High-Efficiency Linear RF Power Amplifiers Development

A Dissertation
Presented to
The Academic Faculty

by

Nuttapong Srirattana

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Electrical and Computer Engineering
Georgia Institute of Technology
11 April 2005

High-Efficiency Linear RF Power Amplifiers Development

Approved by:

Dr. Phillip E. Allen, Advisor
College of Engineering
Georgia Institute of Technology

Dr. John Papapolymerou
College of Engineering
Georgia Institute of Technology

Dr. Joy Laskar, Co-Advisor
College of Engineering
Georgia Institute of Technology

Dr. Ye Li
College of Engineering
Georgia Institute of Technology

Dr. John D. Cressler
College of Engineering
Georgia Institute of Technology

Dr. Paul A. Kohl
College of Engineering
Georgia Institute of Technology

Date Approved: April 11, 2005

To Noppawan, Chongrak, and Nuttapan Srirattana,

my beloved family.

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my advisors, Prof. Phillip E. Allen and Prof. Joy Laskar, for granting me a wonderful opportunity to conduct this interesting research and for their guidance and encouragement throughout the work. I gratefully acknowledge Prof. John Cressler, Prof. John Papapolymerou, Prof. Ye Li, and Prof. Paul Kohl for serving on my defense committee.

I would like to thank the members of Microwave Application Group, past and present, for their assistance, cooperation, and encouragement. Special thanks are given to Deukhyoun Heo and Arvind Raghavan for introducing me to RF power amplifier design, to Hyun-Min Park for valuable discussions and help with measurements, and to Sebastien Nuttinck, Chang-Ho Lee, Kyu-Tae Lim, Sang Woong Yoon, Padmanava Sen, and Rajarshi Mukhopadhyay for collaborations in power amplifier design and device modeling.

I would also like to thank National Semiconductor Corporation for giving me an internship opportunity during 2001 and 2002. I would like to thank Vikram Krishnamurthy (now with VT Silicon), Emery Chen, and Andy Mclean (National Semiconductor Corporation) for the opportunity to work with them and for many helpful discussions. I would like to thank Patrick O'Farrell and Arlo Aude (National Semiconductor Corporation) for their help with circuit layout and submissions, and Stan Halpern (Georgia Tech Research Institute) for help in wire-bonding. I would also like to thank National Semiconductor Corporation and Samsung Electronic for providing CMOS and SiGe BiCMOS processes used in this work.

My family has always given me strong support and encouragement in my studies and in other aspects of my life. I would like to thank and express my deepest love to

everyone in my family (Noppawan and Chongrak Srirattana, my parents, Nuttapan Srirattana, my brother) from the bottom of my heart. Special thanks are also given to Thidarat Tosukhowong for her love, support, understanding, and encouragement. Finally, I would like to thank all my friends for supporting and encouraging me throughout my stay at Georgia Tech.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Next generation mobile communication systems require the use of linear RF power amplifier for higher data transmission rates. However, linear RF power amplifiers are inherently inefficient and usually require additional circuits or further system adjustments for better efficiency. This dissertation focuses on the development of new efficiency enhancement schemes for linear RF power amplifiers.

The multistage Doherty amplifier technique is proposed to improve the performance of linear RF power amplifiers operated in a low power level. This technique advances the original Doherty amplifier scheme by improving the efficiency at much lower power level. The proposed technique is supported by a new approach in device periphery calculation to reduce AM/AM distortion and a further improvement of linearity by the bias adaptation concept.

The device periphery adjustment technique for efficiency enhancement of power amplifier integrated circuits is also proposed in this work. The concept is clearly explained together with its implementation on CMOS and SiGe RF power amplifier designs. Furthermore, linearity improvement technique using the cancellation of nonlinear terms is proposed for the CMOS power amplifier in combination with the efficiency enhancement technique.

In addition to the efficiency enhancement of power amplifiers, a scalable large-signal MOSFET model using the modified BSIM3v3 approach is proposed. A new scalable substrate network model is developed to enhance the accuracy of the BSIM3v3 model in RF and microwave applications. The proposed model simplifies the modeling of substrate coupling effects in MOS transistor and provides great accuracy in both small-signal and large-signal performances.

CHAPTER 1

INTRODUCTION

1.1 The Need for High-Efficiency Linear RF Power Amplifiers

The third-generation (3G) mobile communication systems have become prominent and essential for an ever increasing demand for high data-rate communications. In 3G systems, numerous wireless standards with advanced modulation techniques are employed for faster transmission of voice, video, and multimedia applications up to 2 Mbps. Digital modulation schemes are used to modulate data in both amplitude and phase for the efficient use of frequency bandwidth. Widely adopted techniques include quadrature phase-shift keying (QPSK) and 8PSK used in wideband code-division multiple access (WCDMA) and the enhanced data rates for global evolution (EDGE), which is the next generation of the earlier second-generation (2G) global system for the mobile communication (GSM) standard (see Fig. 1). In EDGE, the signal envelope is not constant and the difference between average power and peak power increases the requirements for linearity of the transmitted signal. For WCDMA, each coded channel utilizes a wide frequency bandwidth that is overlapped onto other channels. As a consequence, linear amplification is required to prevent spurious signals that lead to spectral regrowth of the frequency range.

The emerging fourth-generation (4G) mobile communications have been developed to handle the need for higher data transmission rates beyond the capability of 3G systems. The proposed 4G systems, expected to be launched by 2010, utilize the combination of CDMA and orthogonal frequency-division multiplexing (OFDM) for

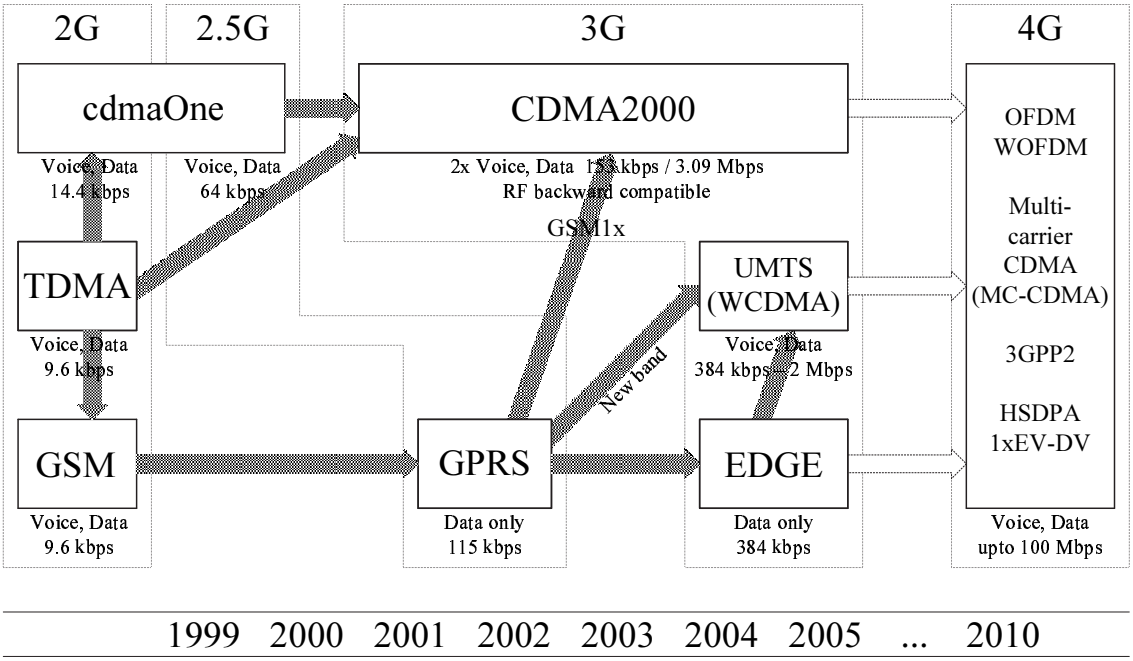


Figure 1: The evolution of mobile communication systems.

extremely high data-rate transmission (up to 100 Mbps.) This multi-carrier modulation helps to reduce intersymbol interference (ISI) but increases the peak-to-average ratio of the signal. Also, this technique still uses QPSK modulation, which requires linear amplification.

From the RF/microwave power amplifier design point of view, achieving high efficiency and high linearity simultaneously is the most challenging problem. It is known that power amplifiers, in general, cannot operate efficiently when used for linear amplification since most of the dc input power is sacrificed in generating head room for a high level of linearity. This problem is aggravated by the probability distribution of the average transmitted power. Frequently, power amplifiers used in many applications are operated at a very low power level. For example, WCDMA has a peak-to-average ratio of 8-12 dB, while the IS-95 CDMD has a peak-to-average ratio of more than 20 dB. Therefore, constantly operating the power amplifier at its maximum level will result in inefficient dc supply usage and affect the longevity of

the equipment operation.

For many years, much effort has been expended in the development of techniques to improve the efficiency of linear RF power amplifiers, especially in the low-power (or "backed-off") region, where the power amplifier normally operates. Several methods, such as envelope elimination and restoration (EER) and bias adaptation, have been explored extensively. However, these methods require the use of external control circuits and signal processing, resulting in an increased level of design complexity. Therefore, it is desirable to develop better efficiency enhancement schemes that overcome the limitations of existing techniques.

For linear power amplifier systems, the Doherty amplifier is normally considered a method to improve efficiency in the low-power regime. The most interesting feature of the Doherty amplifier is that it can function without additional components or circuits. This "self-managing" characteristic of the Doherty amplifier has made its implementation attractive for various applications. However, the Doherty amplifier can achieve an efficiency improvement of only up to 6 dB from maximum system power.

In this work, a new technique called the "multistage Doherty amplifier," which applies the concept of the Doherty amplifier, is developed to extend the efficiency improvement capability beyond the classical design. This work presents a thorough analysis of the multistage Doherty power amplifier and an implementation of a three-stage Doherty power amplifier system with significant efficiency enhancement at backed-off conditions to meet the WCDMA uplink standard specifications. It is shown that, by following the given analysis, performance of the multistage Doherty amplifier can be achieved with great correlation to ideal simulation. The practical issue of AM/AM distortion regarding the choice of device periphery is explored and described mathematically and proven through a design prototype.

In addition to the multistage Doherty amplifier, a new efficiency enhancement

scheme is developed using the concept of device periphery adjustment to minimize dc power consumption. The concept is developed based on previously available techniques, including the reduction of dc current by stage bypassing and current level switching. In this proposed technique, the dc current is optimized between the level required for maximizing device performances and the current needed to produce sufficient required output power. This concept has been further implemented in designing 1.9-GHz linear power amplifiers in CMOS and SiGe processes. The design illustrates a compact integration of a power amplifier and an efficiency enhancement scheme that helps improve efficiency up to a very low-power level. In addition, the linearity enhancement in a FET amplifier using the cancellation of nonlinear current components is adopted to enhance the benefit of using a multiple-device amplifier.

Another important concern in designing RF power amplifier circuits is the availability of an accurate large-signal model. In CMOS RF power amplifier design, the available BSIM3v3 model from process manufacturers does not provide enough accuracy in RF applications because of the lack of gate resistance and substrate resistance models. Therefore, an important part of this thesis is devoted to the MOSFET transistor modeling. In the model development, the BSIM3v3 model is modified to include additional parameters such as gate resistance and substrate resistance components, which are extracted from S-parameter measurements. Modeling accuracy is clarified by the comparison of simulation results to S-parameter measurements and large-signal measurements such as power sweeps and load-pull measurements.

1.2 Organization

The rest of this thesis is organized as follows. Chapter 2 briefly describes the basics of efficiency enhancement techniques in linear RF power amplifiers. The classes of operation in RF power amplifiers are explained in some detail. This shows how the difference in bias point and output termination can lead to the change in amplifier

efficiency. Furthermore, efficiency enhancement techniques are discussed and compared to show the advantages and drawbacks of each approach. Main techniques to be discussed are the envelope elimination and restoration, the bias adaptation, the Doherty amplifier technique, and the outphasing technique. In addition, practical design issues in RF power amplifiers including device modeling, load-pull measurement technique, and other important considerations in RF power amplifier circuits are described to support the conventional approach in RF power amplifier design.

Chapter 3 describes the Doherty amplifier concept and mathematical approaches for understanding the Doherty amplifier function. The basic concept is further exploited to analyze the multistage Doherty amplifier, where the order of the design is enhanced for better efficiency improvement. The design of a three-stage Doherty power amplifier is demonstrated and verified with measurement results that emphasize the success of the approach in improving the efficiency of RF power amplifiers for modern communication systems.

Chapter 4 is devoted to the development of scalable large-signal MOSFET models for RF applications. It starts with a general concept of MOSFET modeling and describes the need for further improvements of the current BSIM model for RF and microwave applications. Issues such as gate resistance, non-quasi static effects, and substrate coupling effects are explained with suggested solutions. The methodology for small-signal modeling and its parameter extraction are also given in details. The extraction results are shown and further used in the implementation of large-signal MOSFET models based on the BSIM3v3 approach. The proposed model incorporates scalable gate resistance and scalable substrate resistance networks to improve the accuracy of the BSIM3v3 model at higher frequencies. The proposed substrate resistance model alleviates the complexity in modeling substrate coupling effects and is also further simplified to be used with the BSIM3v3 model without any adjustment in the core model, thus making it easier for model implementation. Several results from

small-signal S-parameter measurements and large-signal load-pull measurements are made available for model verification.

Chapter 5 focuses on the design and implementation of the device periphery adjustment technique for efficiency enhancement of power amplifiers at the integrated-circuit level. The chapter starts by introducing existing methods used in today's RF power amplifier integrated circuits before the proposed concept is explained. Then, the implementation of the technique in CMOS and SiGe processes is shown and described in details. This also includes the design of matching networks, biasing circuits, and practical design issues. Measurement results are provided to ensure the performance of the design in overcoming the low-efficiency problems in the low-power region. In addition, the analysis of linearity enhancement using the concept of nonlinear cancellation that is possible in FET amplifiers is described and demonstrated with two-tone measurements. Finally, Chapter 6 summarizes the contributions of the dissertation and provides ideas for future research in this area.

CHAPTER 2

BASICS OF EFFICIENCY ENHANCEMENT TECHNIQUES IN RF POWER AMPLIFIERS

2.1 Classes of Operation in RF Power Amplifiers

Power amplifiers are classified into several classes of operation depending on the output current waveform as shown in Fig. 2. Class-A amplifier operates in full input and output ranges where the current waveform has no clipping or distortion, which makes it a perfect operation for an amplitude-modulated signal. However, it requires higher quiescent current and becomes less efficient compared to other classes of operation, thus limiting its use in mobile applications.

In class-B operation, the amplifier draws current during only one-half of the input drive interval and the transistor is on for half of the cycle. Therefore, the transistor consumes less power than class-A operation and is more efficient. Class-B amplifiers are less linear than class-A because of the signal distortion (current clipping). If both linearity and efficiency are concerns, a better option is to operate the amplifier between the class-A and class-B region to alleviate linearity issues while still having better efficiency than class A. This region of operation is called class AB, where the current is drawn for more than half of the cycle but not as much as in class A. Sometimes the behavior of the current waveform is explained by using the term "conduction angle", θ , where 360° is for class-A operation and 180° is for class-B. Class-AB has a conduction angle between 180° and 360°. Since the class-AB operation is more efficient than the perfect class A and is more linear than class-B, it is often used in the applications that require good linearity and efficiency such as 3G mobile applications.

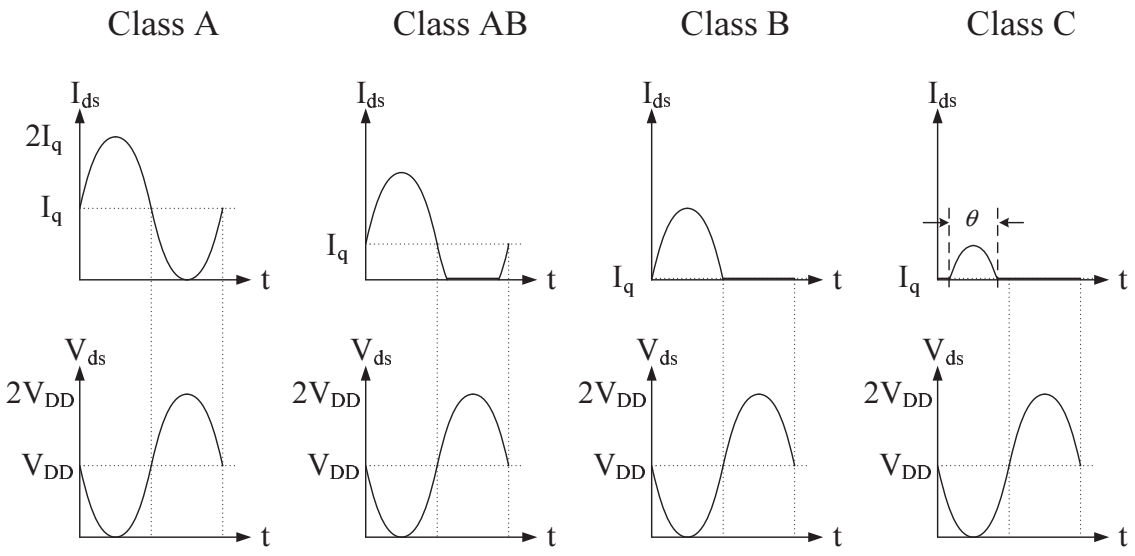


Figure 2: Voltage and current waveforms of power amplifier in different classes of operation.

If the transistor operates for less than half of the cycle or with the conduction angle smaller than 180°, it is called class-C operation. Class-C operation is highly efficient, and mostly used in nonlinear applications, especially in the constant-envelope modulation where only the phase of the signal is used to convey data. Class C is not suitable for linear operation because of strong distortions in signal amplitude.

The maximum drain efficiency (η) of an amplifier can be calculated as a function of the conduction angle by the following expression [32]

$$\eta = \frac{1}{4}\frac{\theta - sin\theta}{sin(\theta/2) - (\theta/2)cos(\theta/2)} \tag{1}$$

The maximum drain efficiency of class-A amplifier ($\theta = 360°$) is 50% and will be higher for smaller conduction angle toward class B ($\theta = 180°$; $\eta = 78.5\%$) and C ($\theta = 0°$; $\eta = 100\%$). However, the output power to the load will drop rapidly when η approaches zero, as given by [32]

$$P_{out} \propto \frac{\theta - sin\theta}{1 - cos(\theta/2)} \tag{2}$$

Therefore, it is not possible to use class-C bias to achieve 100% efficiency while giving full power, thus limiting its use in applications that require both high efficiency and high power.

For constant-envelope operation, efficiency can be improved by the use of harmonics to shape the waveform. Such an approach can be applied to class-A operation by overdriving the input signal and terminating the output harmonics to make the drain voltage similar to a square wave. This is sometimes called an "overdriven class-A amplifier", which is different from class AB since the transistor is still biased in class A (bias point locates at the midway between the cutoff and the saturation.) In this case the drain voltage has sharper edges (smaller rise time and fall time) than a normal sinusoidal waveform, thus reducing the overlap between output current and voltage waveforms, and eventually increasing the efficiency.

The concept of reducing the overlap between voltage and current waveforms has been applied to operate a transistor as a switch, sometimes called a "switching amplifier". In an ideal switch, the current and voltage across the switch are not present at the same time, therefore contributing zero power dissipation. This concept is used in class-D, -E, and -F amplifiers. Class-D amplifiers use transistors to switch between point A and B in Fig. 3, resulting in a square waveform across the switch without the coexistence of current. Each switch conducts a half sine wave that is combined to form a full sine wave in the series LCR resonator. Class-D is normally used in the low-frequency range (10 - 30 MHz region) where parasitic reactances of the switch are not an issue. Practical problems are mainly the control of switches that must be synchronous with no overlapping interval to avoid leakage current and unwanted dc dissipation. Also, even though the ideal efficiency of class-D amplifier can be as high as 100 %, the non-zero saturation voltage of the transistor will cause power dissipation and decrease efficiency.

The class-E amplifier (Fig. 4) uses a transistor to switch the current flowing

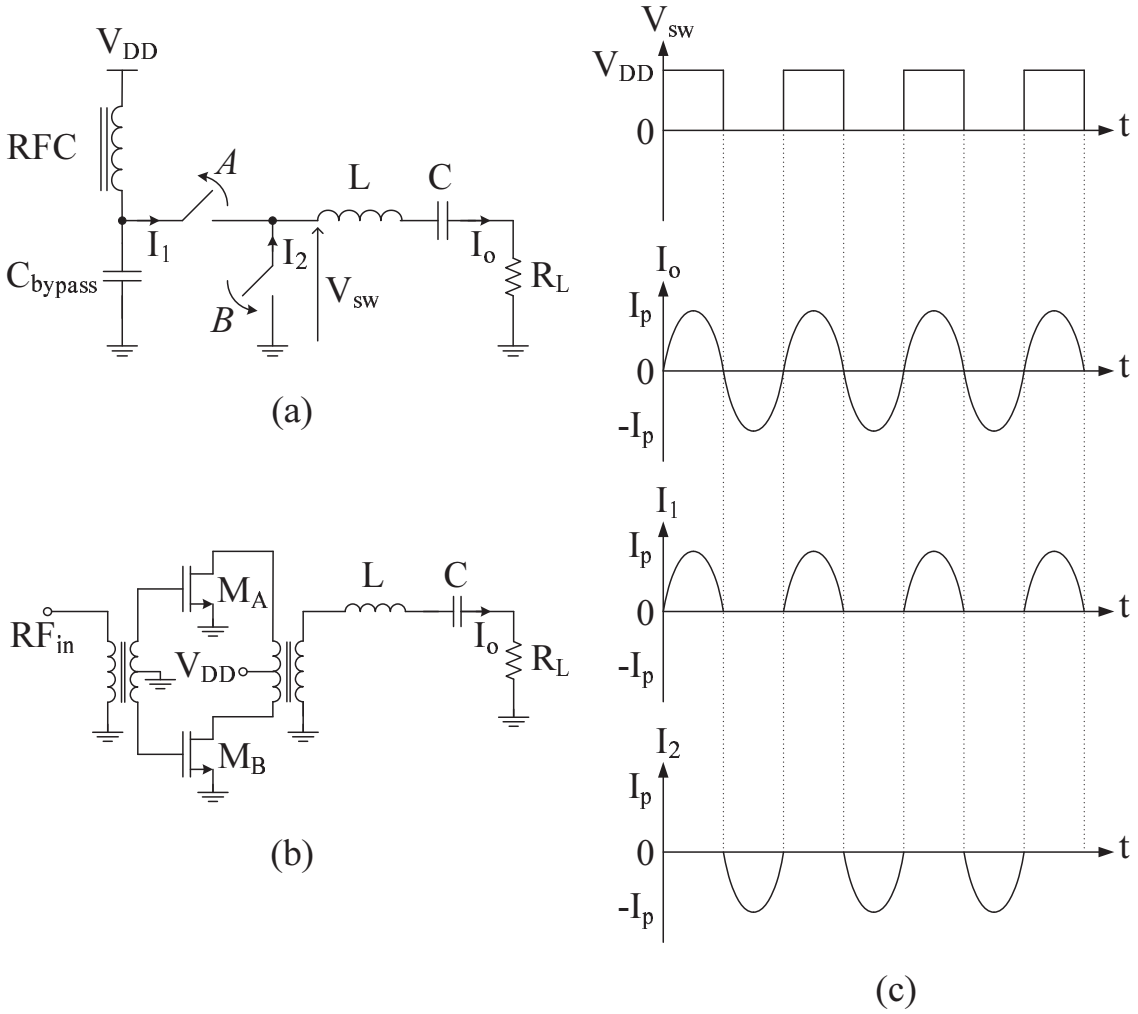


Figure 3: Class-D operation; (a) schematic diagram, (b) push-pull class-D amplifier, (c) voltage and current waveforms of class-D amplifier.

