage on chip. The peak drain voltage is estimated to be slightly above 5V at 2V supply, approaching the drain-gate oxide breakdown limit of the technology.

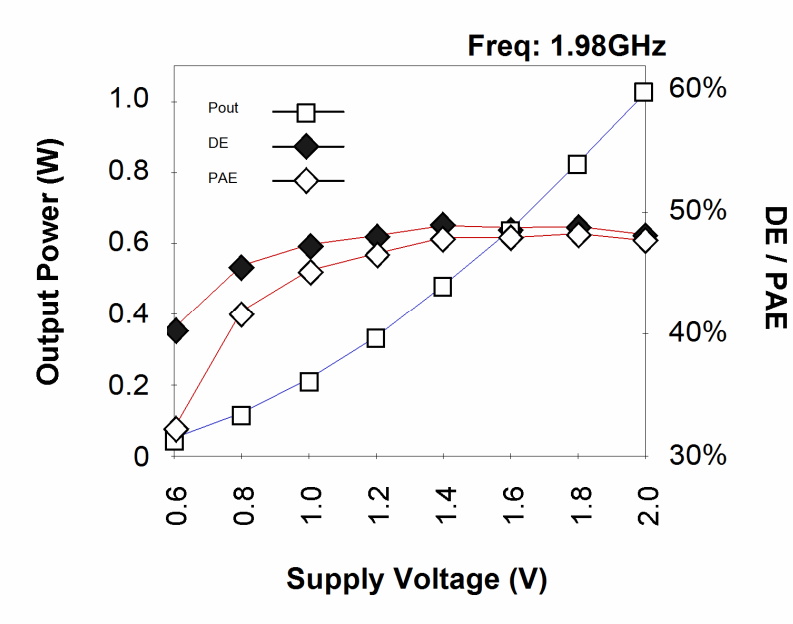


Figure 9-4 Output Power, Drain efficiency, and Power-added efficiency of the CMOS PA prototype

Also shown in Figure 9-4 are the drain efficiency (DE) and the power-added efficiency (PAE), where

$$DE = \frac{\text{output power}}{\text{supply power}}$$
$$PAE = \frac{\text{output power - input power}}{\text{supply power}} \tag{9.1}$$

As expected form the switching nature of the PA, the drain efficiency remains close to the optimal value for most of the output power range; this is except for a slight reduction in the low supply voltage region where the switching action of the transistors becomes less effective. The relatively sharp decline of the PAE in this region is due to the progressively more prominent input power term in the PAE definition. At 2V supply, the PAE was measured to the 48%.

Figure 9-5 shows the output power and PAE at two different supply voltages across the range of frequencies where the amplifier is mode locked successfully. This locking range is measured to be 490 MHz, centering at about 1.9 GHz.

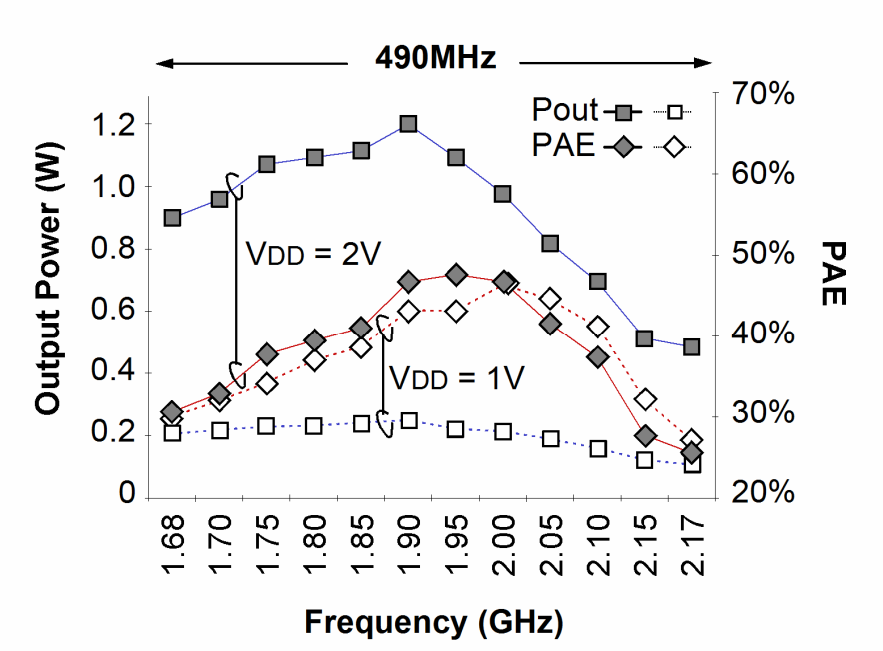


Figure 9-5 Output power and PAE versus Frequency at VDD = 1V and 2V

Sensitivity to Wire Bonding Variation

Since the critical inductors in the PA were realized with bondwires and were potentially sensitive to variation, we would like to understand the dependency of the PA performance on the bonding configuration. In Figure 9-6, we summarize the measured power-added efficiency of a number of test samples with the drain inductor bondwire length close to the optimized value. These data show that the PAE variation is approximately within 4% of the optimized value if the drain inductor bondwires are

within +/- 10% of the nominal length (corresponding to approximately +/-10% variation in the inductance values). From Lee's published data in [Lee98], we expect the machine bonding process to provide inductance accuracy of about 5%. We thus expect the PAE to be fairly robust against bondwire variation in a production environment.

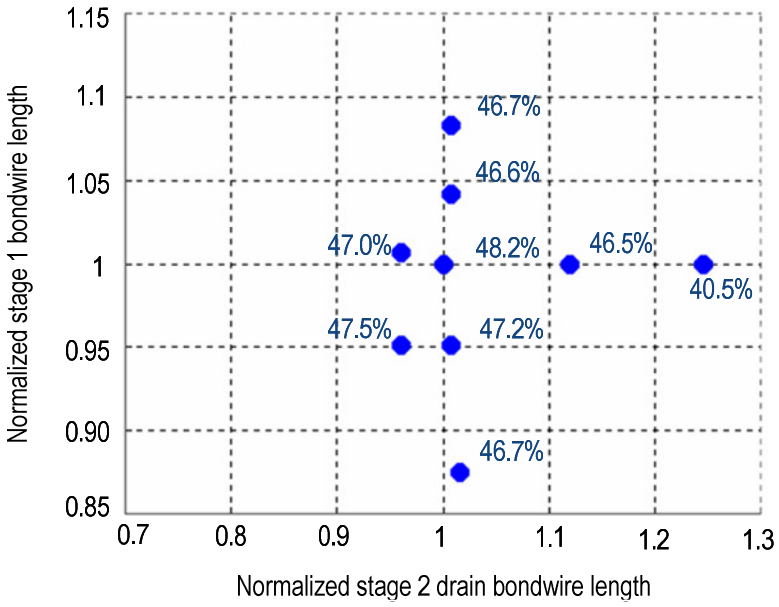


Figure 9-6 PAE sensitivity to wire bonding variation

Testing with GSM Modulated Signal

To confirm its potential in practical communication applications, the injection locked Class-E PA was tested with a Gaussian Minimum Shift Keying (GMSK) [Muro81] modulated input signal. GMSK is a constant envelope modulation scheme in which information is carried in the phase variation of the signal, making them well suited to switching power amplification. This modulation scheme is widely used world-wide in the GSM family of cellular systems (GSM900, DCS1800 and PCS1900). As an example, a random bit sequence was modulated with the modulation parameter corresponding to PCS1900, and then was applied to the PA. Figure 9-7 shows the close-in output spectrum. No observable distortion was found in the signal, which remained confined in the GSM spectral emission specification over the full range of output power.

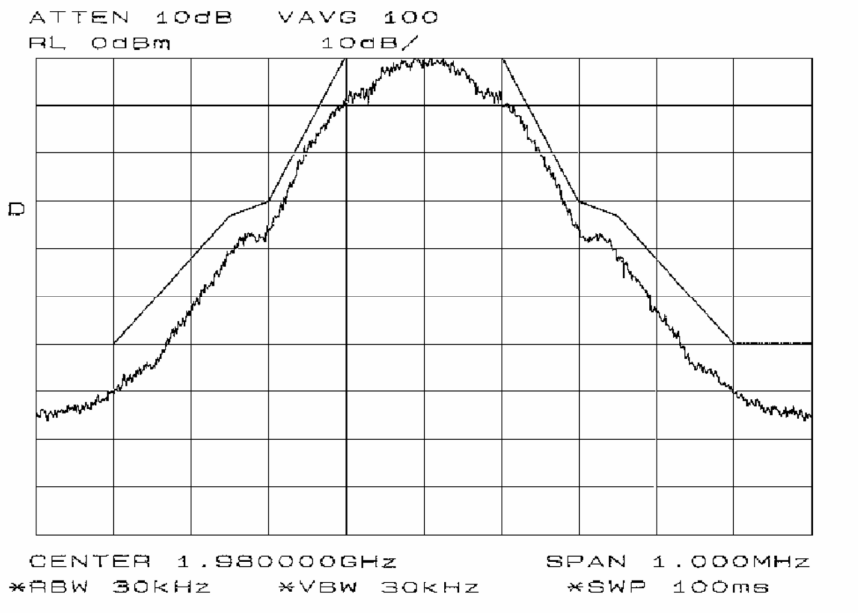


Figure 9-7 Amplified GMSK modulated Signal (BT=0.3) vs. the GSM spectral emission mask

Noise Emission

Besides the close-in modulation spectrum, wide-band noise performance is another key aspect in modern digital communication applications. This is particularly true in high performance systems such as GSM, in which the spurious emissions of the transmitted (up-link) signals that fall in the receiving (down-link) band are tightly specified.

Direct measurement of the PA wide-band noise is difficult because of the limited dynamic range of the spectrum analyzer. In order to measure the noise emission into the

down-link frequency band, we need to high-pass filter the PA output so as to block the strong PA output signal while allowing the noise emission into the down-link frequency band to pass through and be measured at the spectrum analyzer. The measurement set-up is show in Figure 9-8.

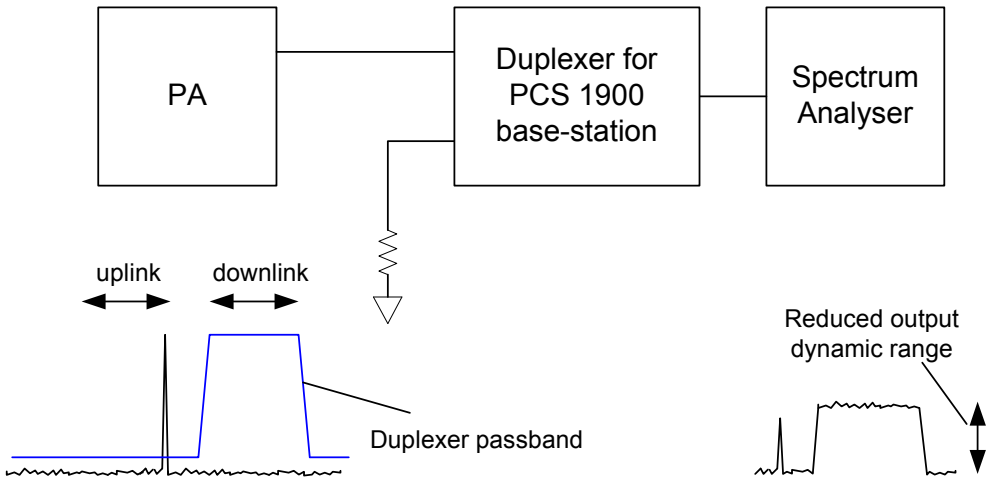


Figure 9-8 Noise emission measurement set-up

A PCS1900 (a version of GSM used in North America, whose frequency band falls in the measurement range of the PA) duplexer was inserted in between the PA and the spectrum analyzer. In this experiment, we have chosen a duplexer designed for PCS1900 base station because it offers very high isolation between the up-link and down-link frequency bands. An amplified GMSK modulated signal centered at the upper limit of the PCS1900

up-link band (1.91GHz) is transmitted from the PA and is passed through the duplexer. The duplexer attenuates the signal tone but allowed the PA noise to pass at a 20MHz offset over a 60MHz bandwidth. As shown in Figure 9-9, the total noise power within this bandwidth is measured to be -52.5dBm (5.6nW), which translates to a noise power density of -130.3dBm/Hz.