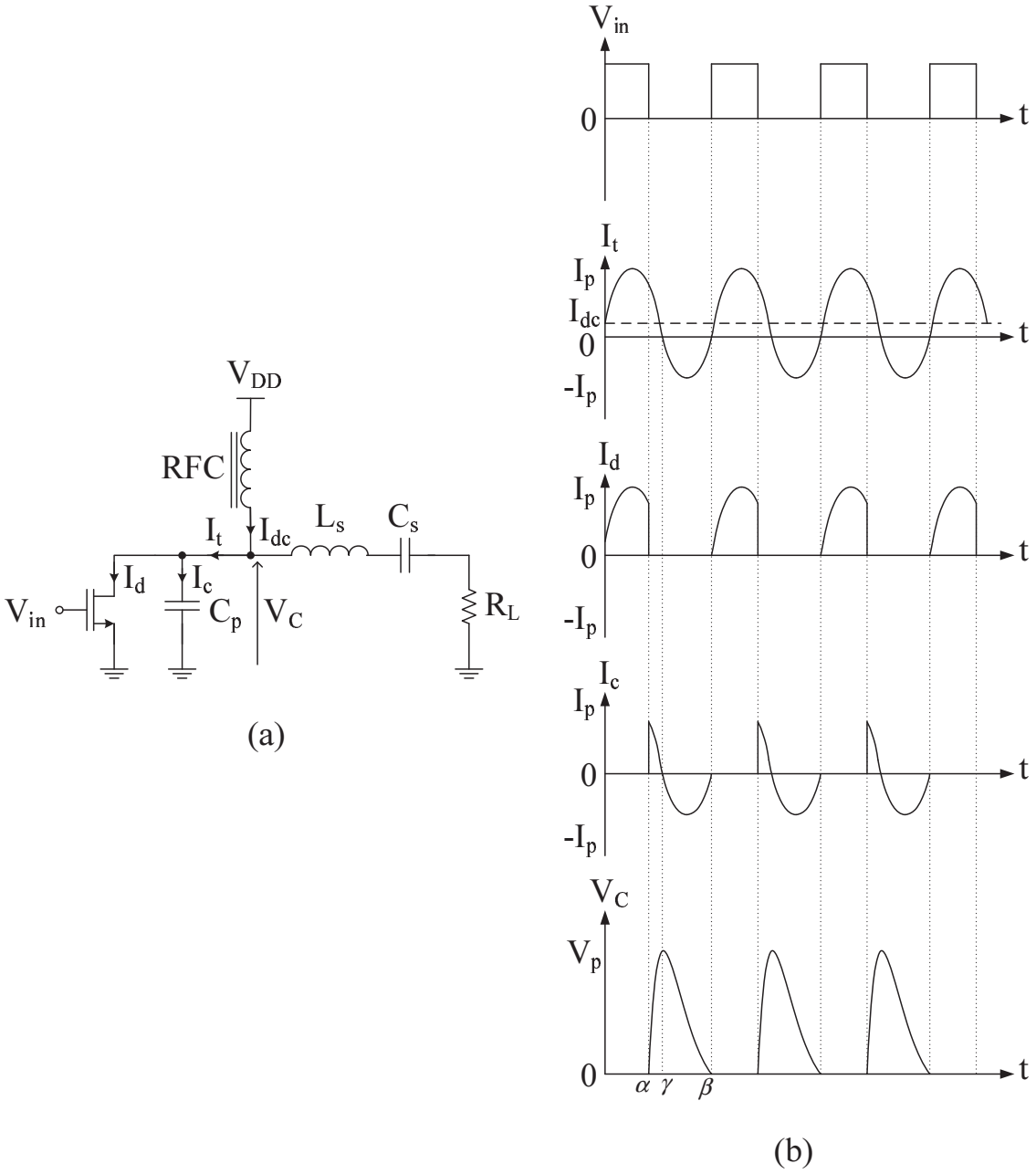


Figure 4: Class-E operation; (a) schematic diagram of class-E amplifier, (b) voltage and current waveforms of class-E amplifier.

through itself and the current charging the output capacitor (C_p) to create output voltage. The voltage waveform across the C_p is given by the following integral equation

$$v_C(\theta) = \frac{1}{\omega C_p} \int_{2\pi-\beta}^{\theta} i_C(\theta) d\theta \tag{3}$$

The voltage peaks at the zero crossing of capacitor current and is a function of the selected angle (α) and closing time of switch (β). Performances of Class-E amplifier are dependent on these factors and also strongly determined by output matching design. L_s and C_s form a resonant circuit to filter the sine wave output to the load R_L. The C_p is important in determining the peak voltage and the maximum operating frequency for the circuit to operate with 100% efficiency. It has to satisfy the condition that no voltage and current overlap, meaning that the C_p must be fully discharged before the next switching cycle starts. A significant downside of the class-E amplifier is the peak voltage (V_p) that can be as high as 3 - 5 times of the supply voltage. This requires the use of high breakdown transistors or some special treatments such as the stacking of transistors. The detail analysis of efficiency in class-E operation can be found in [49, 59].

Another class of switching amplifier, class-F, has a lot of similarities to the overdriven class-A amplifier. The differences are the input drive that should be as close to the square wave as possible to sharpen the edges of drain current waveform (or half sine wave input as of an overdriven class-B bias) and the output termination where either odd or even harmonics are terminated (see Fig. 5). In this case, the fundamental short circuit and the even-harmonic short circuit will form the square-wave voltage across the drain-source terminal of the transistor. This reduces the overlap between output voltage and current, thus boosting the efficiency up to 100% in principle. Similar results can be achieved if the odd harmonics are filtered out and even harmonics are retained (the drain voltage will have a clipped half-sine waveform

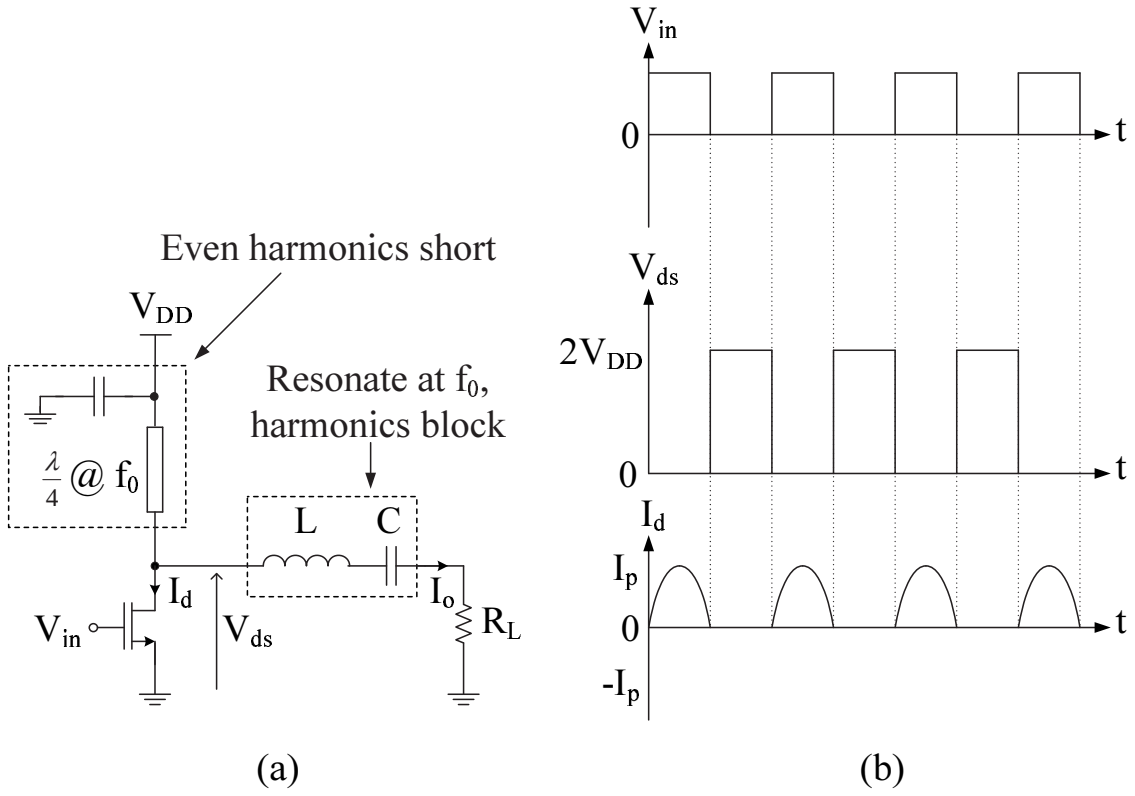


Figure 5: Class-F operation; (a) schematic diagram of class-F amplifier, (b) voltage and current waveforms of class-F amplifier.

instead of a rectangular waveform.) Practically, the quarter-wave transmission line is used to short even harmonics in addition to the main purpose of supplying bias current. The advantage of the class-F amplifier over the class-E amplifier is the maximum drain voltage that is just twice the supply voltage. Further details of class-F design are available in [52].

2.2 Efficiency Enhancement Techniques for RF Power Amplifiers

Conventional linear power amplification has low efficiency at low drive power while having signal distortion at the power level close to the maximum output power. The low efficiency issue can be solved by using efficiency enhancement schemes developed since 1930s and used widely in high-power tube amplifiers. In the case of high-power tube amplifiers, the low efficiency problem leads to excessive heat dissipation at the point where the input power is backed off from the maximum level. This is because the dc power consumption is fixed, while the output power decreases. The consequences are the increase in maintenance cost for cooling systems and electrical usage. Nowadays, not only the low efficiency problem still remains in high-power, solid-state amplifiers for base stations but it has also become a critical issue in low-power linear RF power amplifiers for personal and mobile communications, where the dc power source is limited.

To date, several efficiency enhancement techniques have been developed with different approaches and levels of complexity. Depending on the target application, some techniques may be more suitable than others and sometimes they can be combined for improved results. There are a number of commonly used techniques: envelope elimination and restoration (EER) [27], bias adaptation technique, Doherty amplifier [18], and Chireix's outphasing technique [13]. These techniques are briefly explained below together with their benefits and drawbacks.

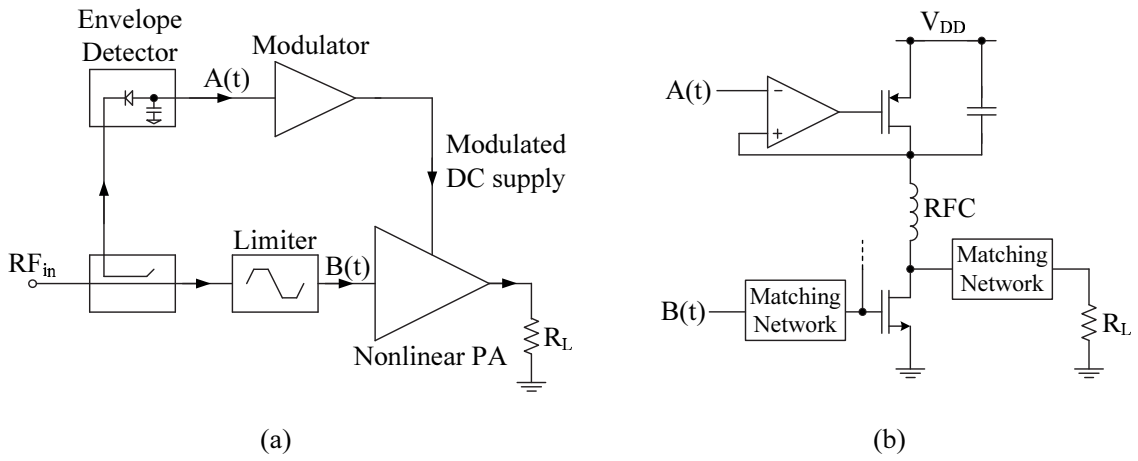


Figure 6: Envelope Elimination and Restoration system; (a) system configuration, (b) implementation of the output stage by low-frequency feedback amplifier.

2.2.1 Envelope Elimination and Restoration

The envelope elimination and restoration (EER) uses the benefit of high efficiency in a nonlinear amplifier to build a linear amplifier system as illustrated in Fig. 6. Input signal is passed through a limiter before amplifying with high-efficiency nonlinear PA. The dc supply of the nonlinear PA is modulated with the envelope amplitude of the input signal. The limiter is used to minimize the chance of AM-PM distortion which may occur if the input signal to the nonlinear amplifier is too high. The output signal in this case will have the amplitude proportional to the dc supply voltage with the phase characteristic of the input signal.

However, there could be many problems in realizing this technique. First, the phase and gain mismatch between two paths must be minimized. This is difficult to obtain from different circuits operated at different frequencies. Second, the dc controller that generates control current and voltage for the PA may not operate at 100% efficiency and has limited bandwidth which may not be adequate for multichannel signal. Other problems are the limiter circuit and nonlinear capacitance of the PA device that, when operated at large-signal conditions, may introduce undesirable

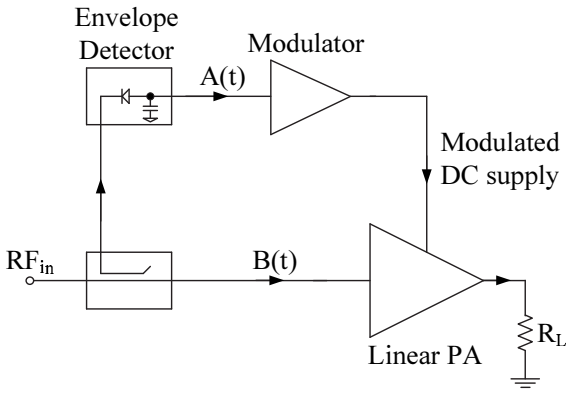


Figure 7: Bias adaptation system.

AM-PM distortions.

2.2.2 Bias Adaptation

Bias adaptation scheme is similar to the EER scheme described previously. However, the limiter circuit is unnecessary, as shown in Fig. 7. Instead of using a nonlinear amplifier, the transistor is operated in a linear condition where the dc supply voltage is modulated with the input signal envelope. The difference between this technique and EER is the RF input signal that contains both amplitude and phase information. The supply voltage control has more flexibility than the EER technique because it does not have to perfectly match with the input envelope, which allows more errors and design relaxation, with a trade-off in lower drain efficiency. Since the amplifier is biased in linear operation (class A or AB), one of the problems is the transistor performance degradation such as the reduction of gain when operated in low-current conditions. Gain reduction is usually observed when the current density of a device is lowered from the optimum value, which consequently reduces PAE. Also, the design of a highly-efficient dc modulator with high output voltage and current is a challenge. Nevertheless, this technique is more attractive because it is simpler and more practical than EER, and has already been implemented in several RF applications in today's communication standards [22, 23].

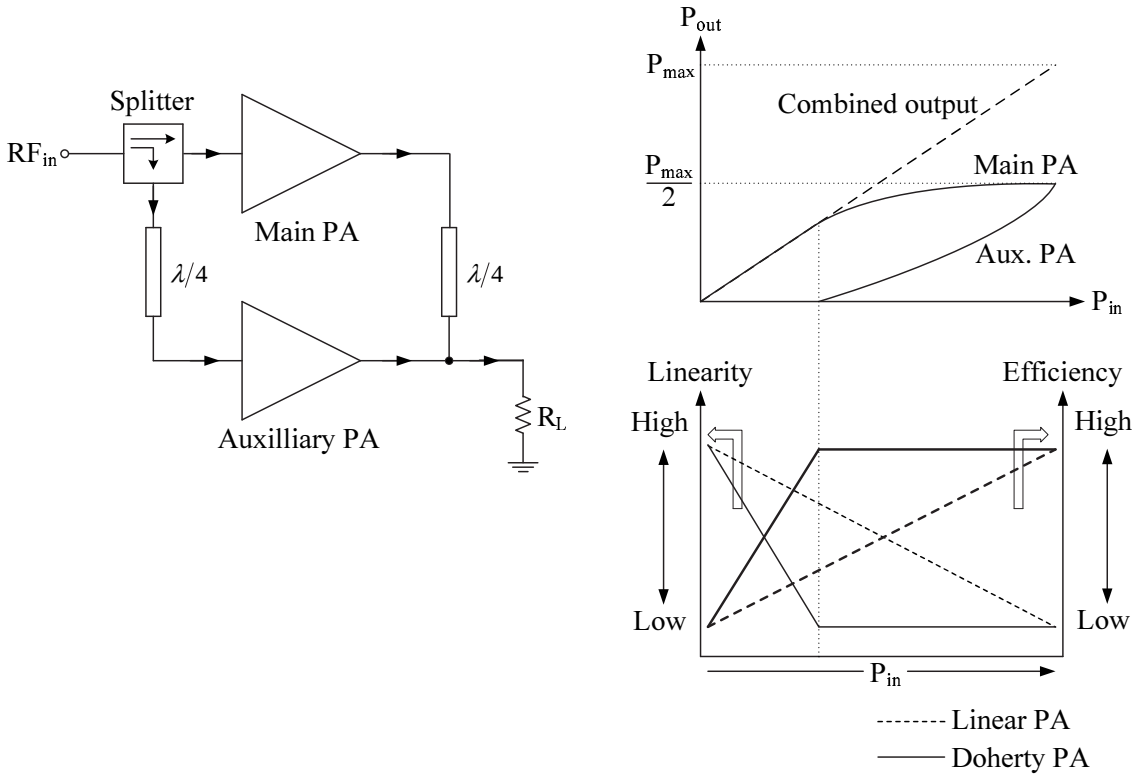


Figure 8: Doherty amplifier system configuration.

2.2.3 Doherty Amplification Technique

The Doherty amplifier is one of the very first techniques that adopts the idea of a multiple-transistor approach to solve the low-efficiency problem of linear RF amplifiers in low-output power region. Linear operation such as class A or AB has nearly-fixed dc power consumption which is not proportional to the RF output power and therefore becomes inefficient at low power. The Doherty amplifier uses more than one amplifier to operate at different power levels. Each amplifier is biased into different bias conditions and designed to have different load terminations so that it can be optimized for multiple power levels. The conventional Doherty amplifier design uses two amplifiers to compromise between efficiency and linearity in low power and high power regions (see Fig. 8).

The operation of Doherty amplifier does not require the manual switching between

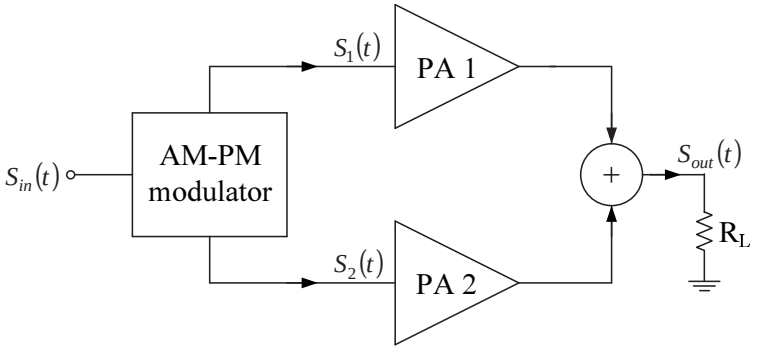


Figure 9: Chireix's Outphasing amplifier system.

each amplifier. Its operation is actually controlled by the input power level, without any use of external control or adjustment. This is the critical design requirement of the Doherty amplifier, which makes it attractive to circuit designers. In addition, the output termination of each amplifier is controlled by the loading of output power from another amplifier and the transformation of impedance over a quarter-wave transmission line. The output termination will be automatically adjusted to maximize the efficiency for a wide range of power levels. The details of the Doherty amplifier design are described in Chapter 3.

2.2.4 Chireix's Outphasing Technique

This technique is sometimes called the linear amplification using nonlinear components (LINC) because it uses two nonlinear amplifiers to amplify two input signals with different phases, which are finally combined at the output to regain an amplitude and phase modulated signal, as illustrated in Fig. 9. The concept of this technique was developed in 1935 by Chireix [13]. The idea initiates from a trigonometric relationship:

$$cos(A) + cos(B) = 2cos\left(\frac{A+B}{2}\right)cos\left(\frac{A-B}{2}\right) \tag{4}$$

in that if the input signals given to each amplifier are

$$S_1(t) = cos(\omega t + cos^{-1}[A(t)])$$
$$S_2(t) = cos(\omega t - cos^{-1}[A(t)]) \tag{5}$$

which are phase-modulated signals generated from the initial input signal S_{in}, where $S_{in}(t) = A(t)cos(\omega t)$; $A(t)$ is the input amplitude. If the gain of each amplifier is G, the output signal after the recombination can be written, using (4) and (5), as

$$\begin{aligned} S_{out}(t) &= G\{S_1(t) + S_2(t)\} \\ &= 2GA(t)cos(\omega t) \end{aligned} \tag{6}$$

which is the amplitude-modulated input signal that is amplified by a factor of $2G$. If the initial input signal is also phase-modulated, by the relationship of (4), the phase information will be restored at the output without any change. High-efficiency nonlinear amplifiers can be used here as long as their output performances are identical.