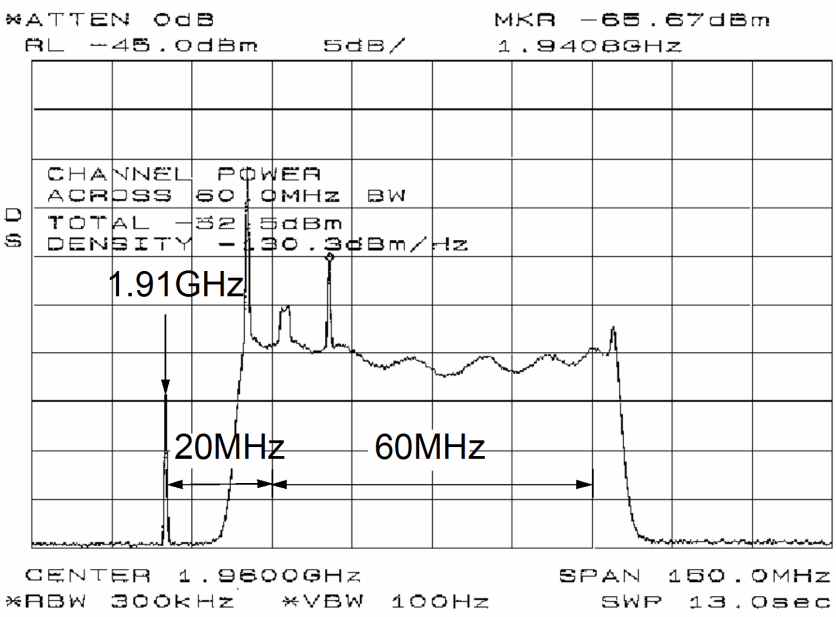


Figure 9-9 PA noise emission spectrum at duplexer output

9.3 Compact Balun prototype

A compact hybrid balun was designed and fabricated in a standard 2-layer printed circuit board process. The PCB substrate is standard FR4 with dielectric constant of 4.6 and thickness of 31mil. The design rule allows a minimum of 10mil metal trace width with 5mil spacing. 0.5oz. and 1oz. copper options are commonly available in PCB processes. The heavier option corresponds to thicker metal layer (1oz copper corresponds approximately to a thickness of 1.4mil) 1oz copper is used in our prototype to minimize conduction loss in the metal trace and to enhance side-wall coupling in the coupled microstrip section. In addition, metal conductivity is further increased by gold plating the copper metal layers. Figure 9-10 illustrates the PCB profile used in the balun prototype.

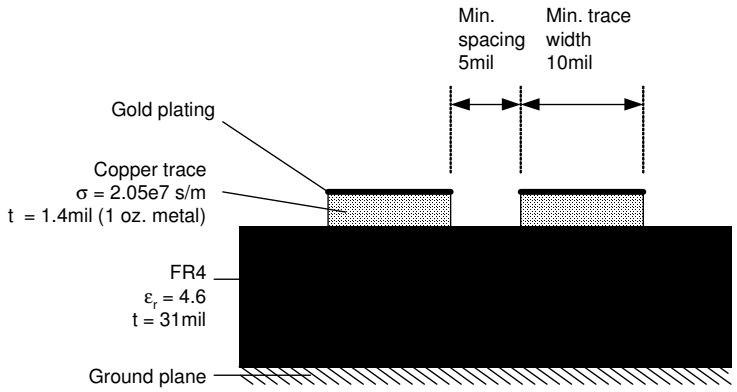


Figure 9-10 PCB profile of the hybrid balun prototype

The final hybrid balun prototype is shown in Figure 9-11. Port 1 is the single-ended port and ports 2 and 3 are the differential ports. The minimum metal trace spacing allowable by the design rule (5mil) is used at the coupled microstrip section to maximize mutual coupling. Two ceramic chip capacitors are used for tuning at the single-ended and differential ports. The diameter of the approximately semi-circular coupled traces is about 3mm, limited chiefly by the physical dimension of the discrete capacitors. The overall dimension of the core portion including the coupled traces and discrete capacitors is approximately 5mm x 6mm. The single-end and differential signals are brought out to their prospective edge mounted SMA connectors by 50Ω microstrips. The overall layout is curved and smooth to avoid unnecessary discontinuity.

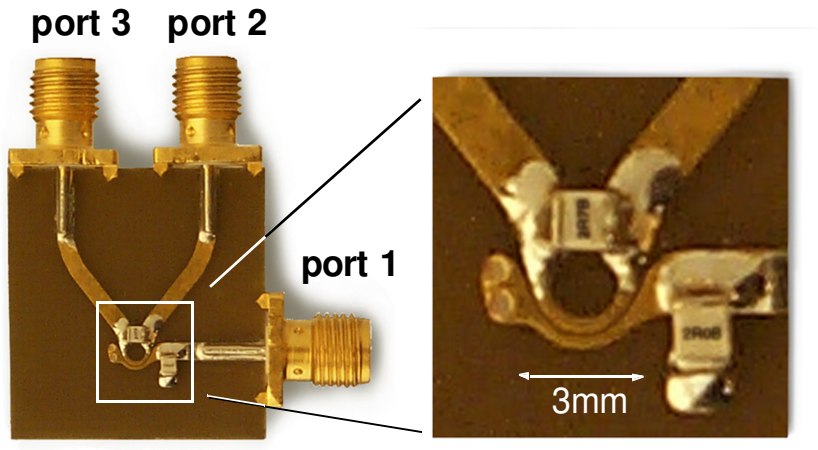


Figure 9-11 Experimental prototype of the compact hybrid balun

9.4 Balun Prototype Measurement Results

Scattering Parameters Measurement

Figure 9-12 depicts the measurement set up for characterizing the hybrid balun prototype. A network analyzer (NA) is used to measure the scattering parameters of the device under test (DUT). The NA is first calibrated against a 3.5mm SMA calibration reference.

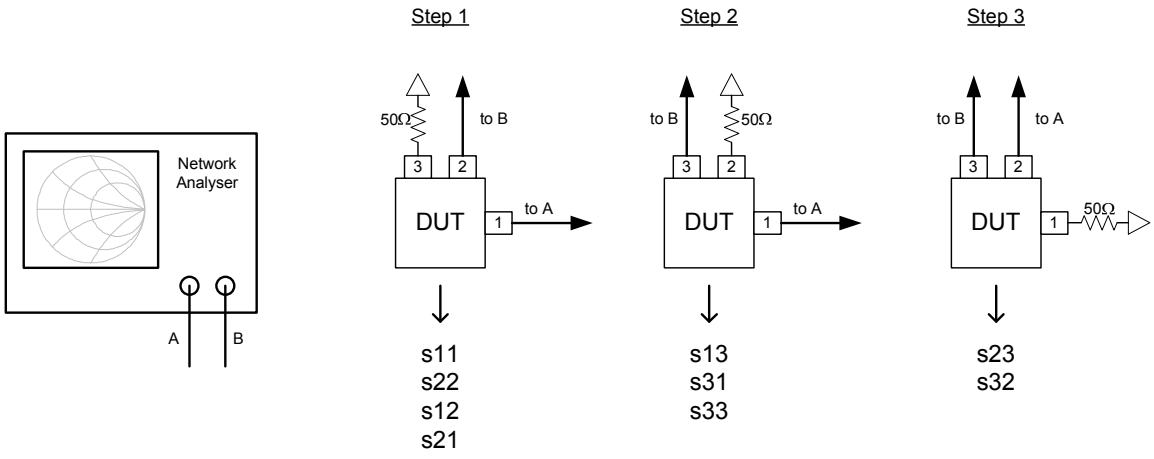


Figure 9-12 Procedure in measuring the 3-port s-parameter of the balun

In order to capture multi-port s-parameter data with the 2-port network analyzer, two out of the 3 ports of the DUT are connected to the NA at a time while the remaining port is terminated by a 50Ω reference termination. Three independent measurements are carried

out to cover all port combinations. Data from each of the nine independent entries in the 3-port scattering parameter matrix are collected against frequency. The raw data are then post-processed to extract the insertion loss, amplitude and phase balance.

Insertion Loss

Figure 9-13 shows the scattering parameter measurements illustrating the transmission, phase balance, and input matching conditions of the balun prototype. When the balun is driven differentially at port 2 and port 3 to produce a single-ended signal at port 1, the insertion loss between the single-ended outgoing power and the total differential incident power can be expressed in terms of the scattering parameters:

$$\text{insertion loss} = \frac{1}{2}\left|s_{12} - s_{13}\right|^2 \tag{9.2}$$

Owing to the reciprocal nature of the balun, and the resulting symmetry in the scattering parameter matrix, the insertion loss can also be expressed in terms of s_{21} and s_{31}. The amplitude of s_{21} and s_{31}, as well as the insertion loss are plotted vs. frequency in Figure 9-13(a).

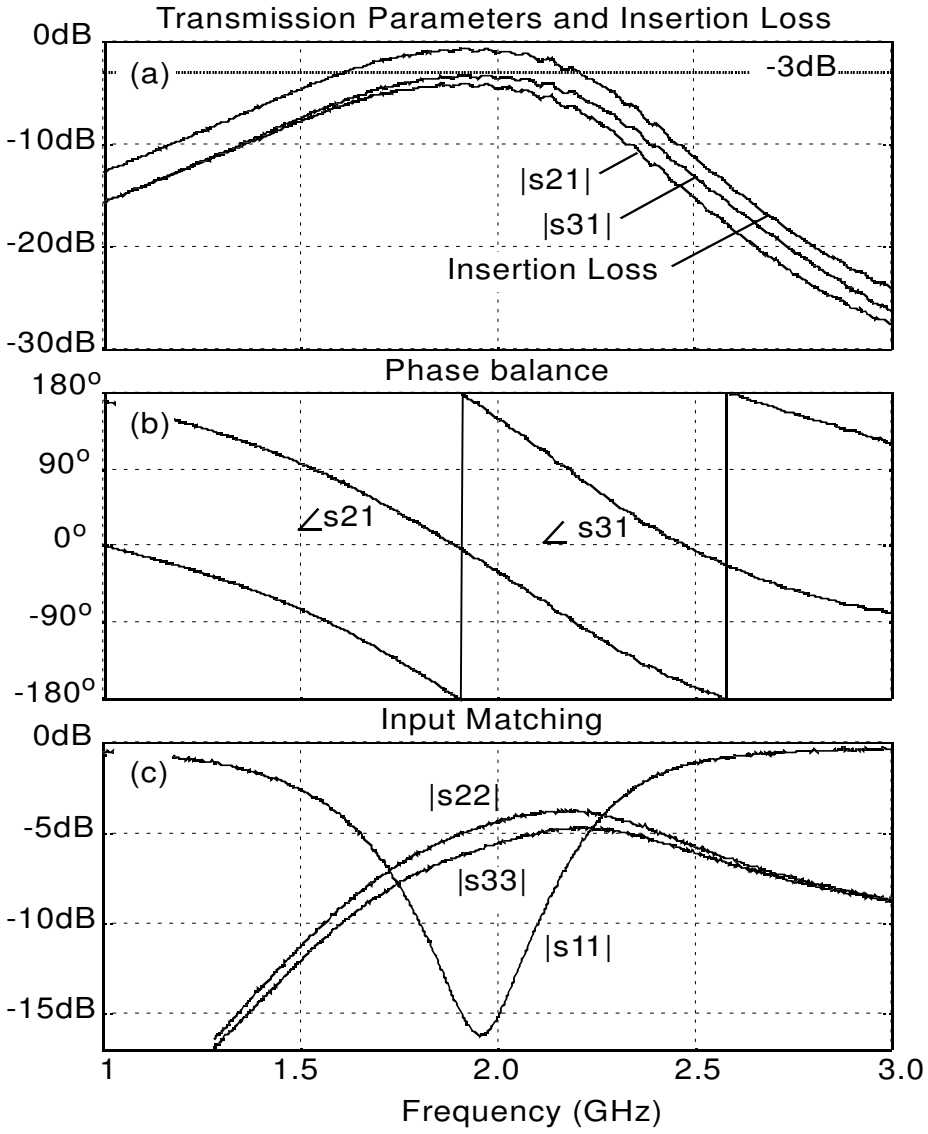


Figure 9-13 Measured scattering parameter of the compact hybrid balun

As is evident in Figure 9-13(a), the amplitude of s_{21} is always less that of s_{31}; this is consistent with (8.18) predicted in our earlier analysis. At the physical dimension of the prototype and the operation frequency, the amplitude imbalance is relatively minor. At 1.98GHz, the balun insertion loss is measured to be 0.7dB which is equivalent to 14.9% of power loss.