The important component in this technique is the AM-PM modulator that is required to generate signal in (5), where the input signal amplitude is transformed into phase deviation. More importantly, the output design of this technique is critical in that a simple signal combiner cannot be used at the output since the signal from each amplifier is not synchronous in phase. This, however, has been solved by a reactance-compensation load design technique that further results in improved efficiency in the back-off region, which is actually the key of this technique. In-depth details of the load design and other practical issues are available in [14, 50, 77].

2.3 Basics of RF Power Amplifiers Design

2.3.1 Transistor Modeling

Device modeling is extremely important to the success in circuit designing. A critical step in the circuit design is how to ensure a close agreement between circuit simulation and actual measurement results. The way to achieve this goal is through the use of accurate device model. In device modeling, there are many concerns ranging from

device physic behaviors up to the electrical interactions between interconnections. In the view of device physics, there are important first-order effects that include the behaviors of charge and field in p-n junctions and second-order effects such as temperature related issues. Modeling of interconnection includes electrical characteristics occurring outside of the p-n junctions to the connecting metal layers and bondwires, which can be very complicate at extremely high frequency.

Transistor modeling for RF power amplifiers is a challenging task since the large-signal behavior is difficult to be modeled accurately. Moreover, transistor device usually operates up to the the maximum limitation of operation between the device breakdown and the triode region, which further aggravates the modeling errors because those regions of operation are more difficult to predict and model than the normal operation region. Therefore, in many cases the large-signal models are developed based-on the curve-fitting of suitable measurement results including dc IV-characteristic, pulse IV measurements, and multi-biased S-parameter parameter extractions. This is sometimes called empirical device modeling as opposed to physical device modeling that models the transistor by physic-based charge and field equations. The empirical device model is used widely for RF large-signal applications [1, 9, 57].

Nevertheless, one benefit of a physical model is that each parameter is physically meaningful, which is important for the understanding of device behavior and easier for identifying model errors. Moreover, it can be helpful for future development of device of the same technology. Therefore, sometimes the benefit of physical model are combined with the accuracy of empirical model. This creates a semi-empirical model that combines the physical insights with some empirical parameters. In some applications, the table-based or look-up model, which is developed experimentally, can be used. However, such model contains parameter values characterized within a specific operation region and thus having a limited use and lack of scalability.

The choice of model development depends on the application it is intended for. Basically, it is a trade-off between model accuracy, the application, and the complexity. In general, physic-based models are suitable in many applications but may not be as accurate as empirical models when coming to the RF power amplifier area. Empirical model can be tailored to accurately model the large-signal and nonlinear device characteristics without having to model the transistor noise behaviors, which are generally not necessary for RF power amplifier design, and thus saving tremendous amount of time and effort. In the other hand, modeling the linearity or memory effects of an amplifier may be too complicated for large-signal model and may need the use of table-based model. Therefore, it is important to select the choice of model considering the final application and the efficient flowing of design cycle.

2.3.2 Load-pull Measurement Techniques

In RF power amplifiers, load termination is an essential factor to determine the maximum output power and efficiency. The calculation of optimum load (R_{opt}) can be started from the dc load-line theory according to Fig. 10, which is given by

$$R_{opt} = \frac{V_{dc} - V_{knee}}{I_{dc}} \tag{7}$$

where V_{dc} and I_{dc} are output bias voltage and current, and V_{knee} is the knee voltage. With this R_{opt} the maximum output power (P_{max}) and maximum drain efficiency (η_{max}) are given by

$$P_{max} = \frac{(V_{dc} - V_{knee})I_{dc}}{2} \tag{8}$$

$$\eta_{max} = \frac{P_{max}}{P_{dc}} = \frac{V_{dc} - V_{knee}}{2V_{dc}} \tag{9}$$

If V_{knee} is zero for an ideal case, the η_{max} will be 50%. Therefore, it is clear that the load-line theory is initially for the load calculation of class-A bias since the

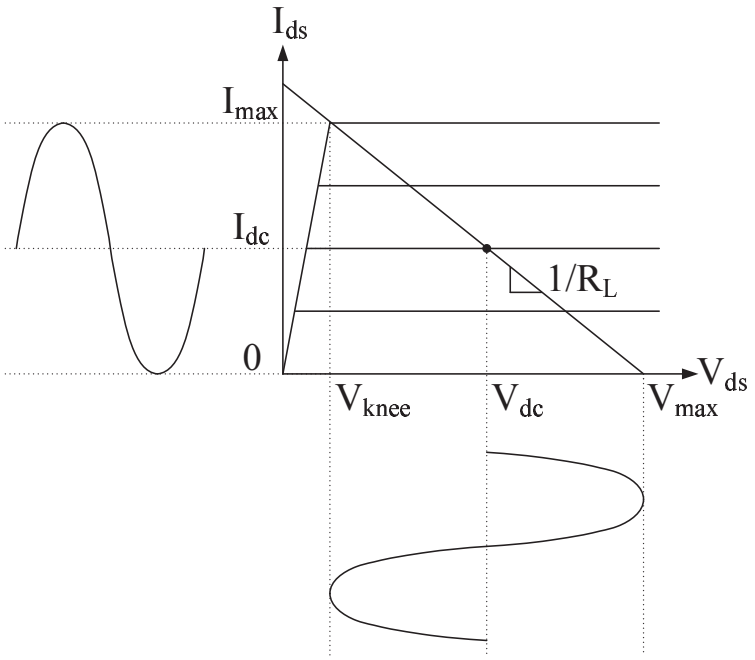


Figure 10: Ideal dc load line for the calculation of optimum load.

maximum efficiency is 50%. The load-line theory assumes the maximum distortionless sinusoidal swing of voltage and current within the range between V_{max} and V_{knee} and between I_{max} and 0. This load line approximates the movement of output voltage and current without considering any transistor performance and characteristic. Once the load or the bias deviates from this calculation, the waveform will be distorted and the amplifier will finally be in other classes of operation.

However, the device actually does have parasitics and non-zero output conductance as well as external parasitic effects from bondwires and package, which eventually alter the load-line contour. Therefore, the load-line theory is used only for getting a rough estimate of the load. A better way, and possibly the most effective way, to achieve an actual value of load termination is from the load-pull measurement.

Load-pull measurement uses impedance tuners to vary the input and output impedances seen by the transistor and measures performances such as reflected and

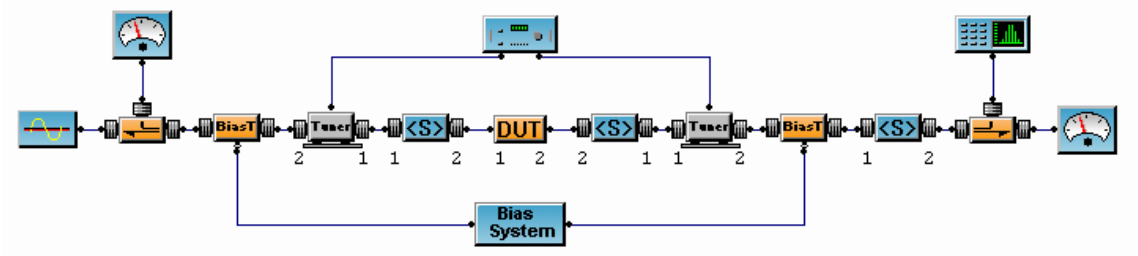


Figure 11: Load-pull system configuration.

delivered powers, output power, dc consumption, power-added efficiency, intermodulation levels, and other parameters. The typical setup of load-pull components is illustrated in Fig. 11. Measured transistor performances are used to construct contours where the impedances of optimum termination with maximum desired performances can be identified at the location around the center of the contours as shown in Fig. 12. The desired performances depend on target application, which could be maximum gain, maximum output power, maximum PAE, or maximum linearity.

Load-pull measurement is an ultimate tool that effectively considers all characteristics of the device and expresses the results from the actual working conditions. The load-pull system normally involves with many equipments and is usually costly and complicate. Even though the load-pull seems to be the best approach for RF power amplifier testing, there are still practical problems, for example, parasitics of the components and biasing networks causing oscillation, which sometimes may not be removed by system calibration. There are also some limitations in the maximum range of reflection coefficient the system can operate up to. Nevertheless, with a careful system setup to minimize parasitics and mismatches, the load-pull measurement is a great tool and still the most favorite approach for RF power amplifier design in both IC and system levels. Some commercial RF simulation software have embedded the load-pulling capability into the circuit simulator, thus enabling the load-pull technique in an economical way. However, this will require an accurate large-signal transistor model. Also, the large-signal envelope simulation is prone to convergence

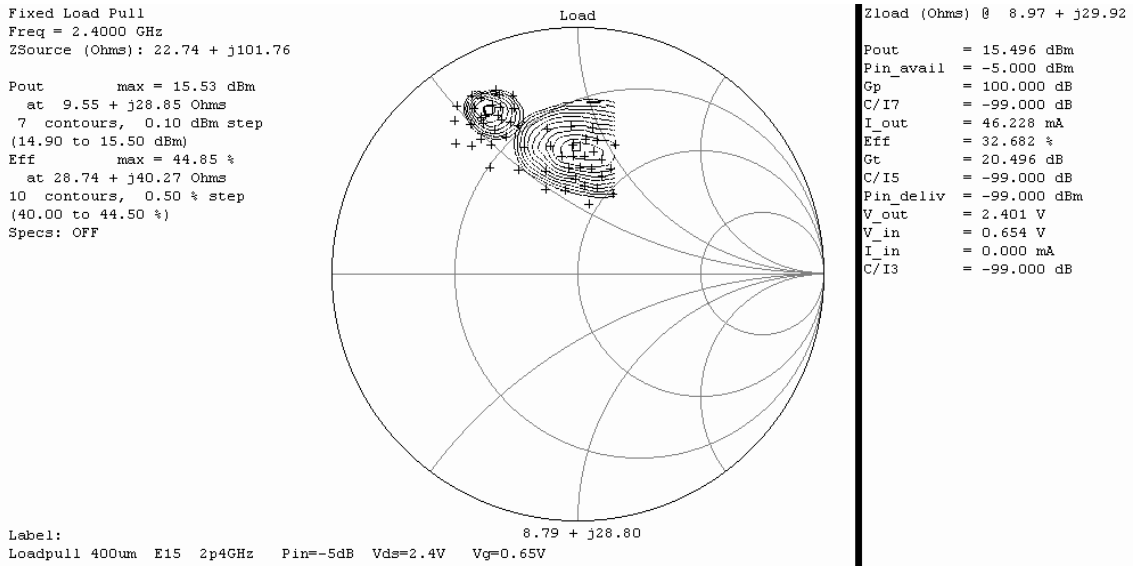


Figure 12: Load-pull measurement result.

problems in the calculation process, which may lead to misleading results. Therefore, it is more effective to use load-pull simulation as a first step before measuring the device in the actual load-pull setup.

2.3.3 Power Amplifier Design Considerations

Designing a power amplifier involves many issues including matching network design, transistor device selection, biasing network, linearity, stability, and so on. If the output power level and the power gain requirement are high, multiple-stage design may be adopted. The choice of device selection is limited to the availability from manufacturers if using off-the-shelf transistors, but not limited for the design in MMIC process. Device size is an important variable for output performances such as output power, power gain, efficiency, and linearity. However, it is difficult to pin point the optimum size for a certain output power since this is dependent on device type (FET, BJT, or HBT) and process (Si, SiGe, GaAs, and etc.), as well as the class of operation. The practical approach is to conduct load-pull measurements on test devices and determines experimentally for certain characteristics. This process will

also give information on bias point and matching conditions that are useful for further designing of bias and matching networks.

The matching network is designed to match the ultimate load or source (50 ohms, for instance) to the optimum impedance at the transistor's input and output. The choice of topology (L-, T-, or π-network) depends on the bandwidth requirements and the possible incorporation of matching components with biasing chokes and dc blocking capacitors. For many times, multiple sections are used to enhance the bandwidth to cover a broad range of communication bands such as those for satellite communications. In personal and mobile communications, small sections of matching network, which offer just enough bandwidth to cover a specific frequency range, are used to minimize the cost and design area.

For multistage power amplifiers, there is always a problem in the design of interstage matching network. To give an example, a two-stage power amplifier has a problem in that the optimum matching impedance for the first stage, obtained from the load-pull simulation or measurement, is not exactly the conjugate match of its output impedance as in the case of the small-signal amplifier. For low-power amplifiers (30 dBm or less), this problem is not substantial as the first stage is assumed to be in a small-signal condition and the conjugate match assumption is acceptable. However, the assumption may be incorrect as the power level increases, and the design may have to rely solely on blind simulations, in which the accurate large-signal model has to play an important role. Therefore, frequently, designing a high-power amplifier is based mainly on brute-force approach and designer experience.

Biasing circuit is also important in controlling bias current level to satisfy output power requirements. For class-AB operation, it also controls the degree of gain expansion and nonlinear cancellation leading to intermodulation sweet-spot, which is a result of the cancellation between signals generated from nonlinear components in

the transistor. Moreover, the input impedance of biasing circuit has a strong influence on the memory effect that generates sideband asymmetry. In general, the bias circuit for FET amplifiers can be implemented using voltage-reference circuit. For bipolar amplifiers, current mirror circuit can be designed for class-A operation. For other classes of operation, it has to be designed so that the base bias current can change rapidly in a wide range to satisfy the dynamic movement of bias point. In addition, a temperature-compensated and process-invariant bias circuit is necessary in the product-level development to ensure the power amplifier performances over a possible range of operation [62].

Another important issue in power amplifier design is the stability. Stability depends mainly on the device itself which includes characteristics such as forward and reverse gain or loss, parasitics, and internal feedbacks. These factors can be expressed through a two-port network using S-parameter. In a small-signal amplifier, the device S-parameter is used mainly to calculate conjugate matches and gain of the circuit. In a large-signal circuit, the device behavior is dynamic and cannot rely on a single set of small-signal parameter. However, S-parameter is still a useful information to determine an estimate of amplifier stability and has been used widely. Stability of the amplifier can be ensured if the following expressions are satisfied [56]:

$$k = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + \Delta^2}{2|S_{21}||S_{12}|} > 1 \tag{10}$$

$$|S_{11}| < 1 \tag{11}$$

$$|S_{22}| < 1 \tag{12}$$

where k is the Rollett stability factor (or k-factor) and $\Delta = S_{11}S_{22} - S_{12}S_{21}$.

These equations were developed with a physical meaning that the amplifier, assuming a conjugate matching condition, must not be connected to a negative resistance where passive matching cannot be realized. The negative resistance termination will lead to unstable operation, which is possible if $k < 1$. In the case of device with $k < 1$, modifications can be made to increase the k such as the insertion of negative feedbacks to reduce device gain. Sometimes it is effective to reduce the device parasitic to enhance k-factor by improved layout design. It is also possible that the k-factor is greater than one in the operating frequency but falls off outside the interested range, which usually happens in the low-frequency range. In the high-frequency range, k-factor is usually greater than one because the device exhibits lower gain. Further design techniques such as low-frequency bypass network, for instance, have to be implemented to ensure the condition of $k > 1$ for every frequency. In the multistage amplifier design, k-factor of individual stages and the combined stage must satisfy the above conditions to guarantee the stability.

Electrostatic discharge (ESD) protection is also an important concern for RF power amplifier design. However, ESD protection circuit usually introduces additional parasitics and occasionally has low impedance compared to load and source terminations. Therefore, the use of ESD protection is limited in RF power amplifier design. Some forms of ESD protection such as a series of reverse-biased diodes may be implemented at the input port where the signal amplitude is small. However, signal level at the output is usually much larger where special designs of ESD protection circuit may be used if necessary. More importantly, the main concern at the power amplifier output is the voltage standing-wave ratio (VSWR), which is introduced by the mismatch of the load. This sometimes arises from the absence or wrong conditions of the antenna. Device may not survive if the VSWR is higher than 10:1, where output voltage exceeds breakdown levels. In this case, devices with higher breakdown voltages such as those developed from GaAs-based processes are more favored than

those from SiGe- or CMOS-based processes. Special circuit implementation or device modification such as the engineering of collector is necessary for Si-based device to sustain a high VSWR, which is a big challenge that still needs further investigations.

CHAPTER 3

MULTISTAGE DOHERTY POWER AMPLIFIERS DESIGN

3.1 Doherty Amplifier Technique

The Doherty amplifier was first introduced in 1936 by W. H. Doherty as a method to improve the efficiency of linear amplifiers in the back-off region for tube amplifiers. The basic idea is to provide high impedance termination to force the amplifier to reach an early saturation once in the back-off region. Then the impedance termination is gradually reduced and allow the increase of output power while keeping the amplifier in the saturation. This can maintain maximum drain efficiency until the amplifier reaches its maximum operation level. The concept requires an adjustable impedance transformation network that can be realized using a quarter-wave transmission line and two amplifiers (called a carrier amplifier and a peak amplifier), as illustrated in Fig. 13.

The operation can be mathematically explained by the impedance transformation characteristic of a quarter-wave transmission line as

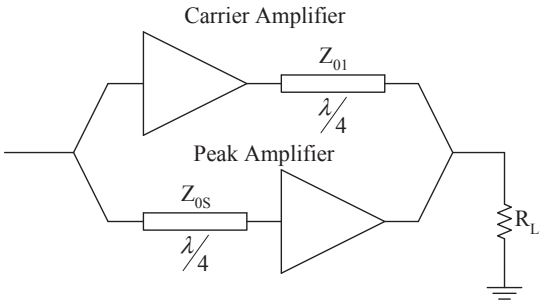


Figure 13: Schematic diagram of the classical Doherty amplifier.

$$Z_{in} = \frac{Z_0^2}{Z_L} \tag{13}$$

where Z_{in} is the input impedance looking into the quarter-wave transmission line, Z_0 is the characteristic impedance of the transmission line, and Z_L is the load impedance connected to another end of the transmission line. In the classical Doherty amplifier (Fig. 13), the transmission line used to perform the Doherty operation (the operation of automatic load-line adjusting) is connected to the output of the carrier amplifier. The quarter-wave transmission line at the input of the peak amplifier is just to compensate for the phase delay between two signal paths, so that the output signal from each amplifier will be added constructively at the load R_L.

The conventional Doherty amplifier has specified the point where the amplifier first saturated to 6 dB below the maximum output power of the system. This means the impedance presented to the carrier amplifier is terminated with $4R_L$ in the low-power region in which the characteristic impedance of $2R_L$ is required for the quarter-wave transmission line. At first, the carrier amplifier, which is biased in class A/AB, is only in operation where the peak amplifier, which is biased below the threshold level (class C), is not yet operating. Once the input level has reached the predetermined level (6 dB back-off), it will turn on the peak amplifier. (To do this, the bias of the peak amplifier needs to be adjusted so that it will promptly turn on at 6-dB back off. However, this may vary from 6 dB in practice.) Consequently, the impedance level presented to the carrier amplifier will be decreased from the turning on of this peak amplifier. With the increased input drive, the impedance transformed to the carrier amplifier will be constantly decreased until the system reaches the maximum power level. At this point, both carrier and peak amplifiers will see a terminating impedance of $2R_L$, while the peak amplifier also reaches its saturation. This implies the same amount of power from each amplifier at the maximum system output power and, since both amplifiers are in saturation, the drain efficiency of the system will

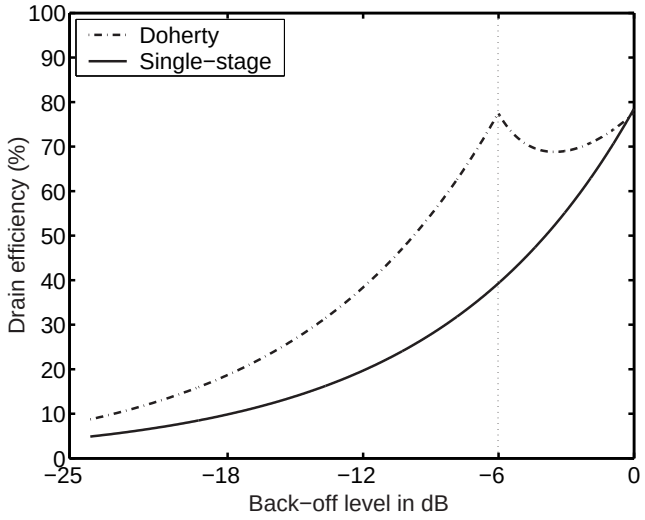


Figure 14: Drain efficiency of the classical Doherty amplifier.

reach the maximum level, as shown in Fig. 14.

3.2 Multistage Doherty Amplifier Technique: Analysis and Design

From the last section, it is shown that the conventional design has an efficiency improvement limitation up to 6 dB from the maximum output power level. For some modern communication standards, the average transmitted output power is in the range of 9-12 dB. Statistically, the power usage profile shows that most of the time the communication equipment operates at a power level far below the maximum power. Therefore, the 6 dB back-off efficiency improvement of the conventional Doherty amplifier is insufficient and results in poor system efficiency. In this section, a method to improve the efficiency of the amplifier in a greater back-off level than a conventional Doherty amplifier is proposed and is called the "multistage Doherty amplifier".