Phase Imbalance

The phases of s_{21} and s_{31} are shown in Figure 9-13 (b). It shows that the 180^{o} phase difference between the differential ports is well maintained over a wide range of frequency. The phase error of the balun is defined as

$$\text{phase error} = \angle s_{12} - \angle s_{13} - 180^{o} \tag{9.3}$$

At 1.98GHz, the phase error is approximately 0.5^{o}.

Input Matching

The input matching at each port is plotted in Figure 9-13(c). As can be seen from the figure, the single-ended port is well matched to the reference port impedance (50Ω). The -10dB bandwidth of s11 is about 308MHz.

Balun Bandwidth

The 3dB bandwidth of the balun insertion loss is approximately 500MHz which is wide enough to cover all cellular and ISM bands currently defined around 2GHz. On the other hand, this bandwidth is much narrower than that of conventional, pure transmission line based baluns. As long as the balun bandwidth is wide enough to cover the frequency band of interest, narrower bandwidth is in fact desirable because it helps to reject high frequency harmonics at the PA output. To demonstrate this point, we show in Figure 9-14 the output harmonic contents, measured when the proposed compact balun is connect to the power amplifier. Compared with an identical measurement using a wide-band commercial power combiner (Mini-circuits ZAPDJ-2S), we observed similar second-order harmonic level, but an additional 16-dB reduction in the third order harmonic.

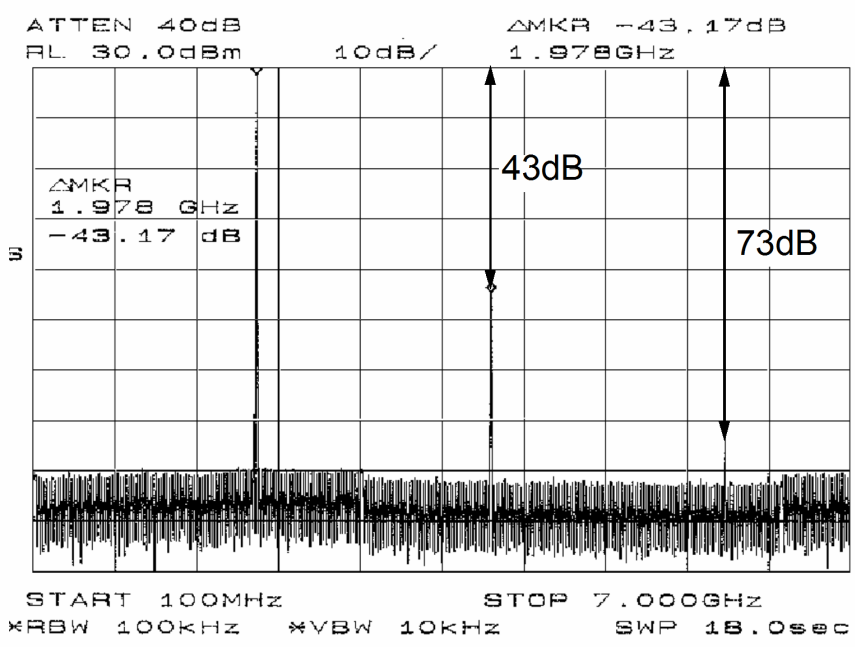


Figure 9-14 Harmonic contents of PA output with proposed compact microstrip balun

9.5 Performance Summary

Table 9-1 summarizes the performance data for the CMOS Class-E PA and the compact balun. Compared with other published CMOS PA works at the same time frame (see Table 9-2), our measurement results demonstrated a new level of performance in terms of operation frequency, output power, and efficiency.

CMOS Class-E Power Amplifier Prototype	
Technology	0.35μm CMOS
Die Size	1.0 x 0.6 mm^2
Package	Chip-on-board
Injection Locking Range	1.68-2.17GHz
Test Frequency	1.98GHz
Supply Voltage	0.6-2.0V
Output Power	50mW – 1W
Input Power	10mW
Power Added Efficiency (PA alone)	48%
Power Added Efficiency (with compact balun)	41%
Second Harmonic (PA alone)	-43dB
Second Harmonic (with compact balun)	-44dB
Third Harmonic (PA alone)	-57dB
Third Harmonic (with compact balun)	-73dB
Noise emission 20-80Mz offset	-130dBm/Hz
Compact Microstrip Balun	
Technology	2-layer FR4 PCB
Core area	7x7mm^2
Center Frequency	1.98GHz
Insertion Loss at center frequency	0.7dB
Phase mismatch at center frequency	0.5^o
3dB Insertion Loss Bandwidth	500MHz
-10dB S11 Bandwidth	380MHz

Table 9-1 Performance Summary

Author	Freq	V_{DD}	Pout	Eff	CMOS Process	Technique
M. Rofougaran VLSI 1994	900MHz	3.0V	20mW	40% drain	1μm	Class-C, Differential
D. Su CICC 1997	850MHz	2.5V	1.0W	42% PAE	0.8 μm	Class-D, single-ended
C. Yoo VLSI 2000	900MHz	1.8V	0.9W	41% PAE	0.25 μm	Class-E, differential, cascode configuration
This work	1.9GHz	2.0V	1.0W	48% PAE	0.35 μm	Class-E, differential, injection lock

Table 9-2 Comparison with other works

Chapter 10 - Conclusion

The demands to reduce cost, size and power consumption of wireless consumer electronics provide a strong motivation for fully integrated CMOS radios. However, despite the wide-spread success in CMOS radio development, realization of efficiency CMOS power amplifiers remains a persisting challenge.

For a broad class of applications that employ constant envelope modulation, switch-mode Class-E operation is identified as a promising solution for CMOS PAs. This is because of the inherently high power efficiency of the switch-mode operation, and its relative

immunity to loss due to charging and discharging of parasitic drain capacitors found in the large CMOS devices. However, for integrated CMOS PAs that deliver beyond 1 Watt of output power at multi-GHz frequencies, direct application of the traditional single-ended Class-E design still involves several key obstacles. Theses obstacles include the input driving requirement of the very large transistors, and the high levels of noise injection into the shared CMOS substrate.

To address these issues, we proposed the use of a multi-stage, fully differential Class-E implementation. We also borrowed the concept of injection locking and applied it to the differential CMOS Class-E PA. To provide an end-to-end solution, a novel hybrid balun was also proposed to efficiently convert the differential PA output signal into signal-ended form. The proposed balun was based on coupled transmission lines and discrete capacitors. It achieved insertion loss comparable to purely distributed baluns, but with a much smaller physical footprint.

To demonstrate the effectiveness of the proposed idea, we implemented a prototype CMOS PA in a 0.35μm single-poly, five-layer metal CMOS process. The PA, intended for applications such as the PCS 1900, delivers 1-W of output power at 1.9GHz with 41% efficiency (including insertion loss of the compact balun). It demonstrated a new level of frequency, output power and power efficiency achievable in CMOS PAs.

In terms of future research, we consider the following area worthy of further exploration:

1. Class-E operation exhibits a relatively high peak drain voltage to supply voltage ratio. We did not experience device failure during the prototype's evaluation, even when the supply was substantially elevated, and the output termination was deliberately mismatched. However, it was not clear whether such stressed operation would cause long-term reliability degradation. Further study on AC reliability limits of CMOS devices is needed in order to understand the PA's true potential and limitation.

2. Class-E PAs feature constant efficiency versus supply voltage, and thus allow the implementation of very efficient power control schemes when a DC-DC converter is used to control the supply voltage. Due to the demand of miniaturizing off-chip loop-filter components, switching frequencies of DC-DC converters are steadily increasing. Exploration in this direction can result in the expansion of supply regulation bandwidth and even enable direct amplitude modulation through the regulator. Such scheme, known as Envelope Elimination and Restoration (ERR), offers an opportunity to expand the usage of switch-mode PAs into a large class of bandwidth efficient, non-constant modulated, signals.

References

CMOS RF integrated Circuits

Technology Trend:

[Cami94] N.Camilleri et al. “New Development Trends for Silicon RF Device Technologies,” *IEEE Microwave and Millimeter-Wave Monolithic Circuits Symposium*, 1994. pp.5-8.

[Gray95] P.R.Gray and R.G. Mayer, "Future Directions in Silicon ICs for RF Personal Communications," *Proc. Custom integrated Circuits Conf.*, May 1995, pp.83-90.

[Stey96] M.Steyaert et al., “RF Integrated Circuit in Standard CMOS Technologies,” *Proc. ESSCIRC*, 1996, pp.11-18.

[Neuv96] Y.Neuvo, “Future Directions in Mobile Communications,” *Proc. ESSCIRC*, 1996, pp.11-18.

[Issa00] R.D. Issac, “The future of CMOS technology,” *IBM Journal of Research and Development*, vol. 44, No. 3, May, 2000, pp.369-378

[ITRS03] ITRS, “Radio Frequency and Analog/Mixed-Signal technologies for Wireless Communications [A section of the Process Integration Chapter]”, 2003

[Raza96] B.Razavi, “Challenges in Portable RF Transceiver Design,” *Circuits and Devices*, Sep. 1996, pp. 12-25.

[Song86] Bang-Sup Song, “CMOS RF Circuits for Data Communications Applications,” *IEEE J. Solid-State Circuits*, Vol.21, No.2, Apr.1986, pp. 310-317.

[Cave98] R.H.Caverly et al., “CMOS RF Circuits for Integrated Wireless Systems,” *IEEE MTT-S Digest*, 1998, pp. 1851

[Huan98] Q.Huang, F.Piazza, P.Orsatti, T.Ohguro, “The impact of scaling down to deep submicron on CMOS RF circuits,” *IEEE J. Solid-State Circuits*, Vol.33, No.7, July 1998, pp1023-1036.

CMOS Transceivers:

[Sheng96] S. Sheng et al., “A Low-Power CMOS Chipset for Spread-Spectrum Communications,” *Digest ISSCC*, 1996, pp. 946.

[Rude97] J. Rudell, et al., “A 1.9-GHz wide-band IF double conversion CMOS receiver for cordless telephone applications”, *IEEE J. solid State Circuits*, vol. 32, Dec. 1997, pp. 2071-2088.

[Rofo98] A. Rofougaran, et al., “A single-chip 900-MHz spread-spectrum wireless transceiver in 1-µm CMOS. I. Architecture and transmitter design,” *IEEE J. Solid State Circuits*, vol.33, Apr.1998, pp. 512-534.

[Rofo98b] A. Rofougaran, et al., “A single-chip 900-MHz spread-spectrum wireless transceiver in 1-µm CMOS. II. Receiver design,” *IEEE J. Solid State Circuits*, vol.33, Apr.1998, pp. 512-534.

[Cho99] T. Cho, et al., “A single-chip CMOS direct-conversion transceiver for 900 MHz spread-spectrum digital cordless phones”, *ISSCC Dig. Tech. Papers*, San Francisco, CA, Feb. 1999.

[Dara01] H. Darabi, et al. “A 2.4GHz CMOS transceiver for Bluetooth”, *IEEE J. Solid State Circuits*, vol. 36, Dec 2001, pp. 2016-2024.

PAs for Wireless Communications:

History and Overviews:

[Espe59] Espenscheid, L. et al., “History of some Radio Foundations,” *Proc. IRE*, vol. 47, 1959, pp. 1253-1268

[Soka95] N. O. Sokal, “RF power amplifiers-classes A through S,” *IEEE Program Proc. Professional Electro’95 Int.*, 1995, pp. 335–400.

[Raab99] F. Raab, “Transmitter Technology for the Second Century of Wireless,” Presentation 1999.

[Gupt99] R.Gupta, “Fully Monolithic CMOS RF Power Amplifiers: Recent Advances,” *IEEE Communications Magazine*, April 1999, pp.94-98.

[Crip99] S. Cripps, “RF Power Amplifiers for Wireless Communications,” Artech House, 1999.

[Abul01] M. Abulet, “RF Power Amplifiers,” Noble Publishing, 2001.

[Matt97] S. Mattisson, “Minimizing Power Dissipation of Cellular Phones,” *Proc. Int. Sym. Low Power Electronics Design*, 1997, pp.42-45.

[Sevi97] J. F. Sevic, “Statistical characterization of RF power amplifier efficiency for CDMA wireless communication systems,” *Proc. 1997 Wireless Communications Conf.*, 1997, pp. 110–113.