The multistage Doherty amplifier (Fig. 15) uses more than one peak amplifier, with quarter-wave transmission lines to combine their output power. An active load-pulling effect is created as different amplifiers turn on at different power levels. In the design of a multistage Doherty amplifier, it is important to find the characteristic

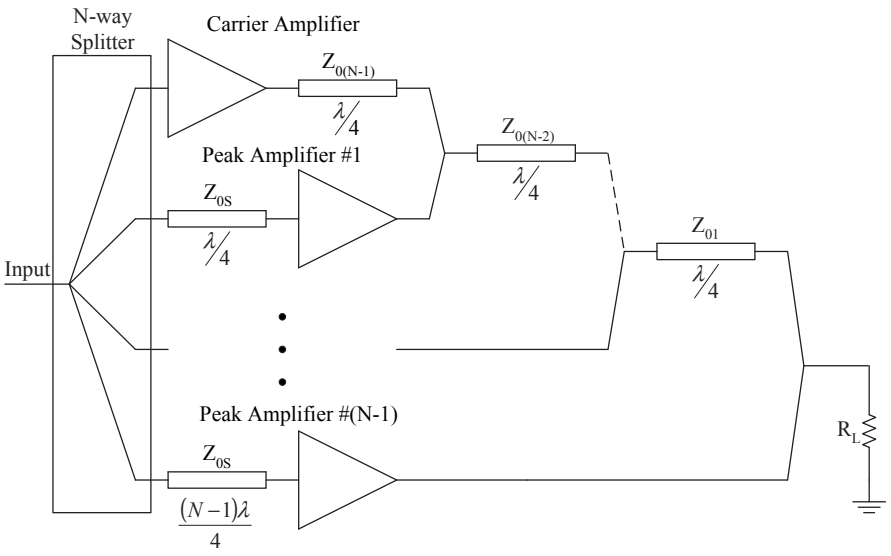


Figure 15: Schematic diagram of the multistage Doherty amplifier.

impedance of the quarter-wave transmission lines required for Doherty operation. These impedance values depend on the level of back off and can be calculated using the following set of equations:

$$Z_{0i} = R_L \cdot \prod_{j=1}^{i} \gamma_j \tag{14}$$

$$\prod_{j=k}^{(i+k)/2} \gamma_{(2j-k)} = 10^{(B_i/20)} \tag{15}$$

where i = 1,2,3,. . .,N-1; $k = 1$ (for odd i) or 2 (for even i); N is the total number of amplifier stages; and B_i is the back-off level (positive value in dB) from the maximum output power of the system at which the efficiency will peak. The maximum level of back off (B_{N-1}) is set by the carrier amplifier. The number of efficiency peaking points is directly proportional to the number of amplifier stages used in the design.

3.2.1 Principle of Operation

To simplify the analysis, the equivalent circuit of the three-stage Doherty amplifier with an ideal current source to represent each amplifier, as shown in Fig. 16, is used

in the derivation. The design equations for an N-stage Doherty amplifier can easily be generalized from this analysis.

From Equations (14) and (15), the characteristic impedance of each output quarter-wave transmission line is found to be

$$Z_{01} = \gamma_1 R_L \tag{16}$$

$$Z_{02} = \gamma_1 \gamma_2 R_L \tag{17}$$

where

$$\gamma_1 = 10^{(B_1/20)} \tag{18}$$

$$\gamma_2 = 10^{(B_2/20)} \tag{19}$$

The phase of the peak amplifier output currents, i_{P1} and i_{P2}, must lag that of the carrier amplifier output current, i_{C1}, by 90 and 180 degrees, respectively, for proper Doherty amplifier operation. This is achieved by inserting additional transmission lines at the input of each peak amplifier. By doing so, the signal from each amplifier will be added constructively at the output load.

The operation of a three-stage Doherty amplifier can be separated into three regions: low-power operation, where only the carrier amplifier is turned on; medium-power operation where the carrier amplifier and peak amplifier #1 are both turned on; and high-power operation where all the power amplifiers are turned on. At low-power operation (Fig. 16(a)), where all the peak amplifiers are in the off state and appear as open circuits, the carrier amplifier will see an impedance given by

$$R_{C1}^{l} = \frac{Z_{02}^2}{R_{C2}^{l}} = \gamma_2^2 R_L \tag{20}$$

where

$$R_{C2}^{l} = \frac{Z_{01}^2}{R_L} = \gamma_1^2 R_L \tag{21}$$

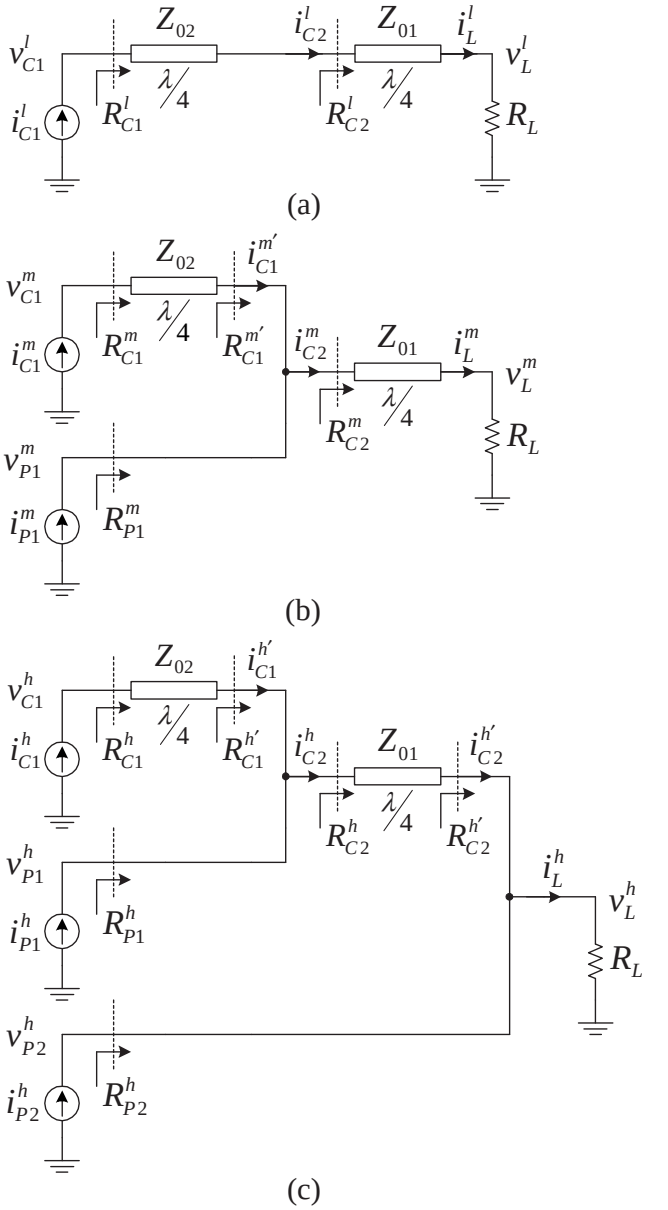


Figure 16: Equivalent circuit of the three-stage Doherty amplifier at (a) low-power operation, (b) medium-power operation, (c) high-power operation.

Assuming the system supply voltage is V_{DD}, the maximum output power at the load R_L, when the carrier amplifier is in saturation, is

$$P_L = \frac{V_{DD}^2}{2\gamma_2^2 R_L} \tag{22}$$

For the case of the three-stage Doherty amplifier, with peak efficiency at 6 and 12 dB back off ($B_1 = 6$ and $B_2 = 12$), the values of γ_1 and γ_2 are 2 and 4, respectively. At this point, the maximum power from the carrier amplifier will be 1/16 of the maximum possible system power. Since the other amplifiers are still turned off and the carrier amplifier is in saturation, the entire output power will be delivered only from this amplifier and therefore the overall efficiency is equal to the maximum efficiency of the carrier amplifier.

In the medium-power operation region (Fig. 16(b)), the impedances presented at the output of each amplifier can be analyzed using power conservation analysis. It can be considered that the output power delivered to the load R_L (in Fig. 16(b)) is equal to the combination of the output power delivered from the carrier amplifier and the peak amplifier #1, which can be expressed as

$$\frac{(v_L^m)^2}{R_L} = \frac{(v_{P1}^m)^2}{R_{C2}^m} = \frac{(v_{C1}^m)^2}{R_{C1}^m} + \frac{(v_{P1}^m)^2}{R_{P1}^m} \tag{23}$$

or

$$\frac{a_1^2}{R_{C2}^m} = \frac{1}{R_{C1}^m} + \frac{a_1^2}{R_{P1}^m} \tag{24}$$

where $a_1 = v_{P1}^m / v_{C1}^m$. R_{C2}^m is the parallel combination of $R_{C1}^{m'}$ and R_{P1}^m, which is equal to Z_{01}^2/R_L. Since $R_{C1}^{m'}$ is equal to Z_{02}^2/R_{C1}^m, the expression for R_{C2}^m can be written as

$$\frac{1}{R_{C2}^m} = \frac{R_{C1}^m}{Z_{02}^2} + \frac{1}{R_{P1}^m} \tag{25}$$

Using (24) and (25), R_{C1}^m and R_{P1}^m can be derived as

$$R_{C1}^{m} = \frac{Z_{02}}{a_1} \tag{26}$$

$$R_{P1}^{m} = \frac{a_1 Z_{02} R_{C2}^{m}}{a_1 Z_{02} - R_{C2}^{m}} \tag{27}$$

From (26) and (27), it can be written as

$$R_{C1}^{m} = \frac{\gamma_1 \gamma_2}{a_1} R_L \tag{28}$$

$$R_{P1}^{m} = \frac{a_1 \gamma_1^2 \gamma_2}{a_1 \gamma_2 - \gamma_1} R_L \tag{29}$$

where $a_1 = v_{P1}^{m}/v_{C1}^{m}$, which has a value between γ_1/γ_2 (when only the carrier amplifier saturates) and 1 (when both carrier and peak amplifier #1 saturate), assuming every amplifier uses the same drain bias voltage of V_{DD}. For the case of the three-stage Doherty amplifier, with $\gamma_1 = 2$ and $\gamma_2 = 4$, R_{C1}^{m} will reduce from $16R_L$ to $8R_L$ at the saturation of peak amplifier #1 because of the load-pulling effect resulting from the turning-on of peak amplifier #1. Also, peak amplifier #1 is presented with a transformed impedance R_{P1}^{m} of $8R_L$. Therefore, the total output power, which is delivered equally from both amplifiers, will be 1/4 of the maximum output power (i.e., 6 dB back off) and will result in an efficiency peak (equal to the maximum efficiency) since both amplifiers are in saturation.

In the high-power region (Fig. 16(c)), peak amplifier #2 is in operation and produces a load-pulling effect on the other amplifiers. An analysis of the terminating impedance to the output of each amplifier can be done similarly to the medium power case, resulting in the following design equations:

$$R_{C1}^{h} = \gamma_1 \gamma_2 R_L \tag{30}$$

$$R_{P1}^{h} = \frac{\gamma_1 \gamma_2}{a_2 \gamma_2 - 1} R_L \tag{31}$$

$$R_{P2}^{h} = \frac{a_2 \gamma_1}{a_2 \gamma_1 - 1} R_L \tag{32}$$

where $a_2 = v_{P2}^h / v_{P1}^h$, which has a value between $1/\gamma_1$ (when peak amplifier #1 saturates) and 1 (when all amplifiers saturate). For the three-stage Doherty amplifier with $\gamma_1 = 2$ and $\gamma_2 = 4$, R_{P1}^h decreases from $8R_L$ to $8R_L/3$ when peak amplifier #2 has reached saturation, while the R_{C1}^h remains unchanged. At saturation, R_{P2}^h is equal to $2R_L$, which enables half the total system output power to be delivered from peak amplifier #2. The rest of the output power is delivered from the carrier amplifier and peak amplifier #1 in the ratio of 1:3.

It is suggested that the value of R_L should be determined by considering the amount of power delivered to the load R_L from the peak amplifier stage #(N-1), since this peak amplifier will contribute more power compared to other stages. Using (32), peak amplifier #(N-1) will be terminated with the optimum matching condition when the value of R_L is calculated by

$$R_L = \frac{\gamma_1 - 1}{\gamma_1} R_{opt\_P\#(N-1)} \tag{33}$$

where $R_{opt\_P\#(N-1)}$ is the optimum termination of peak amplifier #(N-1).

3.2.2 Efficiency Calculation

The efficiency of the three-stage Doherty amplifier can be calculated using power conservation analysis. For low-power operation (Fig. 16(a)), using (20) and the relationship of the carrier amplifier dc current (I_{C1}^l) and voltage across the load (v_L^l) given by [51]

$$I_{C1}^l = \frac{v_L^l}{\gamma_2 R_L} \tag{34}$$

the ideal drain efficiency using a class-B amplifier for the carrier amplifier can be formulated as

$$\begin{aligned}\eta^l &= \frac{\dfrac{(v_L^l)^2}{2R_L}}{(2/\pi)I_{C1}^l V_{DD}} \\ &= \frac{\pi}{4}\frac{\gamma_2 v_L^l}{V_{DD}}, \qquad 0 \le v_L^l \le \frac{V_{DD}}{\gamma_2}\end{aligned} \tag{35}$$

This implies a peak efficiency at $v_L^l = V_{DD}/\gamma_2$ or $P_L = 1/\gamma_2^2$ of the maximum output power of the system ($\eta^l = 78.5\%$ for a class-B amplifier). For medium-power operation (Fig. 16(b)), dc current consumption of the carrier amplifier and peak amplifier #1 can be calculated as

$$I_{C1}^m = \frac{v_{P1}^m}{\gamma_1\gamma_2 R_L} = \frac{v_L^m}{\gamma_2 R_L} \tag{36}$$

$$I_{P1}^m = I_{C2}^m - I_{C1}^{m'} = \frac{v_L^m}{\gamma_1 R_L} - \frac{V_{DD}}{\gamma_1\gamma_2 R_L} \tag{37}$$

Therefore, for class-B carrier and peak amplifiers, the drain efficiency can be formulated as

$$\begin{aligned}\eta^m &= \frac{\dfrac{(v_L^m)^2}{2R_L}}{(2/\pi)(I_{C1}^m + I_{P1}^m)V_{DD}} \\ &= \frac{\pi}{4}\frac{\gamma_1\gamma_2\left(\dfrac{v_L^m}{V_{DD}}\right)^2}{(\gamma_1+\gamma_2)\left(\dfrac{v_L^m}{V_{DD}}\right) - 1}, \quad \frac{V_{DD}}{\gamma_2} < v_L^m \le \frac{V_{DD}}{\gamma_1}\end{aligned} \tag{38}$$

From (38), there exists another peak in efficiency at $v_L^m = V_{DD}/\gamma_1$ or $P_L = 1/\gamma_1^2$ of the maximum output power of the system. Similarly, for high-power operation (Fig. 16(c)), the carrier and peak amplifier dc currents are derived as

$$I_{C1}^h = \frac{V_{DD}}{\gamma_1\gamma_2 R_L} \tag{39}$$

$$I_{P1}^h = I_{C2}^h - I_{C1}^{h'} = \frac{v_L^h}{\gamma_1 R_L} - \frac{V_{DD}}{\gamma_1\gamma_2 R_L} \tag{40}$$

$$I_{P2}^h = I_L^h - I_{C2}^{h'} = \frac{v_L^h}{R_L} - \frac{V_{DD}}{\gamma_1 R_L} \tag{41}$$

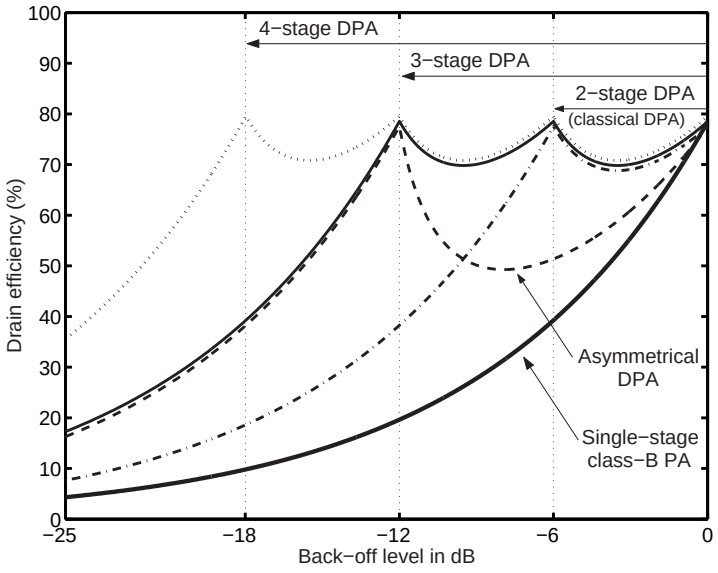


Figure 17: Drain efficiency of the multistage Doherty power amplifier (DPA) using class-B amplifiers ($\gamma_1 = 2$, $\gamma_2 = 4$, $\gamma_3 = 4$; $\gamma = 4$ for asymmetrical DPA [25]).

Finally, the drain efficiency is given by

$$\begin{aligned} \eta^h &= \frac{\dfrac{(v_L^h)^2}{2R_L}}{(2/\pi)(I_{C1}^h + I_{P1}^h + I_{P2}^h)V_{DD}} \\ &= \frac{\pi}{4}\frac{\gamma_1\Big(\dfrac{v_L^h}{V_{DD}}\Big)^2}{(\gamma_1+1)\Big(\dfrac{v_L^h}{V_{DD}}\Big) - 1}, \quad \frac{V_{DD}}{\gamma_1} < v_L^h \leq V_{DD} \end{aligned} \tag{42}$$

Equation (42) shows a peak in efficiency at $v_L^h = V_{DD}$, which is at maximum system power. For the general case of the N-stage Doherty amplifier, the drain efficiency can be calculated similarly to the above analysis. The simulated drain efficiency of the multistage Doherty amplifier using class-B amplifiers is illustrated in Fig. 17.

3.3 Practical Considerations in Multistage Doherty Amplifier Design

In the Doherty amplifier, peak amplifiers are biased below the threshold so that they will not be turned on before the input power has reached a predetermined level. This means the operation of peak amplifiers has been forced into class C, which is

known to typically have lower gain compared to the class-A/AB operation of the carrier amplifier. Therefore, to achieve the required output power in each operation region, two factors can be controlled: the input power to each amplifier and the size of each transistor, which is related to current gain. It is possible to provide greater input power to the peak amplifiers, which are biased in class C, to achieve the required output power without creating AM-AM distortion. Nevertheless, the unequal division of input power will reduce the amount of output power from other transistors, thus reducing the overall gain. To avoid this, the current gain of each transistor must be adjusted to compensate for the gain reduction in the class-C amplifier with minimum alteration of the input power division. To illustrate this, the three-stage Doherty amplifier described in the last section is used in the following analysis.

To maintain the gain level until maximum power operation, the periphery of the peak amplifiers needs to be increased to compensate for fundamental current reduction (assuming a fixed bias configuration is used). The fundamental component of RF current (I_{fund}) is given by the following expression [14]:

$$I_{fund} = \frac{I_{sat}(\theta - sin\ \theta)}{2\pi(1 - cos(\theta/2))} \tag{43}$$

where I_{sat} is the maximum current swing for class-A operation and θ is the conduction angle.

Since I_{sat} is proportional to the device periphery, using (39), (40), and (41), and assuming that, in a general case, peak amplifier transistors are approaching class-B bias at the peak system operation, the relative device periphery can be calculated to be

$$\text{Device periphery ratio} = 1 : \gamma_2 - 1 : \gamma_2(\gamma_1 - 1) \tag{44}$$

(carrier amplifier : peak amplifier #1 : peak amplifier #2)

In practice, the periphery of the transistor for peak amplifiers may need to be larger

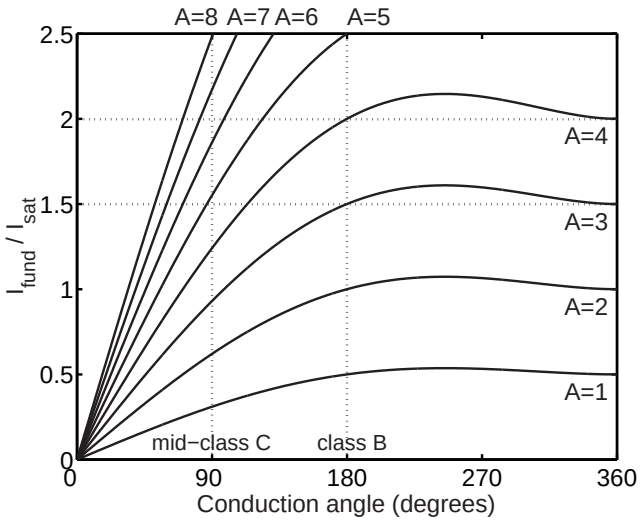


Figure 18: Fundamental current as a function of conduction angle (A = relative device periphery).

than the above calculated values as they may still be in class-C bias at maximum system operation, as shown in Fig. 18. This can be determined experimentally depending on the extent of back-off efficiency improvement and bias point design. Nevertheless, the periphery of the carrier amplifier must be adjusted to provide sufficient output power at the system's maximum operating level and used as a reference for the peak amplifiers. The periphery of the carrier amplifier device can be calculated by

$$DP_C = \left(\prod_{j=1}^{N-1} \gamma_j \right)^{-1} \times DP_{P_L max} \tag{45}$$

where DP_C is the device periphery of the carrier amplifier and $DP_{P_L max}$ is the device periphery of a class-A biased transistor that can deliver the system's maximum power to the load R_L.

3.4 Measurement Results

To verify the analysis, a three-stage Doherty power amplifier with $\gamma_1 = 2$ and $\gamma_2 = 4$ was designed using GaAs FET devices and microstrip-based power combining elements on an FR-4 printed circuit board. The schematic diagram is shown in

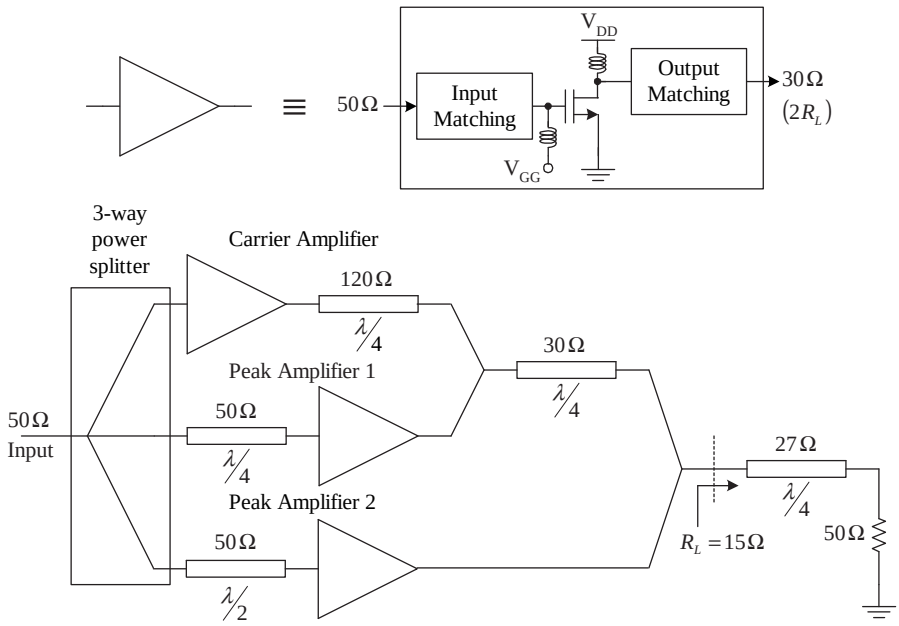


Figure 19: Schematic diagram of the three-stage WCDMA Doherty power amplifier.

Fig. 19, with the actual board layout shown in Fig. 20. The three-stage Doherty power amplifier is targeted for a class-1 WCDMA uplink standard with -33 dBc ACLR1 and -43 dBc ACLR2 linearity requirements with 33 dBm maximum output power. The input signal is divided equally by a microstrip Wilkinson power divider network before feeding to the input matching network of each amplifier. Delay lines of 90 and 180 degrees are inserted at the input of peak amplifiers #1 and #2, respectively. The design is tested with a single-tone signal at 1.95 GHz, the results of which are shown in Fig. 21.

The choice of device periphery is calculated from (44) with the predetermined back-off level improvement, which results in a ratio of 1:3:4. However, because of the limitation of device availability, a device size ratio of 1:2:4 was chosen with the carrier amplifier device W/L being 2400 μm/0.6 μm. The drain bias voltage for all GaAs FET devices in this design is 10 V. The gate bias voltage of the carrier amplifier is set to -1.64 V, which is above the pinch-off voltage of -2 V, for class-AB biasing. The gate bias voltages of peak amplifiers #1 and #2 are adjusted so that the peak amplifiers

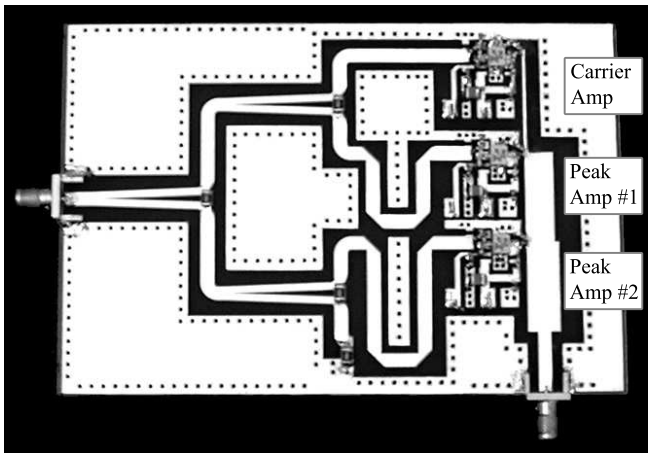


Figure 20: Board layout of the three-stage WCDMA Doherty power amplifier prototype using a device periphery ratio of 1:2:4. Each GaAs FET device has the same package size of 3.8 mm $\times$ 4.2 mm.

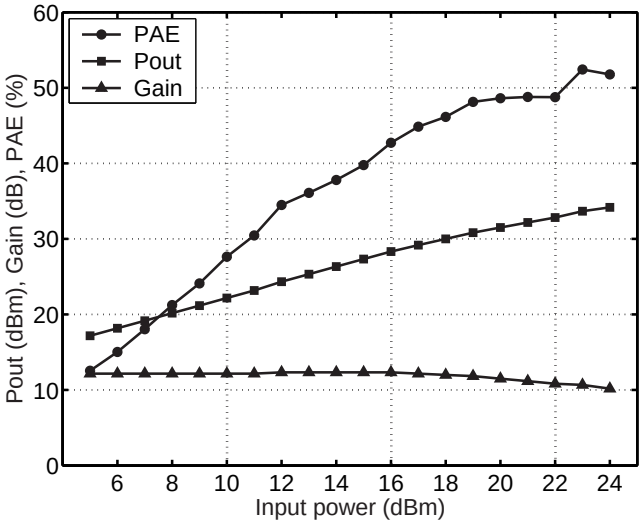


Figure 21: Measured output power, gain, and PAE of the three-stage WCDMA Doherty power amplifier using a device periphery ratio of 1:2:4 at 1.95 GHz.

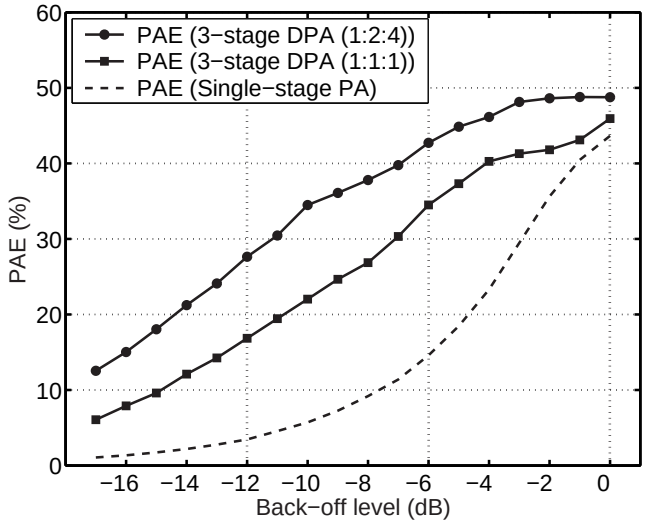


Figure 22: Comparison of PAE measurement results of the three-stage WCDMA Doherty power amplifiers using different device periphery ratios at the frequency of 1.95 GHz (0-dB back off corresponds to output power of 33 dBm).

#1 and #2 are turned on at the backed-off levels of 12 dB and 6 dB, respectively.

The design achieved a linear power gain of 12.2 dB with 1-dB output compression at 33 dBm. The 33-dBm output power level is defined as the 0 dB back-off level in this work. At that point, the power-added efficiency (PAE) was measured to be 48.5%. Also, the PAE was measured to be 42% at 6 dB back off, and 27% at 12 dB back off, which represents a PAE improvement of 2.5 times and 7.5 times, respectively, compared to a single-stage class-AB design. The PAE at 12-dB back off, shown in Fig. 22, is lower than the ideal simulation mainly because of the soft turn-on characteristic of the peak amplifier devices, which contributes to more dc power consumption at the low-power level. Nevertheless, these results show an impressive PAE improvement in the back-off region. In Fig. 22, a 3-stage Doherty power amplifier with $\gamma_1 = 2$ and $\gamma_2 = 4$ using the device periphery ratio of 1:1:1 was designed and measured for giving a comparison. It is seen from Fig. 22 that a lower device periphery ratio (1:1:1) results in poorer PAE improvement at the same back-off level.

The three-stage Doherty power amplifier is tested with a real-time WCDMA 3GPP

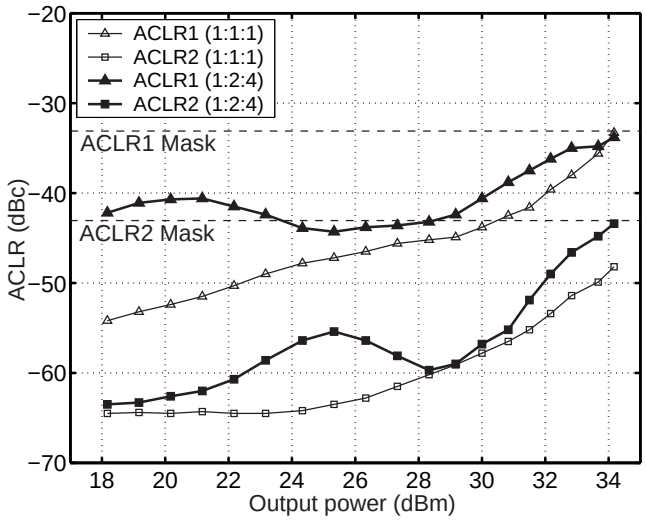


Figure 23: Measured ACLRs of the three-stage WCDMA Doherty power amplifiers using different device periphery ratios at 1.95 GHz.

signal using a chip rate of 3.84 Mcps. The output signal is measured with a raised root cosine (RRC) filter with α of 0.22 and a bandwidth equal to the chip rate. The adjacent channel power leakage ratio (ACLR) measurement results in Fig. 23 show that the design meets the WCDMA ACLR requirements of -33 dBc (ACLR1) and -43 dBc (ACLR2) at 5 MHz and 10 MHz offset, respectively, up to a power output of 34 dBm. The ACLR1 and ACLR2 at the 33-dBm output power are measured to be -35 dBc and -47 dBc, which provide a few dB of margin over the linearity requirement. Moreover, it can be noticed that the ACLR levels of the 3-stage Doherty power amplifier with the device periphery ratio of 1:2:4 are higher than that of the 1:1:1 design since each amplifier stage is operated closer to its saturation, resulting in better overall PAE. The 3-dB bandwidth is measured to be 160 MHz, dominated by the quarter-wave transformer characteristic, with only ± 0.5 dB output power variation from center frequency in the WCDMA uplink frequency range (1.92 - 1.98 GHz), as shown in Fig. 24. The ACLR is still within the specification in this frequency range (Fig. 25).

Thus, the device size ratio of 1:2:4 obtained from ideal calculations results in

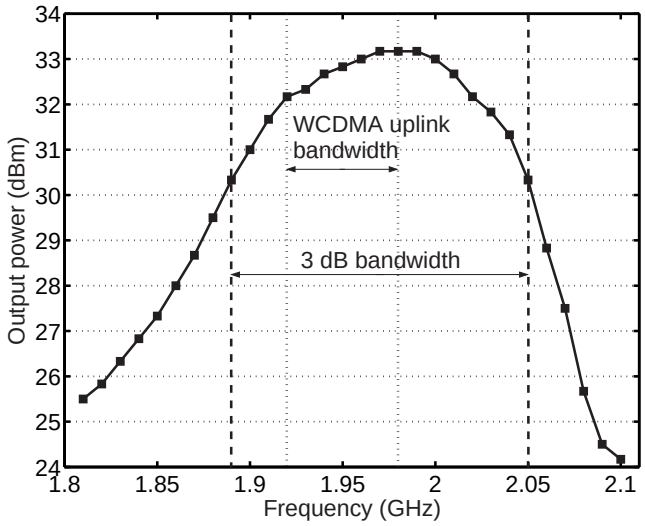


Figure 24: Measured output power of the three-stage WCDMA Doherty power amplifier with a device periphery ratio of 1:2:4 versus frequency at the maximum output power level (33 dBm at 1.95 GHz).

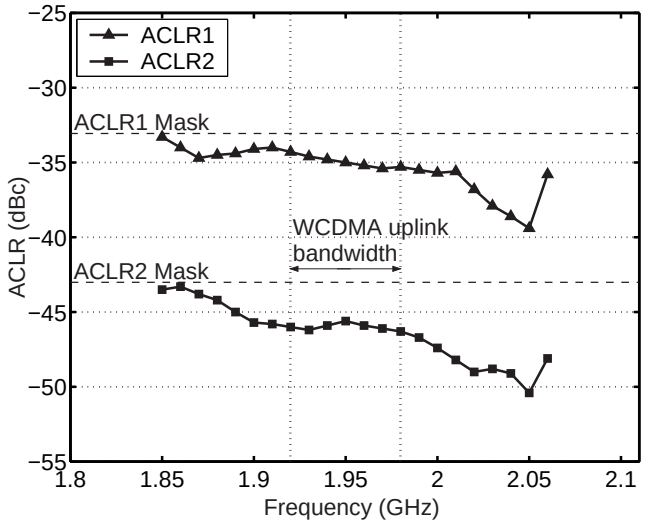


Figure 25: Measured ACLR of the three-stage WCDMA Doherty power amplifier with a device periphery of 1:2:4 versus frequency at the maximum output power level (33 dBm at 1.95 GHz).

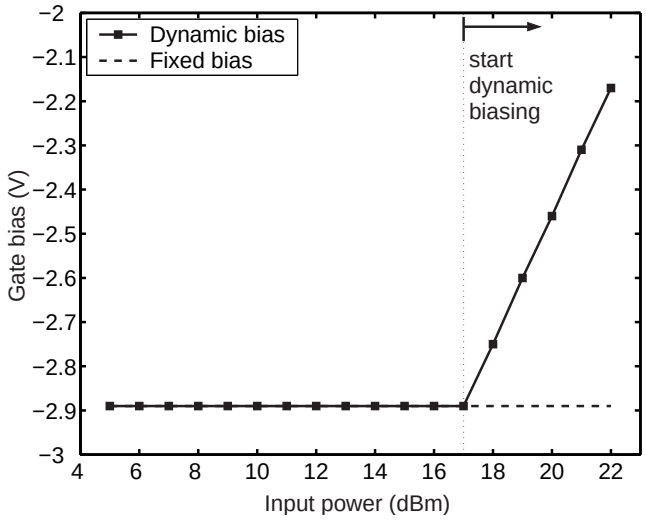


Figure 26: Bias voltage adjustment of the peak amplifier #2 with increasing input drive level to alleviate AM-AM distortion.

a performance sufficient to meet the stringent WCDMA requirements. However, in reality, the peak amplifier devices may not have reached class-B operation, as assumed in the ideal-case calculations. It is possible to further enhance the performance of the Doherty power amplifier by moving the bias point of the peak amplifier from class C in the back-off region toward class B as much as possible by adjusting the gate voltage with increasing input drive. In this design, peak amplifier #2, which was biased in very deep class C, is now biased manually with a dynamic bias profile, as illustrated in Fig. 26.

The measurement results of output power characteristics and linearity of the dynamic-biased three-stage Doherty power amplifier are shown in Fig. 27 and Fig. 28. It is seen that the results have improved in all respects. The ACLR levels at the output power of 33 dBm have reduced from -35 dBc to -40 dBc and -47 dBc to -55 dBc for ACLR1 and ACLR2, respectively. This is expected from the improved AM-AM characteristics of the dynamically biased amplifier compared with the fixed bias amplifier, as observed from the output power curves in Fig. 27. It can be summarized

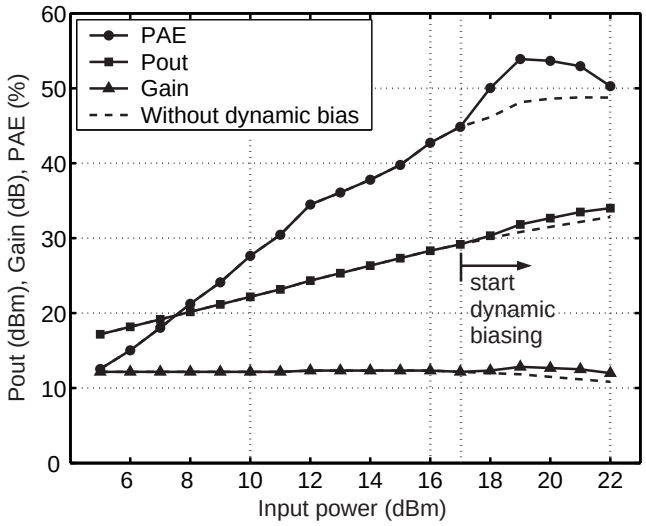


Figure 27: Measured output power, gain, and PAE of the 3-stage WCDMA Doherty power amplifier with dynamic biasing applied to the peak amplifier #2 (at 1.95 GHz).

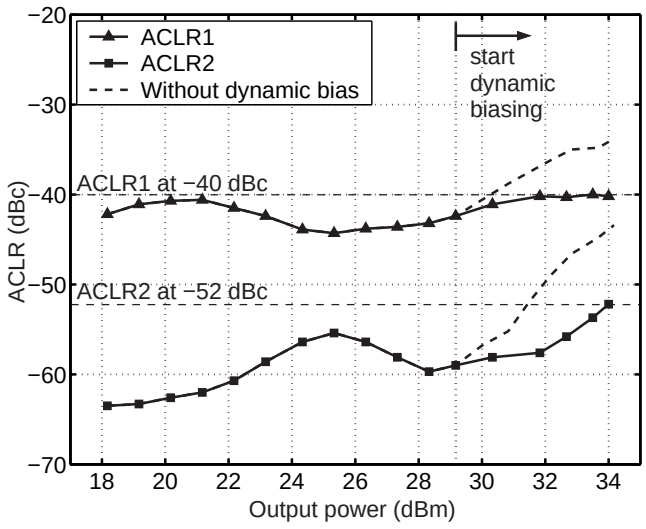


Figure 28: Measured ACLR of the 3-stage WCDMA Doherty power amplifier with dynamic biasing applied to the peak amplifier #2 (at 1.95 GHz).

that with a correct choice of device periphery and appropriate biasing, the performance of the multistage design can be as good as a perfect class A/AB single-stage amplifier at maximum operating power, with improved PAE in the low-power region.

3.5 Summary

This work presents the design and analysis of a three-stage Doherty power amplifier for WCDMA application, which shows a significant improvement in PAE in the low-power region, compared to a single-stage design, while satisfying all the WCDMA requirements. The design equations derived in this work can easily be generalized to the design of an N-stage Doherty amplifier that mitigates efficiency degradation up to even higher output power back-off levels. The analysis of power device selection enables the achievement of near-perfect Doherty amplifier operation. For the design of communication systems where efficiency enhancement is needed, the multistage Doherty amplifier can be an interesting alternative to existing reported techniques.

CHAPTER 4

LARGE-SIGNAL RF MOSFETS MODELING

4.1 MOSFET Device Models

Accurate device model is the key to a successful circuit design. In a transistor modeling, device characteristics such as the behavior of charges and capacitances at dc and higher frequency are important in obtaining accurate device model. There are a number of reported models since the large expansion in development of digital ICs in the late sixties. Today, there are several MOSFET models used in the integrated circuit design industry. Some models are well known and recognized as an industry's standard model. An example of which is the BSIM model. In contrast, some models are well known to a smaller group of users (e.g. MOS Model 9 of Philips [47]). In this work, the most widely used model, the BSIM model, is studied and used as a core model. Even though the BSIM3 is dominant in MOSFET circuit design, it will be shown from a theory and experiment that, still, the BSIM3 needs some necessary improvements to be suitable for CMOS circuit design at RF and microwave frequencies.

4.2 Introduction to Modified-BSIM3v3 Model

The BSIM3 model is an industry standard MOSFET model for deep sub-micrometer applications. The BSIM3 version 3 model (BSIM3v3) has been distributed since 1996 and used world-wide for device modeling and CMOS IC design [17]. It has been validated for digital applications at several hundred megahertz but has not been substantiated in RF and Microwave frequency applications. Recent studies have shown that MOSFET behaves as a distributed device and must be modified for RF

applications. Modifications of device behavior should be made to incorporate the following effects:

1) distributed channel or the non-quasi-static (NQS) effect,

2) distributed gate resistance, and

3) distributed substrate resistance.

4.2.1 NQS Effects

An understanding of quasi-static operation in MOS transistor under large-signal dynamic operation of device intrinsic part is important for analyzing NQS effects. In dc operation, a MOS transistor is driven by four dc voltages, V_D, V_G, V_B, and V_S, defined with respect to ground. The channel of the device is formed by the transport electrons flowing from source terminal to drain terminal (assuming NMOS device) in the inversion layer. Here, the device is assumed to be operated quasi-statically when the charges per unit area in the inversion layer (under the gate conductor and in the oxide surface) and the depletion region are unchanged. This can happen when the variation of terminal voltages is sufficiently slow.

In dc and low-frequency applications, this assumption is valid and the model, which is mostly developed under the quasi-static operation, can be utilized with negligible errors. However, when terminal voltages change more rapidly, such as those in high-frequency operation, the charges flowing in the inversion layer do not have sufficient time to follow the terminal voltages and the quasi-static assumption is no longer valid. Therefore, for high-frequency applications, the model must be developed specifically by considering the non-quasi-static operation.

Non-quasi static (NQS) effects are important for high-frequency CMOS circuit designs, which have been experimented in [58, 68]. There are several ways to account for the NQS effects in MOSFET modeling. The first method is to consider these effects using a solution of the current continuity equation [4, 8, 42, 43, 67, 68, 69, 71].

Basically, in this method, the channel is sectioned into many small pieces, where each can be solved as a single-section in quasi-static manner. However, this approach requires extensive mathematical exercises and thus relatively more complex than other methods. Moreover, this method is mostly valid for long-channel MOSFET devices but not accurate for short-channel devices.

Nevertheless, an easier approach to model the NQS effects by using a single resistor connecting to the gate of the device was proposed in [67]. This resistor introduces a delay to the charge travelling in the channel as described in [24, 41]. This method is more attractive since it does not only model the NQS effects but also simultaneously model the gate-induced thermal noise. In BSIM3v3, the NQS effects are modeled and can be selected by setting the NQSMOD parameter in the BSIM3 model to 1. However, this is only available for BSIM3v3 version 3.2 or later.

4.2.2 Distributed Gate Resistance

The effect of gate resistance is significant in the microwave frequency range. Large gate resistance tends to increase thermal noise and decrease the maximum available gain or S_{21}. In a deep sub-micron salicided CMOS process, the metal vapor is deposited to the gate polysilicon, known as silicided gate. This process dramatically reduces the gate sheet resistance to less than 10 ohms per square. Still, the total gate resistance depends on the physical structure of the device, and the accurate modeling of gate resistance is important for high-frequency circuit simulation. The gate poly has become a distributed R-C circuit at high frequency where the gate resistance can be modeled as a lumped element using transmission line theory given by [55]

$$R_g = \Big(\frac{1}{3}\Big) R_{\square} \Big(\frac{W}{L}\Big) \tag{46}$$

where $R_{\square}$ is the gate sheet resistance and W and L are gate finger width and length, respectively. This equation assumes that the gate is connected from one side.