CMOS Realization

[Rofo94] M. Rofougaran, A. Rofougaran, C. Olgaad and A.A. Abidi, "A 900 MHz CMOS RF Power amplifier with programmable output," *IEEE Symp. VLSI Circuits Dig. Tech. Papers*, 1994, pp.133-134.

[Rofo95] M. Rofougaran, “A 900MHz RF Power Amplifier in 1µm CMOS for a Spread-Spectrum Communication Transceiver,” M.S. Thesis UCLA, 1995.

[Su97] D. Su and W. Mcfarland, "A 2.5-V, 1-W Monolithic CMOS RF Power Amplifier," *Proc. Custom Integrated Circuits Conf.*, May 1997, pp.189-192.

[Tsai99] K.C. Tsai, P.R. Gray, "A 1.9-GHz 1-W CMOS Class-E Power Amplifier for Wireless Communications," *IEEE J. Solid-State Circuits*, Vol. 34, Jul. 1999, pp.962-970.

[Yoo01] C. Yoo; Q. Huang, “ A common-gate switched 0.9-W class-E power amplifier with 41% PAE in 0.25-µm CMOS,” *IEEE J. Solid-State Circuits*, Vol.36, No.5, May 2001, pp823-830.

[Gupt01] R.Gupta, B.M. Ballweber, D.J. Allstot, “Design and optimization of CMOS RF power amplifiers,” *IEEE J. Solid-State Circuits*,Vol.36, No.2, Feb. 2001, pp.166-175.

[Mert02] K.L.R. Mertens, M.S.J. Steyaert, “A 700-MHz 1-W fully differential CMOS class-E power amplifier,” *IEEE J. Solid-State Circuits*, Vol.37, No.2, Feb. 2002. pp.137 – 141.

[Aoki02] I.Aoki, S.D.Kee, D.B.Rutledge, A. Hajimiri, “Fully integrated CMOS power amplifier design using the distributed active-transformer architecture,” *IEEE J. Solid-State Circuits*, Vol.37, No.3, Mar. 2002, pp371-383.

Linearity and Linearization:

[Kahn52] L. Kahn, “Single-Sideband Transmission by Envelope Elimination and Restoration,” *Proc. of the I.R.E.*, Jul. 1952, pp.803-806.

[Su98] D. Su, W.J. McFarland, “An IC for linearizing RF power amplifiers using envelope elimination and restoration,” *IEEE J. Solid-State Circuits*, Vol. 33, No.12, Dec. 1998, pp2252-2258.

[Ariy90] S. Ariyavisitakul, T.-P. Liu, “Characterizing the Effects of Nonlinear Amplifiers on Linear Modulation for Digital Portable Radio Communications,” *IEEE Tran. Vehicular Technology*, Vol. 39, Issue. 4, Nov 1990, pp.383-389.

[Bocc95] J.Boccuzzi, “Performance Evaluation of Non-Linear Transmit Power Amplifiers for North American Digital Cellular Portables,” *IEEE Tran. Vehicular Tech.*, Vol.44, No.2, May 1995, pp.220-229.

[Sowl97] T. Sowlati, Y.M. Greshishchev, C.A.T. Salama, “Phase-correcting feedback system for class E power amplifier,” *IEEE J. Solid-State Circuits*, Vol.32, No.4, April 1997, pp544-549.

Constant Envelope Modulations

[Muro81] K. Murota, K. Hirade, “GMSK modulation for digital mobile radio telephony,” *IEEE Trans. Communications*, vol. COM-29, pp. 1044–1050, July 1981.

[Ande91] J.Anderson, C.-E.W.Sundberg, “Advances in Constant Envelope Coded Modulation,” *IEEE Communications Magazine*, Dec. 1991, pp.36-45.

Class-E PAs:

Classical Analyses

[Soka75] N.O. Sokal, A. Sokal, “Class E—A new class of high-efficiency, tuned single-ended switching power amplifiers,” *IEEE J. Solid-State Circuits*, vol. SC-10, pp. 168–176, June 1975.

[Soka77] N.O. Sokal, F.H. Raab, “Harmonic output of class-E RF power amplifiers and load coupling network design,” *IEEE J. Solid-State Circuits*, Vol.12, No.1, Feb 1977, pp86-88

[Raab77] F. H. Raab, “Idealized operation of the class E tuned power amplifier,” *IEEE Trans. Circuits and Systems*, vol. CAS-24, Dec. 1977, pp. 725–735

[Raab78] F. H. Raab," Effects of circuit variations on the class E tuned power amplifier,” *IEEE J. Solid-State Circuits*, Vol.13, No.2, Apr 1978, pp239-247

[Raab78b] F. H. Raab, N. O. Sokal, “Transistor power losses in the class E tuned power amplifier,” *IEEE J. Solid-State Circuits*,Vol.13, No.6, Dec 1978, pp912-914.

[Kazi83] M.Kazimierczuk, “Effects of the collector current fall time on the class E tuned power amplifier,” *IEEE J. Solid-State Circuits*, Vol.18, No.2, Apr 1983, pp181-193.

[Moln84] B. Molnar, “Basic limitations on waveforms achievable in single-ended switching-mode tuned (Class E) power amplifiers,” *IEEE J. Solid-State Circuits*, Vol.19, Issue: 1, Feb 1984, pp144-146.

[Kazi86] M. Kazimierczuk, “Class E tuned power amplifier with non-sinusoidal output voltage,” *IEEE J. Solid-State Circuits*, Vol.21, No.4, Aug 1986, pp575-581.

[Char89] P.Charampos et al, “Analysis and Design of a Generalized Class E Tuned Power Amplifier,” *IEEE Tran. Circuits and Systems*, Vol.36, No.8, Aug.1989, pp.1068-1079.

State-Space Analysis and Simulation:

[Mand90] J. Mandojana, K. Herman, and R. Zulinski, “A discrete/continuous time domain analysis of a general class E amplifier,” *IEEE Trans. Circuits and Systems*, vol. 37, Aug. 1990, pp. 1057–1060.

[Mand92] J. C. Mandojana and R. E. Zulinski, “Computer-aided design of class E amplifiers,” *Proc. 34th Midwest Symp. Circuits and Systems*, 1992, vol. 2, pp. 866–869.

[Herm90] K. J. Herman, R. E. Zulinski, and J. C. Mandojana, “An efficient computer program for the exact analysis of class E amplifier,” *Proc. 32nd Midwest Symp. Circuits and Systems*, 1990, vol. 1, pp. 478–481.