If both sides of gate poly are connected, factor of 1/3 will become 1/12.

This calculation, however, requires further adjustment for interconnections between each gate poly in multi-finger devices that have not yet been accounted for. Furthermore, this calculation should be verified with the extracted value from measurement result to ensure correctness. For BSIM3v3, the gate resistance must be added externally to increase the model accuracy at microwave frequency. This gate resistance can be included together with the resistance accounted for the NQS effects mentioned earlier, and thus reducing additional components.

4.2.3 Distributed Substrate Resistance

Substrate resistance has significant impacts on device performances at high frequency. In RF CMOS IC design, the non-epitaxial (non-EPI) process is preferred to the epitaxial process (EPI) because of the higher substrate resistance. In a non-EPI CMOS process, a high substrate resistance (6 ohm-cm to 10 ohm-cm) helps to reduce the parasitic capacitance to substrate and the unwanted signal coupling between devices. The substrate resistance of the highly doped EPI process is minute (0.01 ohm-cm to 0.02 ohm-cm), which helps protecting the circuit from a failure mechanism called "latch-up"– the state of having a virtual PNPN device in the substrate that regeneratively draws enormous current from external network, trigged by unexpectedly large current flowing through the substrate such as that resulted from over-stress (e.g. spike) in supply voltage or temperature. This highly doped substrate reduces the current gain of the parasitic PNP and NPN BJTs, which reduces the possibility of latching-up. Even though there is no difference in the modeling of substrate resistance in both processes, this thesis focuses on the modeling of NMOSFET in non-EPI CMOS-9 (0.18 μm) process.

Substrate resistance can be calculated using 2D device simulator such as MEDICI

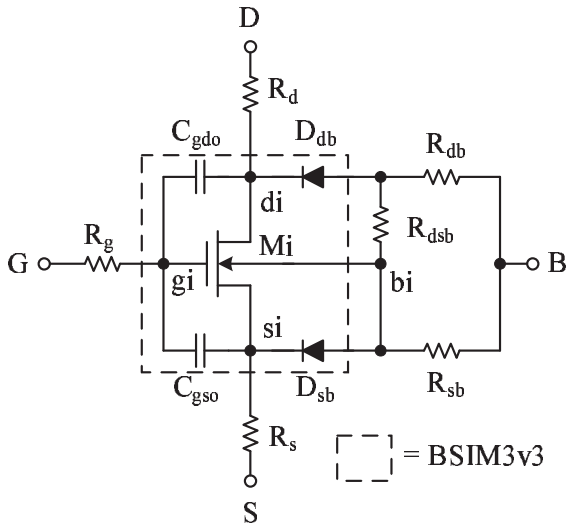
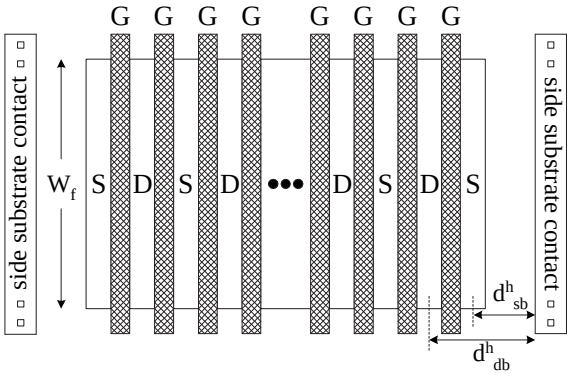


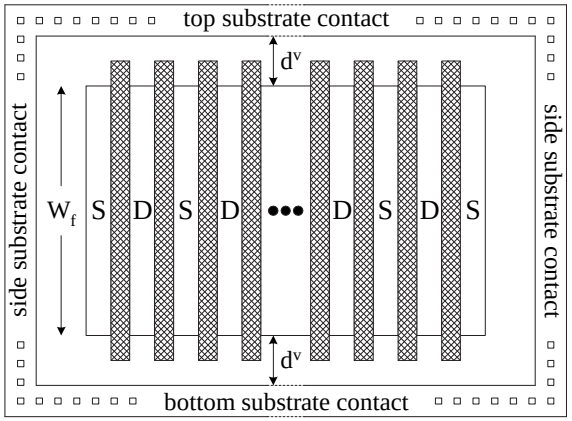
Figure 29: Modified BSIM3v3 MOSFET model with added substrate resistance network.

[66]. However in a real device, substrate coupling effects are distributed to every direction under the P-substrate, and thus difficult to be captured by any kind of physical device simulator. Therefore, the direct extraction of S-parameter (or Y-parameter) is usually considered as one of the best solutions to obtain accurate substrate parameters. In theory, the substrate coupling effects can be modeled as a serial connection of capacitor and resistor. The capacitor is used to model the depletion capacitance from PN junction (N+ drain and source to P-substrate for N-type MOSFET) and the resistor is for the substrate resistance. In the BSIM3v3 model, a diode connected in reverse fashion is used, instead of a fixed capacitor, to model the bias dependence of the depletion capacitor. Still, the BSIM3v3 has not included the substrate resistance in the model yet. As a result, this requires an addition of external resistors from the internal bulk terminal to the external bulk pin, as illustrated in Fig. 29.

Silicon substrate can be modeled as a single resistor for a frequency range up to 10 GHz, and as a parallel R-C circuit for higher frequency [11, 33, 34]. The main advantage of the single-resistor network over the parallel R-C network is the simplicity in parameter extraction. The modeling of substrate can be very complicate, or even



(a)



(b)

Figure 30: Layout of the RF MOSFET device used in substrate modeling, (a) having only side substrate contacts used in [11, 12, 19, 21, 26, 33, 34, 37, 46, 67], (b) having ring-shaped substrate contact used in this work. (d^h and d^v represent the distance from active junctions to the substrate contact in horizontal and vertical directions.)

impossible, if too many unknown parameters exist. It has been proven in [11, 33, 34], that a resistive network is adequate for modeling the silicon substrate up to 20 GHz with negligible errors.

The degree of substrate coupling strongly depends on the device geometry or the distance from the source and drain junctions to the bulk connection. The placing of bulk connection in RF MOS circuit is usually done to minimize the back-gate modulation and latch-up. In both cases, the shortest path between the intrinsic device and substrate connection with minimum substrate resistance is preferred. Therefore, a device is usually surrounded by a ring of substrate contact (sometimes called guard

ring or guard band) with as many substrate connections as possible. However, previous study of substrate parameter extraction has reported results that considered only side substrate contacts [11, 12, 19, 21, 26, 33, 34, 37, 46, 67]. None of the previous work has shown the extraction under the condition of having substrate contacts surrounding the MOS device (see Fig. 30). This is mainly because of the complexity in 3D extraction of distributed parameters, thus limiting accuracy in the actual circuit design where devices are always surrounded by guard rings.

Therefore, to minimize mismatches between the device model and the device used in the real circuit, the objective of this research is to develop a novel scalable MOSFET model including substrate network, where the transistor is completely surrounded by substrate contacts. The detailed methodology and step-by-step extraction procedure of this model are discussed in detail in the following section. More importantly, the scalability of each parameter, which is the key in developing a scalable large-signal model for MOSFET, is explored thoroughly and demonstrated as a state-of-the-art MOSFET model.

4.3 Scalable Modified-BSIM3v3 MOSFET model

4.3.1 Scalability of MOSFET model

Scalability of model parameters is one of the most important issues for any transistor device. In MOSFET, it was reported in [10, 60] that the scalability depends on geometry of device. The small-signal equivalent circuit shown in Fig. 31 is discussed to illustrate the scalability of each parameter. Normally, intrinsic parameters, including gate-source capacitance (C_{gs}), gate-drain capacitance (C_{gd}), small-signal transconductance (g_m), drain-source resistance (r_{ds}), and R_i – the internal resistance to model NQS effects [67] – are linearly scalable with gate geometry (total gate width and each gate finger width) since they are proportional to the underlying channel area. R_g, R_d, and R_s, which are extrinsic parameters (not a part of channel), also

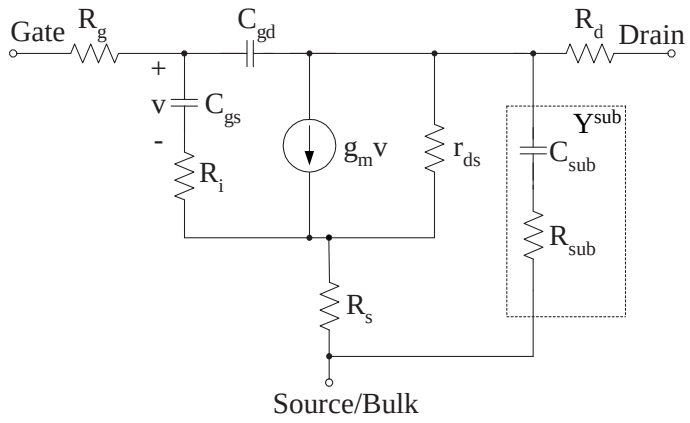


Figure 31: Equivalent circuit of RF MOSFET including substrate network (shown as Y^{sub}.)

linearly scale with gate geometry. The substrate parameters, C_{sub} and R_{sub}, however, do not scale with gate geometry because of their distributed behavior, scattering in every direction from N+ drain and source (for N-MOSFET) to the bulk contacts located on the side and/or top and bottom of the device. Therefore, the modeling of substrate coupling parameters is more complicated than other parameters.

Previously, MOSFET substrate model extraction and its scalability have been studied and reported in [11, 12, 21]. However, the study of substrate parameter scalability has been limited just to a small range of device periphery (up to 500 μm total gate width) and considered only the substrate coupling from drain and source junctions to side body contacts (see Fig. 30(a)). Also, such structure is used in most of the reported MOSFET modeling work [11, 12, 19, 21, 26, 33, 34, 37, 46, 67]. This is mainly due to the increased complexity in parameter extraction of larger devices. However, large devices (up to several millimeters) are required in many applications such as driver amplifiers and power amplifiers, where the accuracy of the parameter extraction and scalability is crucial. Moreover, the ring-shaped substrate contact (see Fig. 30(b)) commonly used for reducing body effect in most RF circuit designs has not been considered in RF MOSFET substrate modeling.

From the initial investigation, it is found that the discrepancy in parameter scalability using the previous model [11, 21] and the new model with ring-shaped substrate contact is due to the fact that the substrate coupling from the drain and source junctions to the top and bottom substrate contacts was neglected. This additional substrate coupling is significant, yet difficult, to be modeled with compact analytical expressions because of its distributed nature throughout the device. Empirical expressions can be used in modeling the substrate parameter scalability as reported previously [61]. However, the use of empirical equations limits the model scalability to certain device geometries used in the modeling process and makes it difficult to extend the study for different device structures.

In this work, the analytical expressions for substrate parameter modeling based on device geometry such as number of gate fingers (N_f), gate finger width (W_f) and length (L_f), and parameters related to the geometry of the ring-shaped substrate contact are proposed. For the first time, the proposed model considers the effect of substrate coupling to the entire ring-shaped substrate contact and approximates it to substrate coupling in the vertical and horizontal directions. It was found that the proposed method can accurately model the scalability of substrate parameters for a large range of device sizes up to 6 mm total gate width. The parameter extraction procedure and scalability modeling are presented in the following sections.

4.3.2 Extraction of Small-Signal Model Parameters

By comparing the small-signal model (Fig. 31) and BSIM3v3 model (Fig. 29), the small-signal model is actually a reduced form of BSIM3v3 when the source and bulk are tied together. Therefore, the extraction of substrate resistance and gate resistance for BSIM3v3 can be done by first extracting the substrate resistance and gate resistance using the small-signal equivalent model. Then, the parameters are incorporated to the BSIM3v3 model and further optimized to fit the measurement results.

Each small-signal model parameter of Fig. 31 can be extracted by analyzing the two-port Y-parameter of the device. First, the S-parameter of N-type MOSFET test devices, with the total gate width ranging from 200 μm to 6 mm, is measured by the HP8510 vector network analyzer. N-type MOSFETs with gate finger width of 20 um were fabricated on a standard five-metal layer CMOS process with gate length of 0.4 um. As commonly used for RF applications, the substrate contact is formed as a ring surrounding the active area, as illustrated in Fig. 30(b) to minimize the back gate effect. The substrate contact is tied to the source terminal which is common for most RF circuits. Multiple gate fingers of the same size were formed to construct a larger FET without changing the layout style in order to minimize the discrepancy of the parasitic associated with the layout. Fig. 32 and 33 show photographs of the test-devices layout used in this work.

The device parameter extraction is similar to the method described in [11, 12] where high frequency extraction up to 40 GHz was performed to find parasitic resistances R_g, R_d, and R_s. Before the Y-parameter calculation of intrinsic parameters can be done, the substrate network Y-parameter needs to be calculated and subtracted out. At zero-biased condition ($V_g = 0$ V; $V_{ds} = 0$ V), substrate parameters can be extracted separately from the intrinsic parameters. The substrate parameter extraction is done through the Y-parameter analysis (in Fig. 31), resulted in the following expressions for R_{sub} and C_{sub}:

$$R_{sub} = \frac{Re\{Y_{22}^c\}}{\left(Im\{Y_{22}^c\} + Im\{Y_{12}^c\}\right)^2} \tag{47}$$

$$C_{sub} = \frac{Im\{Y_{22}^c\} + Im\{Y_{12}^c\}}{\omega} \tag{48}$$

where Y^c is the Y-parameter after de-embedding the pad and interconnect parasitic as well as removing the gate resistance (R_g) and drain contact resistance (R_d). The R_{sub} is a bias independent value which can be readily included in the equivalent

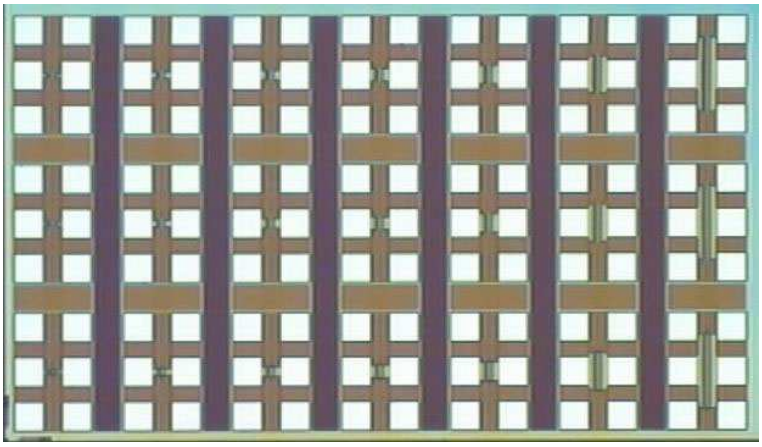


Figure 32: NMOSDG device layout photograph showing test structures having total gate width of 200 μm up to 6000 μm. (The first and second row show shorted- and opened-structures for pad and interconnect parasitics de-embedding. The third row shows the actual devices.)

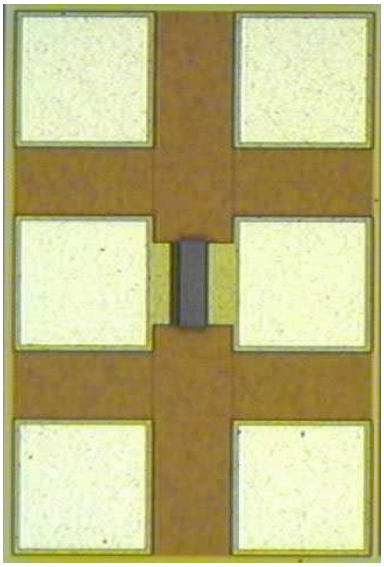


Figure 33: 1200-μm NMOSDG device layout photograph.

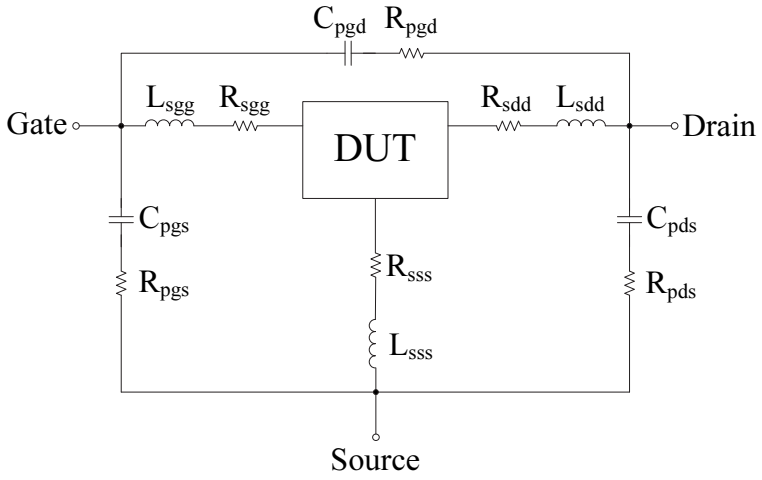


Figure 34: Pad and interconnect parasitics of MOS test devices.

small-signal model at the desired bias condition [12]. The schematic diagram of pad and interconnect is shown in Fig. 34 and equations for finding pad and interconnect parasitics are given as follow:

$$C_{pgd} = -\frac{1}{\omega}\left(\frac{Re\{Y_{12}^o\}^2}{Im\{Y_{12}^o\}} + Im\{Y_{12}^o\}\right) \tag{49}$$

$$C_{pgs} = \frac{1}{\omega}\left(\frac{Re\{Y_{11}^o + Y_{12}^o\}^2}{Im\{Y_{11}^o + Y_{12}^o\}} + Im\{Y_{11}^o + Y_{12}^o\}\right) \tag{50}$$

$$C_{pds} = \frac{1}{\omega}\left(\frac{Re\{Y_{22}^o + Y_{12}^o\}^2}{Im\{Y_{22}^o + Y_{12}^o\}} + Im\{Y_{22}^o + Y_{12}^o\}\right) \tag{51}$$

$$R_{pgd} = \frac{1}{\omega C_{pgd}}\sqrt{\frac{-\omega C_{pgd}}{Im\{Y_{12}^o - 1\}}} \tag{52}$$

$$R_{pgs} = \frac{1}{\omega C_{pgs}}\sqrt{\frac{\omega C_{pgs}}{Im\{Y_{11}^o + Y_{12}^o - 1\}}} \tag{53}$$

$$R_{pds} = \frac{1}{\omega C_{pds}}\sqrt{\frac{\omega C_{pds}}{Im\{Y_{22}^o + Y_{12}^o - 1\}}} \tag{54}$$

$$R_{sgg} = Re\{Z^s_{11} - Z^s_{12}\} \tag{55}$$

$$R_{sdd} = Re\{Z^s_{22} - Z^s_{12}\} \tag{56}$$

$$R_{sss} = Re\{Z^s_{12}\} \tag{57}$$

$$L_{sgg} = \frac{Im\{Z^s_{11} - Z^s_{12}\}}{\omega} \tag{58}$$

$$L_{sdd} = \frac{Im\{Z^s_{22} - Z^s_{12}\}}{\omega} \tag{59}$$

$$L_{sss} = \frac{Im\{Z^s_{12}\}}{\omega} \tag{60}$$

where C_{pgd}, C_{pgs}, C_{pds}, R_{pgd}, R_{pgs}, and R_{pds} are pad parasitic capacitances and resistances associated with each pad; R_{sgg}, R_{sdd}, R_{sss}, L_{sgg}, L_{sdd}, and L_{sss} are interconnect parasitic resistances and inductances (excluded from the intrinsic device). The pad and interconnect parasitics removal must be done prior to the extraction of intrinsic device parameters. To remove these parasitics, the following procedure can be followed;

$$Y_{DUT,OPEN} = Y_{DUT} - Y_{OPEN} \tag{61}$$

$$Y_{SHORT,OPEN} = Y_{SHORT} - Y_{OPEN} \tag{62}$$

$$Z_{DUT,OPEN,SHORT} = Z_{DUT,OPEN} - Z_{SHORT,OPEN} \tag{63}$$

where $Y_{DUT,OPEN}$ is the Y-parameter of the device after the de-embedding of pad parasitics using opened structure (structure with transistor removed), $Y_{SHORT,OPEN}$ is the Y-parameter of shorted structure (remove the transistor and insert a small piece of metal to connect every path together) after de-embedding of pad parasitics, and

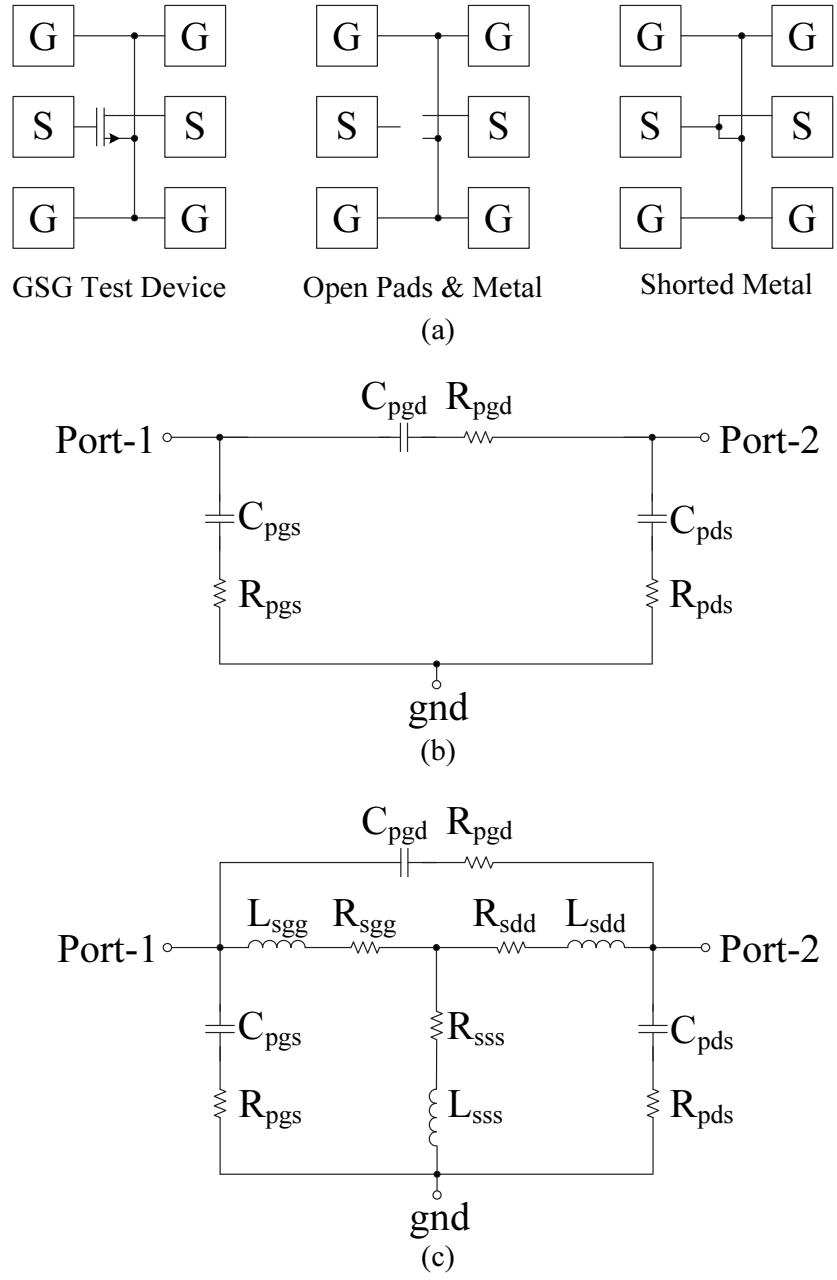


Figure 35: Structures required for two-step pad and interconnect parasitics removal procedure; (a) test device, opened, and shorted structures, (b) opened-structure parasitic components, (c) shorted-structure parasitic components.