PA Realization Technologies:

[Jos01] R.Jos, “Technology developments driving an evolution of cellular phone power amplifiers to integrated RF front-end modules,” *IEEE J. Solid-State Circuits*, Vol.36, No.9, Sep. 2001, pp.1382-1389.

[Goli03] M. Golio, “RF and Microwave Semiconductor Device Handbook,” CRC Press, 2003.

[Chan02] K. Chang et al. “RF and Microwave Circuit and Component Design,” Wiley-Interscience, 2002

[Berg99] O. Berger, “GaAs MESFET, HEMT and HBT Competition with Advanced Si RF Technologies,” GaAs Mantech, 1999.

[Matt74] J.Matthews and A. Blakeslee, Defects in Epitexial Multilayers,” *J. Crystal Growth*, 27, p.118-125, 1974.

[Meye96] B.Meyerson et al., “Silicon: Germanium Heterojuction Bipolar Transistor; from Experiment to Technology,” Chapter of “Current Trends in Heterojunction Bipolar Transistors,” World Scientific, 1996, pp.367.

[Patt90] G.L.Patton et al., “75GHz fT SiGe base heterojunction bipolar transistors”, *IEEE Electron Device Letter*, 1990, pp.171-173.

[Schu95] A. Schuppen et al., “Enhanced SiGe heterojunction bipolar transistor”, *IEDM Technical. Digest*, 1995, pp.743-746.

[Oda97] K. Oda et al., “130 GHz fT SiGe HBT technology”, *IEDM Technical Digest*, 1997, pp.791-794.

[King00] C. King, “SiGe Bipolar Transistors for High Speed Optical Networking and Wireless Communication Applications,” Presentation, Jul, 2000.

CMOS Reliability

[Wolt85] D.R.Wolters and J.J.Van der Schoot, “Dielectric Breakdown in MOS devices. Part I: Defect Related and Intrinsic Breakdown,” *Philips J. Res.* 40, pp.115, 1985.

[Lee88] J.Lee, I.-C.Chen, C.Hu, “Modeling and Characterization of Gate Oxide Reliability,” *IEEE Tran, in Electronic Devices*, Dec. 1988, pp.2268-2278.

[Moaz89] R.Moazzami, J.C.Lee and C.Hu, “Temperature acceleration of time-dependent dielectric breakdown,” *IEEE Trans. on Electron Devices*, 1989, pp.2462.

[Sabr90] A.G. Sabnis, VLSI Reliability, VLSI Electronics Microstructure Science, Vol. 22, Academic Press, 1990.

[Hu96] C. Hu, “Gate Oxide Scaling Limits and Projection,” *IEDM Technical Digest*, 1996, pp.319-322.

[Ohri98] M.Ohring, “Reliability and Failure of Electronic Materials and Devices,” Ch.6 and Ch.12, Academic Press, 1998.

[Hu98] C. Hu, “Reliability Phenomena under AC Stress,” *Microelectronic Reliability* vol.38, 1998, pp.1-5.

Injection Locking:

[Adle46] R. Adler, “A Study of Locking Phenomena in Oscillators,” *Proc. of I.R.E. and Waves and Electrons*, Jun. 1946, pp.351-357.

RF Baluns:

[Marc44] Marchand N., “Transmission-Line Conversion,” *Electronics*, Dec. 1944, pp142-145.

[Delv02] L.M. Delvin et al., “A monolithic, 2 to 18 GHz upconverter,” *Microwave Symposium Digest, 2002 IEEE MTT-S International*, Vol. 1, Jun 2002, pp. 257-260.

[Jone56] E. M. T. Jones and J. T. Bolljahn, “Coupled-strip-transmission-line filters and directional couplers,” *IEEE Trans. Microwave Theory Tech.*, vol. MTT-4, Apr. 1956, pp. 75–81.

[Fook90] Fooks E.H. and Zakarevicius R.A., “Microwave Engineering Using Microstrip Circuits.” New York Prentice Hall, 1990, pp. 3191.

[Roge91] J. Rogers and R. Bhatia, “A 6 to 20 GHz planar balun using a Wilkinson divider and Lange couplers,” *IEEE MTT-S Int. Microwave Symp. Digest*, 1991, vol. 2, pp. 865–868.

[Poza93] D. M. Pozar, “Microwave Engineering,” Addison-Wesley, 1993.

[Nguy93] C. Nguyen, D. Smith, “Novel miniaturized wideband baluns for MIC and MMIC applications,” *Electronics Letters*, Vol. 29, No. 12, Jun 1993, pp. 1060 – 1061.

[Ojha96] S. Ojha, G. R. Branner, and B. P. Kumar, “A miniaturized lumped distributed balun for modern wireless communication systems,” *Proc. 39th Midwest Symp. Circuits and Systems*, 1996, vol. 3, pp. 1347–1350.

[Kuma96] B. P. Kumar, G. R. Branner, G. Razmafrouz, “An improved planar balun design for wireless microwave and RF applications,” *IEEE 39th Midwest symposium Circuits and Systems*, 1996, Vol. 1, 18-21 Aug. 1996, pp. 257-260.

[Delv02] L. M. Delvin, et al. “A monolithic, 2 to 18 GHz upconverter,” *IEEE MTT-S International Microwave Symposium Digest*, 2002, Vol. 1, Jun 2002, pp257-260.

Monolithic and bondwire Inductors:

[Gree74] H. Greenhouse, “Design of Planar Rectangular Microelectronic Inductors,” *IEEE Tran. on Parts, Hybrids, and Package*, Vol. PHP-10, No. 2, Jun. 1974, pp.101-109.

[Nguy90] N. Nguyen, R. Meyer, “Si IC-Compatible Inductors and LC Passive Filters,” *IEEE J. Solid-States Circuits*, Vol.25, No.4, Aug.1990, pp.1028-1031.

[Lee98] Y. G. Lee et al., “Novel high-Q bondwire inductor for MMIC,” *IEDM Technical Digest*, 1998, pp.548-551.

[Cami94b] N. Camilleri et al., “Bonding Pad Models for Silicon VLSI Technologies and Their Effects on the Noise Figure of NPNs,” *IEEE MTT-S Digest*, 1994, pp.1179-1182.