$Z_{DUT,OPEN,SHORT}$ is the Z-parameter of the intrinsic device after removing all the pad and interconnect parasitics. Details of the parasitic components are illustrated in Fig. 35. The transformation between Z-, Y-, and S-parameter can be found in [48].

Next, the gate, drain, and source resistances (R_g, R_d, and R_s) can be extracted using the cold-FET method – a high-frequency parameter extraction from zero-biased measurements – by the following equations:

$$R_g = Re\{Z_{11}^{cold} - Z_{12}^{cold}\} \tag{64}$$

$$R_d = Re\{Z_{22}^{cold} - Z_{12}^{cold}\} \tag{65}$$

$$R_s = Re\{Z_{12}^{cold}\} \tag{66}$$

The extraction is done with an assumption that the impedance of intrinsic capacitances at very high frequency is negligible($\gg$ 20 GHz for C_{gs} and $C_{gd} \leq$ 100 fF), thus shorting all capacitors and leaving only R_g, R_d, and R_s.

Sometimes, for higher frequency applications (beyond 20 GHz), the terminal inductances (L_g, L_d, and L_s) may have comparable impedance to terminal resistances and should be included for better accuracy. These inductances can be calculated by

$$L_g = \frac{Im\{Z_{11}^{cold} - Z_{12}^{cold}\}}{\omega} \tag{67}$$

$$L_d = \frac{Im\{Z_{22}^{cold} - Z_{12}^{cold}\}}{\omega} \tag{68}$$

$$L_s = \frac{Im\{Z_{12}^{cold}\}}{\omega} \tag{69}$$

These terminal resistances and inductances are removed before the extraction of R_{sub} and C_{sub} using (47) and (48). Then, to finish all the intrinsic parameter extractions, the R_{sub} and C_{sub} are first removed from the Y-parameter of the device after de-embedding terminal resistances and inductances in (64) - (69). Finally, the remaining intrinsic parameters can be extracted using the following equations:

$$C_{gd} = -\frac{Im\{Y_{12}^d\}}{\omega} \tag{70}$$

$$C_{gs} = \frac{Im\{Y_{11}^d + Y_{12}^d\}}{\omega} \tag{71}$$

$$R_i = \frac{Re\{Y_{11}^d\}}{(\omega C_{gs})^2} \tag{72}$$

$$g_m = Re\{Y_{21}^d\} \tag{73}$$

$$\tau = -\frac{g_m(Im\{Y_{21}^d - Y_{12}^d\})}{\omega} - R_i C_{gs} \tag{74}$$

$$r_{ds} = \frac{1}{Re\{Y_{22}^d\}} \tag{75}$$

where Y^d is the Y-parameter of the device after de-embedding terminal resistances and inductances and substrate network, C_{gd} is the gate-drain capacitance, C_{gs} is the gate-source capacitance, R_i is the internal resistance of the gate-source junction, g_m is the small-signal current gain, τ is the channel time delay, and r_{ds} is the drain-source output resistance.

These parameters are extracted from a number of devices in several dies. Then, the most consistent extracted values were carefully selected in order to minimize the measurement uncertainties without doing any optimization. Most of the extracted parameters show a linear scaling ability. This includes parasitic resistances (R_g, R_d, and R_s), gate-source capacitance (C_{gs}), gate-drain capacitance (C_{gd}), output resistance (r_{ds}), R_i, and small-signal current gain (g_m). R_g, R_d, R_s, R_i, and r_{ds} scale in inverse proportion to W as they are dependent on the number of fingers connected in parallel (see Table 1 for values.) C_{gs}, C_{gd}, and g_m are observed to be the least

Table 1: Extracted intrinsic and substrate parameters of N-MOSFET devices at V_{gs} = 1.2 V; V_{ds} = 2.4 V.

parameters	200 μm	800 μm	2400 μm
$R_g(\Omega)$	12.7	3.3	1.3
$R_d(\Omega)$	6.5	1.8	0.7
$R_s(\Omega)$	7.8	1.9	0.6
$R_i(\Omega)$	9.9	2.7	0.8
C_{gs} (fF)	246	993	2953
C_{gd} (fF)	55	227	703
g_m (mS)	42	170	513
$R_{ds}(\Omega)$	856	204	63
C_{sub} (fF)	58	226	716
$R_{sub}(\Omega)$	158	82	35

sensitive parameters to extraction uncertainties and also show good linear scalability, which is in proportion to W (see Fig. 36).

4.3.3 Scalable Substrate Network Modeling

The analysis of the substrate network is based on a set of 0.4-μm multi-finger NMOS thick-oxide devices having gate finger width of 20 μm with ring-shaped body contacts surrounding the active device. This type of substrate connection has been used widely in RF CMOS circuit design since it minimizes the discrepancy of voltages between source junction and substrate. Nevertheless, this increases the difficulty in developing an accurate scalable substrate model since the distributed nature of substrate coupling is not limited to any direction. In this study, we propose the simplified model of substrate resistance that considers two main paths: the substrate resistance from the junction to the side substrate contacts (horizontal), R^h_{db} and R^h_{sb}, and to the top and bottom substrate contacts (vertical), R^v_{db} and R^v_{sb} (see Fig. 37), instead of considering the distribution to the entire substrate contact. This approximation is acceptable because the smaller resistance, which could be one of the two directions as stated, dominates the larger resistance in other directions. By this simplification, the equation for approximating substrate resistance can be formulated as follows:

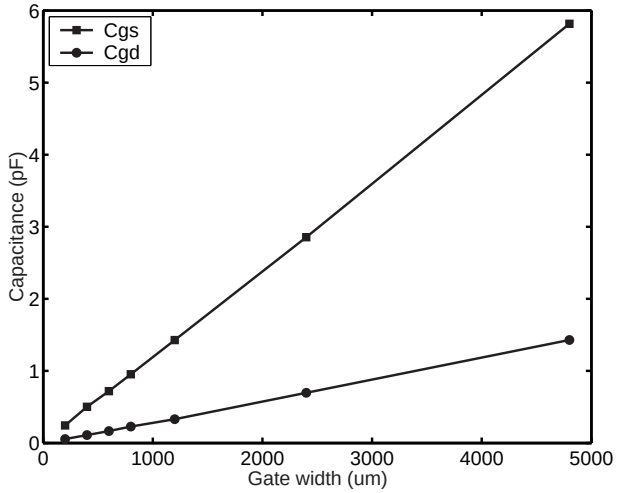


Figure 36: Scaling of C_{gs} and C_{gd} with different number of fingers.

$$\frac{1}{R_{db}^t} = \frac{1}{R_{db}^h} + \frac{1}{R_{db}^v} = \sum_{k=1}^{Nd} \left(\frac{1}{R_{db,k}^h} + \frac{1}{R_{db,k}^v} \right) \tag{76}$$

$$\frac{1}{R_{sb}^t} = \frac{1}{R_{sb}^h} + \frac{1}{R_{sb}^v} = \sum_{k=1}^{Ns} \left(\frac{1}{R_{sb,k}^h} + \frac{1}{R_{sb,k}^v} \right) \tag{77}$$

$$\frac{1}{R_{dsb}^t} = \sum_{k=1}^{Nf} \frac{1}{R_{dsb,k}} \tag{78}$$

where R_{db}^t and R_{sb}^t are the total drain-to-bulk and total source-to-bulk resistances, R_{dsb}^t is the equivalent resistance between drain and source underneath the channel in the substrate, and $R_{db,k}^h$, $R_{db,k}^v$, $R_{sb,k}^h$, $R_{sb,k}^v$, and $R_{dsb,k}$ are the equivalent resistances corresponding to each drain and source junction in the horizontal (h) and vertical (v) directions. N_f, N_d, and N_s are the number of gate, drain, and source fingers, respectively.

R_{sub} is an approximation of the substrate resistance from the substrate network equivalent circuit in Fig. 38(a).

$$R_{sub} \approx \frac{R_{db}^t (R_{sb}^t + R_{dsb}^t)}{R_{db}^t + R_{sb}^t + R_{dsb}^t} \tag{79}$$

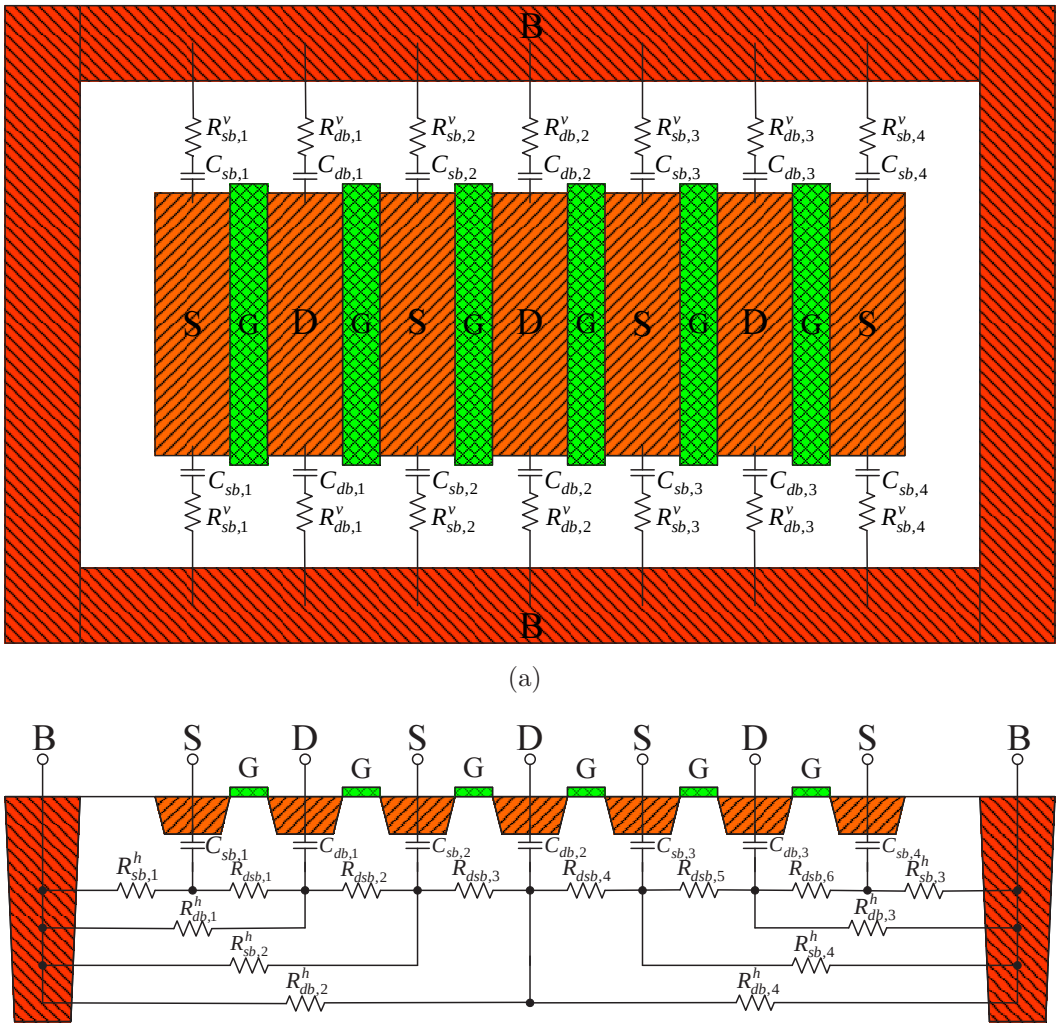


Figure 37: Substrate coupling in MOSFET device with surrounded body contacts, (a) vertical coupling, (b) horizontal coupling.)

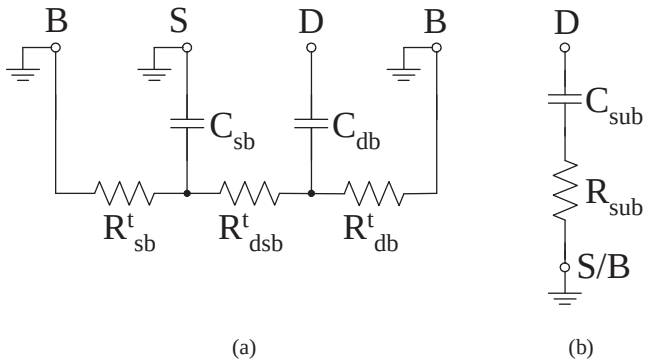


Figure 38: (a) Equivalent circuit of the substrate network, (b) approximated substrate network.

The total substrate capacitance can be calculated from the following equation:

$$C_{sub} \approx \sum_{k=1}^{Nd} C_{db,k} \tag{80}$$

with the assumption that $(\omega C_{sb})^2 R_{sb}^2 \ll 1$ (valid for the frequency range up to 10 GHz).

The extraction of each substrate resistance parameter can be done using the extracted R_{sub} from several devices with different number of fingers. The contribution of the substrate resistance of the inner drain and source junctions to the side substrate contacts is very small compared to that from the side most junctions [11, 19]. Therefore the R_{db}^h and R_{sb}^h can be approximated as a contribution from the side most junctions to reduce the extraction complexity. For a symmetric device layout (even number of gate fingers), The value of R_{db}^h, R_{sb}^h, R_{db}^v, R_{sb}^v, and R_{dsb}^t can be calculated as

$$R_{db}^h \approx \frac{r_{db}^h}{2} \cdot \frac{d_{db}^h}{W_f} \tag{81}$$

$$R_{sb}^h \approx \frac{r_{sb}^h}{2} \cdot \frac{d_{sb}^h}{W_f} \tag{82}$$

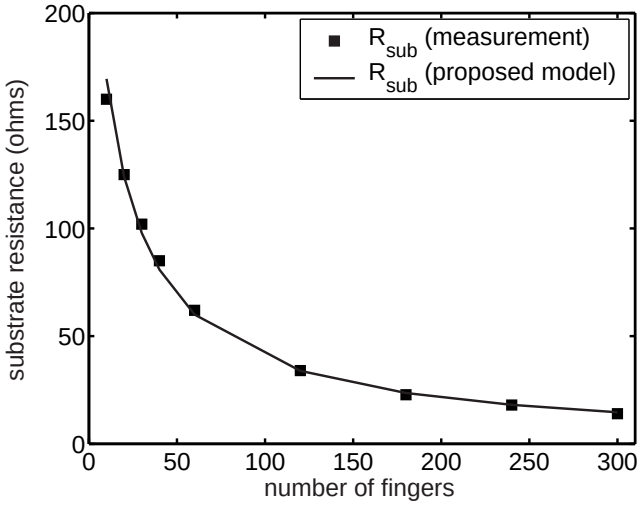
$$R_{db}^{v} = \alpha \cdot \frac{r_{db}^{v}}{2N_d} \cdot \frac{d^v}{L_d} \tag{83}$$

$$R_{sb}^{v} = \alpha \cdot \frac{r_{sb}^{v}}{2N_s} \cdot \frac{d^v}{L_s} \tag{84}$$

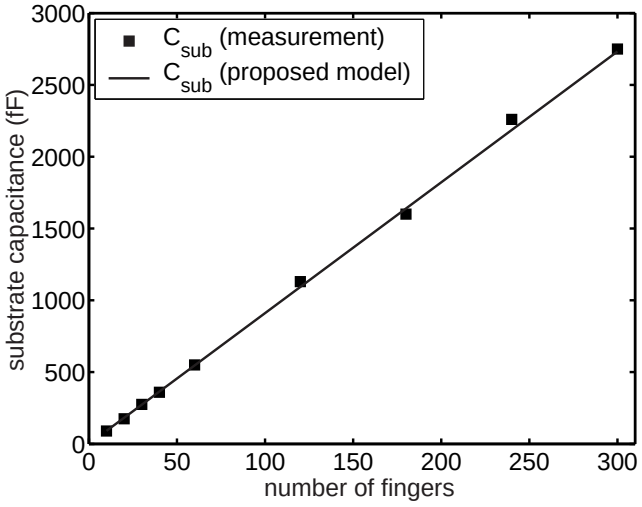
$$R_{dsb}^{t} = \frac{r_{dsb}}{N_f} \cdot \frac{L_f}{W_f} \tag{85}$$

where r_{db}^{h}, r_{db}^{v}, r_{sb}^{h} , r_{sb}^{v}, and r_{dsb} are unit substrate resistances in the horizontal and vertical directions corresponding to each junction. d_{db}^{h} and d_{sb}^{h} are the distances between the center of drain and source outmost finger and the side substrate contact. d^v is the distance from the upper or lower edge of drain and source finger to the top or bottom substrate contact. L_f, L_d, and L_s are the device channel length, and the length of the drain and source junction, respectively. The parameter α in (83) and (84) is used to correct the distributed behavior of R_{db}^{v} and R_{sb}^{v} beneath the drain and source junctions coupled to the top and bottom substrate contacts ($0.5 \leq \alpha \leq 1$).

The substrate network parameters, R_{sub} and C_{sub}, of 0.4-μm multi-finger NMOS thick-oxide devices with total gate width of 200 μm to 6 mm and having a ring-shaped substrate contact are extracted and shown in Fig. 39. Each device uses identical gate finger width of 20 μm, d^v= 2.8 μm, d_{db}^{h} = 6.6 μm, d_{sb}^{h}= 5.6 μm, and $L_d = L_s$ = 0.6 μm. The correction parameter α is obtained from the extraction of (81)-(85), by assuming the same unit substrate resistance is used in both directions. In this case, the parameter α is determined to be 0.5. It is clearly seen that the new model can predict substrate parameters very accurately. The extraction of R_{db}^{t}, R_{sb}^{t}, and R_{dsb}^{t} in Fig. 40 illustrates the scalability for a variety of device sizes using the same gate finger width. It is observed from Fig. 41 that the substrate resistance components in vertical direction (R_{db}^{v} and R_{sb}^{v}) are much smaller than those in the horizontal direction (R_{db}^{h} and R_{sb}^{h}), thus dominating the R_{sub} value as the device size increases.



(a)



(b)

Figure 39: Substrate network parameters extraction results, (a) substrate resistance, (b) substrate capacitance. ($V_{ds} = 0V$; $V_{gs} = 0V$)

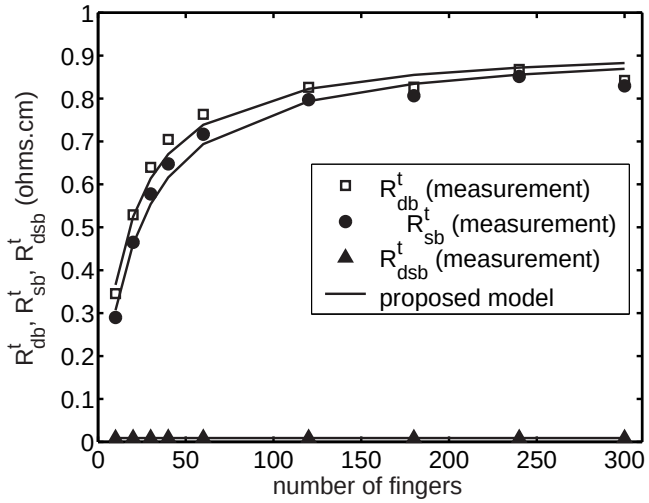


Figure 40: Scalability of substrate resistance components. ($V_{ds} = 0V$; $V_{gs} = 0V$)

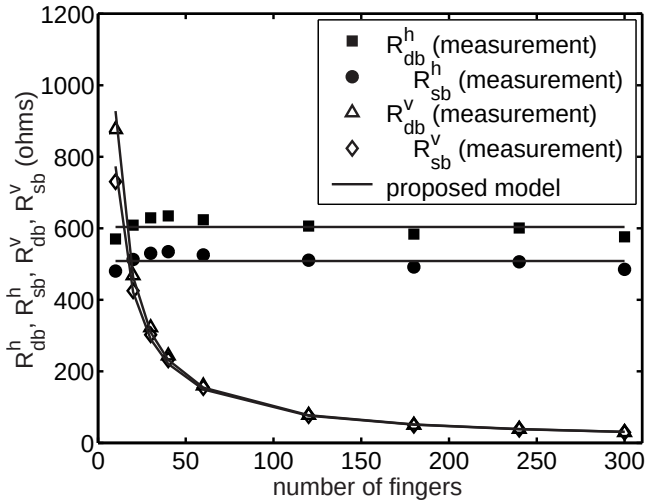


Figure 41: Substrate resistance components in vertical and horizontal directions. ($V_{ds} = 0V$; $V_{gs} = 0V$)

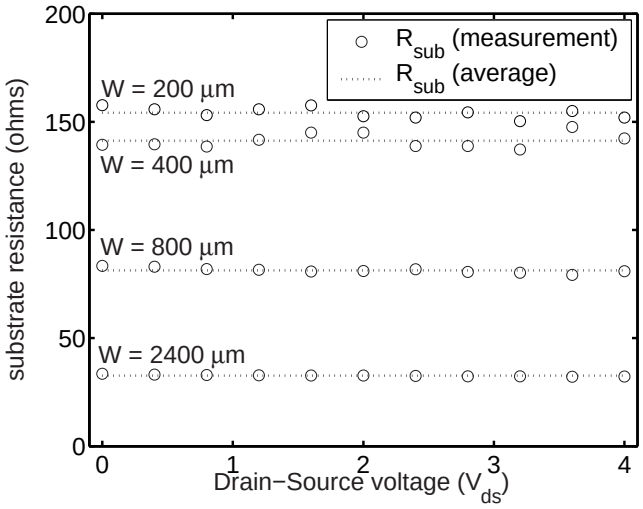


Figure 42: Bias dependence of substrate resistance when $V_{gs} = 0$ V.

This confirms the necessity of including the substrate coupling in the vertical direction to the modeling of substrate resistance.

The substrate parameters are extracted with the change in drain bias voltage to show their bias dependence. The results in Fig. 42 show that the substrate resistance is bias independent as mentioned in [11, 12], while the substrate capacitance is bias dependent. The characteristic of C_{db} (or C_{sb}) can be modeled by

$$C_{db} = \frac{C_{db0}}{\left(1 + \frac{Vdb}{\phi_{bi}}\right)^{\alpha_j}} \tag{86}$$

where C_{db0} is C_{db} at zero biased condition, ϕ_{bi} is the built-in potential which is obtained from the fitting of measurement results to the model, and is equal to 0.78 V. α_j is a fitting parameter which is equal to 0.3 in this case. The results in Fig. 43 shows a good agreement between this model and the measurement.

In addition to the previous results, substrate resistance and capacitance of 2400-μm NMOS devices with different finger width, ranging from 10 μm to 50 μm, are extracted and verified using the proposed model, from which the results are shown in Fig. 44 and Fig. 45. It can be seen that the model can predict the parameter

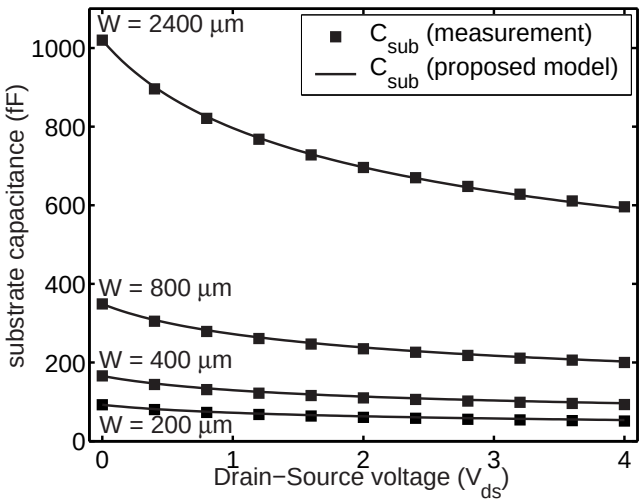


Figure 43: Bias dependence of substrate capacitance when $V_{gs} = 0$ V.

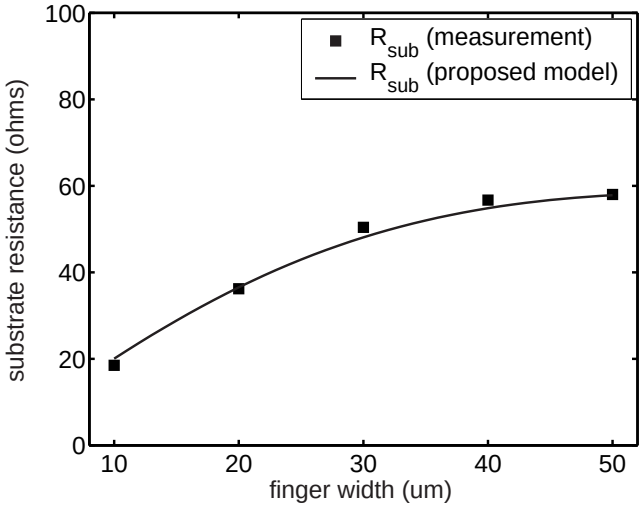


Figure 44: Substrate resistance of 2400 μm devices with different finger size. ($V_{ds} = 0V$; $V_{gs} = 0V$)

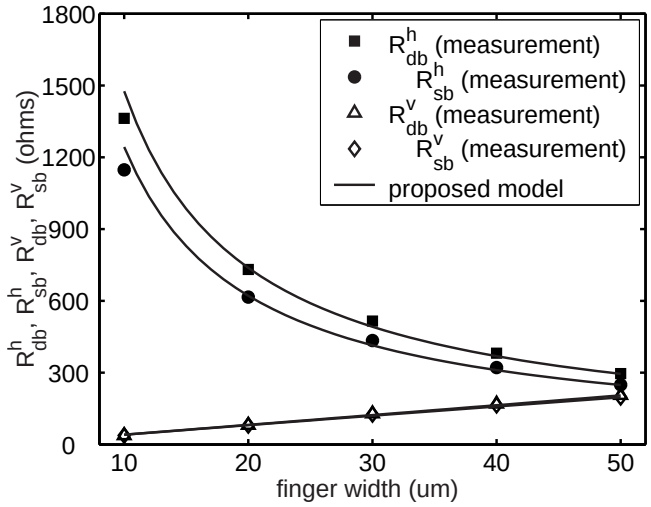


Figure 45: Substrate resistance components in vertical and horizontal directions of 2400 μm devices with different finger size. ($V_{ds} = 0V$; $V_{gs} = 0V$)

scalability correctly.

The MOSFET parameters in Fig. 31 are extracted to validate the substrate model when the transistor is in biased condition (V_{gs} = 1.2 V; V_{ds} = 2.4 V). S-parameter simulations, illustrated in Fig. 46, 47, and 48, show a good agreement with the measurement values obtained from many devices (W = 200, 800, and 2400 μm). Additionally, the Y_{22} (Fig. 49), which is strongly dependent on substrate parameters, has shown a close match between simulation and measurement results.

4.3.4 Modified-BSIM3v3 Model

In the BSIM3v3 model, intrinsic parameters have been readily included except the R_g and R_{sub}. While R_g can be found similarly to the method suggested in previous section, there are many different ways to incorporate R_{sub} [37, 41, 67].

From the device cross section (Fig. 37), the substrate network can be derived analytically and modeled as shown in Fig. 50. R_{db} and R_{sb} are the combination of R_{db} and R_{sb} in vertical and horizontal directions. This model has junction capacitances (C_{db} and C_{sb}) that are not included in the intrinsic model (M_i). This model can be

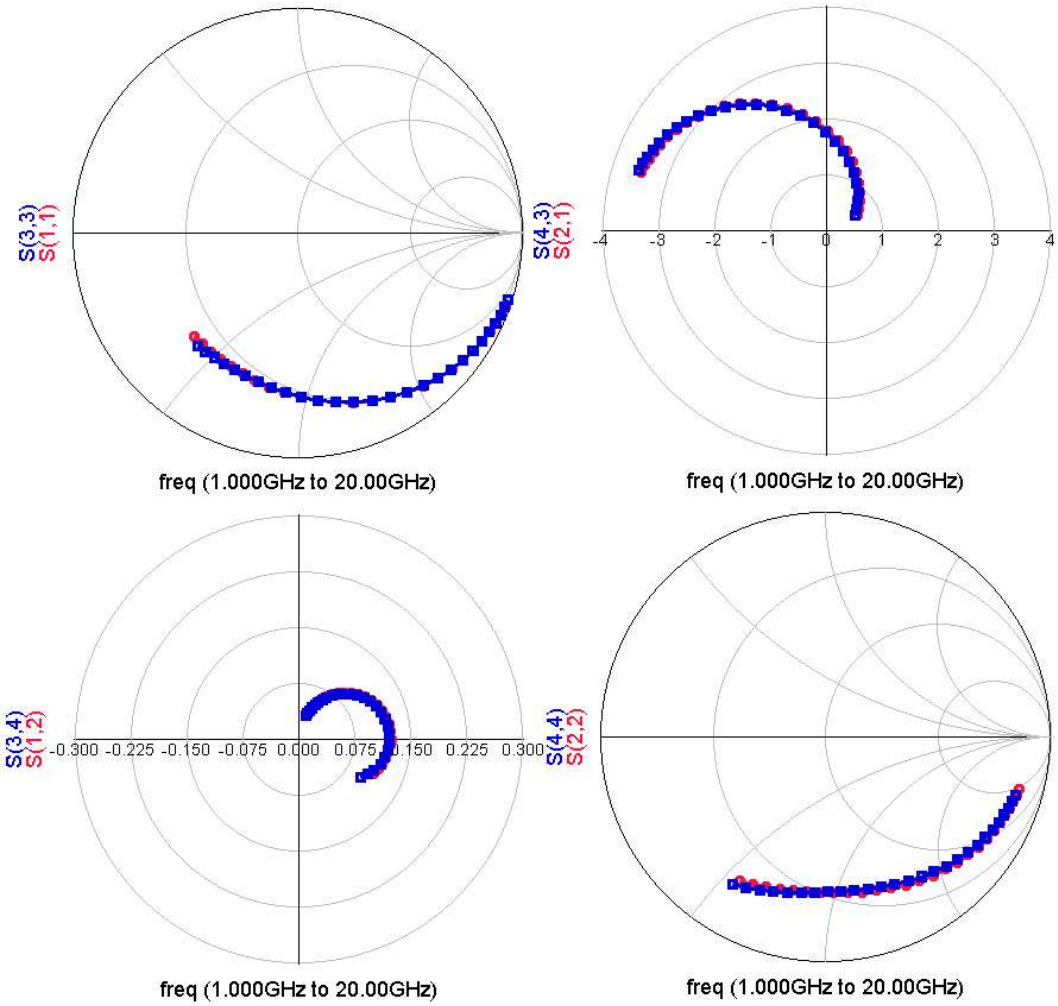


Figure 46: S-parameter of 0.4 μm thick-oxide NMOS devices with W of 200 μm. Measurement is shown in circle and simulation in square. ($V_{ds} = 2.4V$; $V_{gs} = 1.2V$)

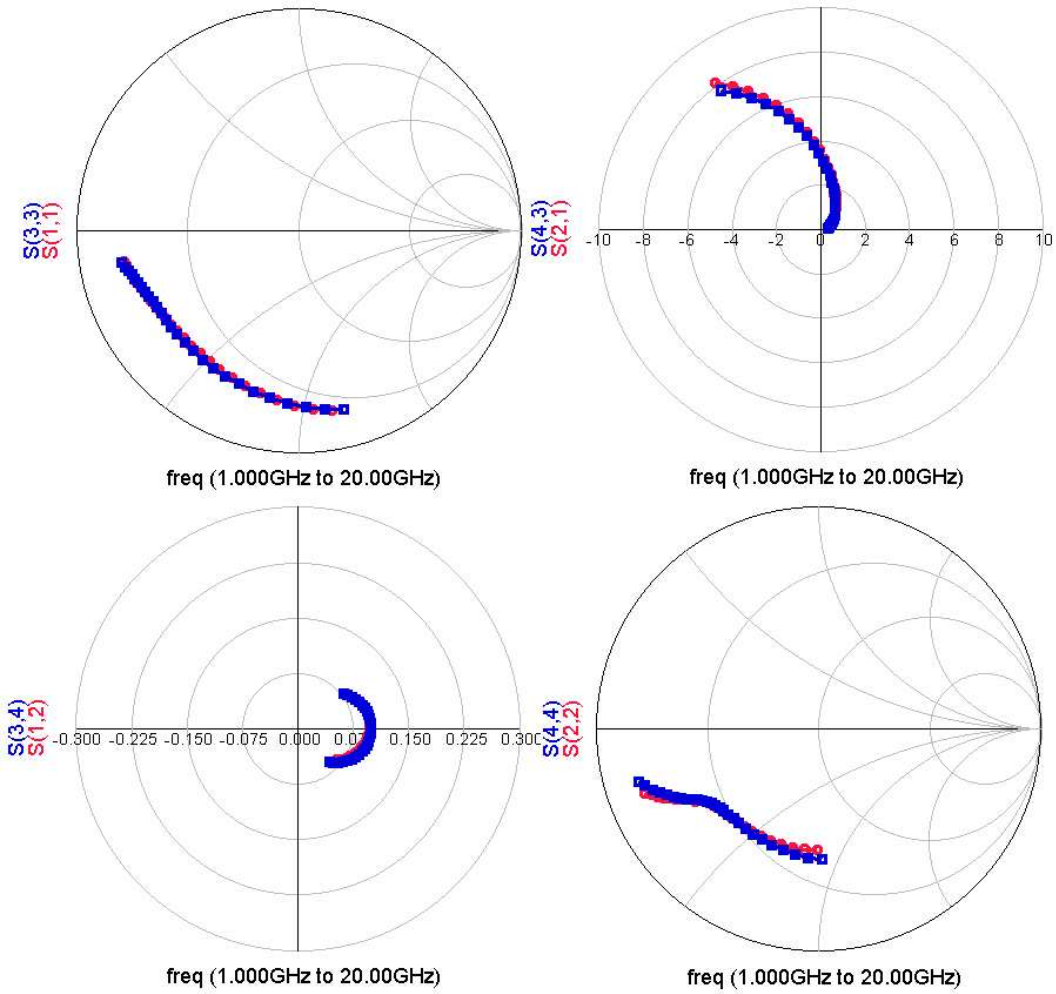


Figure 47: S-parameter of 0.4 μm thick-oxide NMOS devices with W of 800 μm. Measurement is shown in circle and simulation in square. ($V_{ds} = 2.4V$; $V_{gs} = 1.2V$)

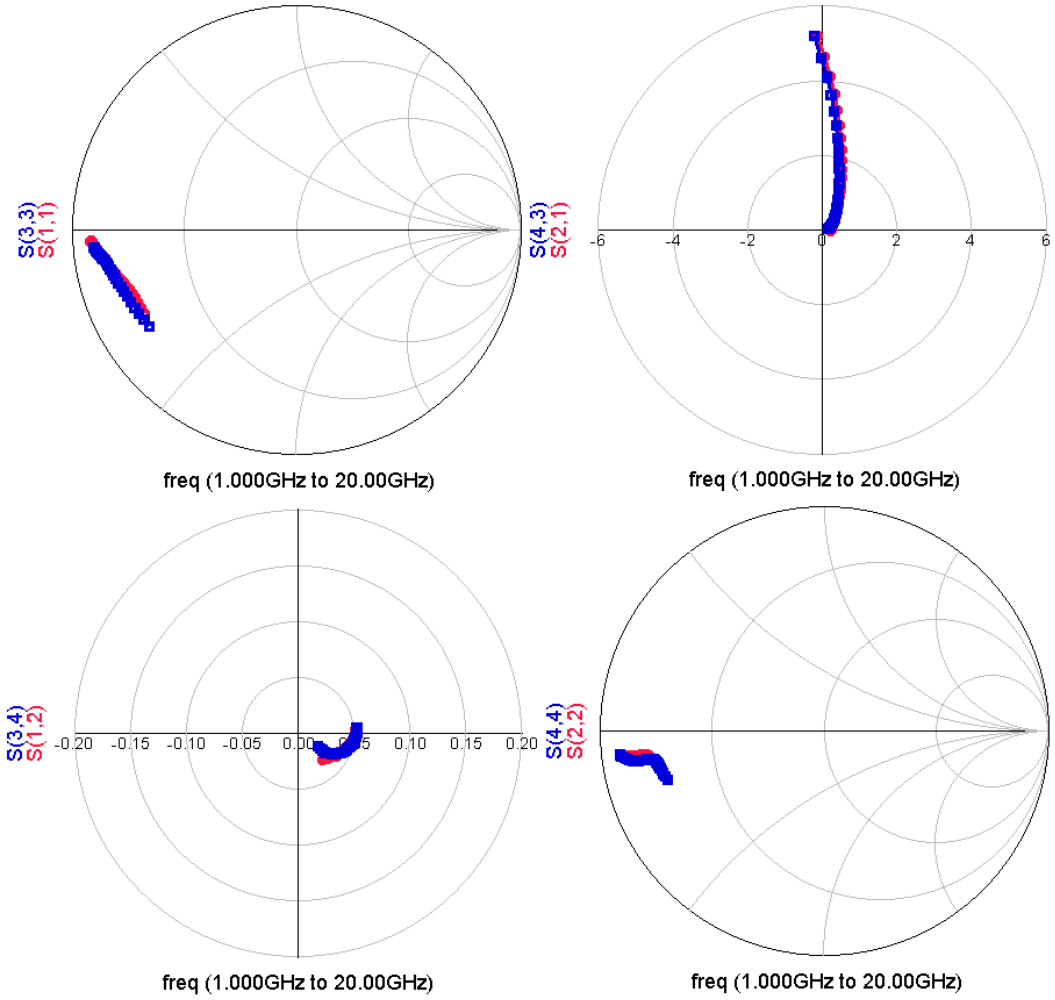
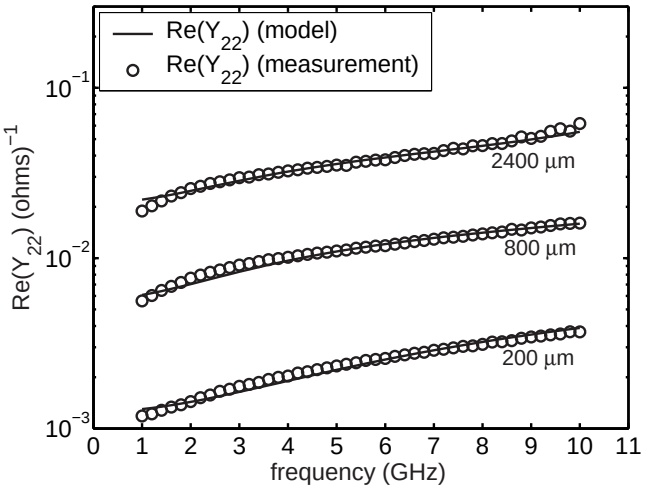
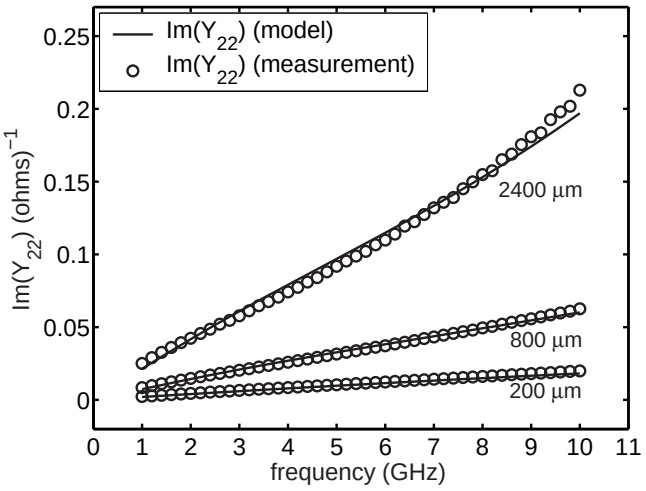


Figure 48: S-parameter of 0.4 μm thick-oxide NMOS devices with W of 2400 μm. Measurement is shown in circle and simulation in square. ($V_{ds} = 2.4V$; $V_{gs} = 1.2V$)



(a)



(b)

Figure 49: Y_{22} of the equivalent model compared to the measurement results, (a) real part, (b) imaginary part.

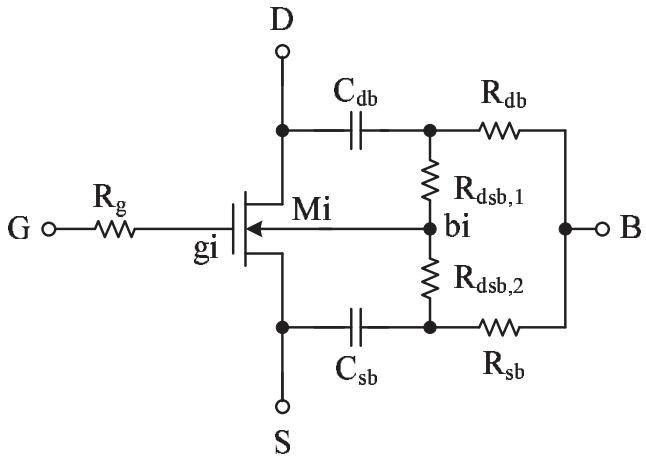


Figure 50: Analytical MOSFET model including gate resistance and substrate network.

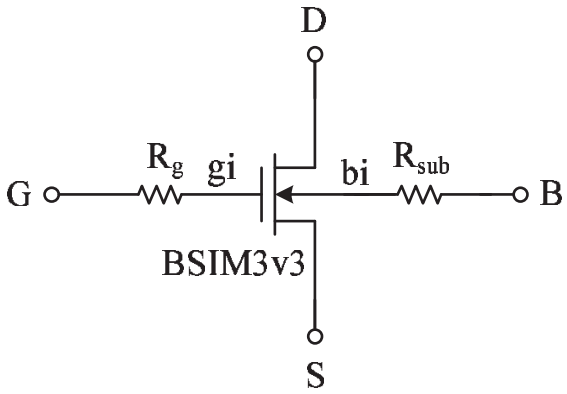


Figure 51: BSIM3v3 MOSFET model with gate resistance and simplified substrate network.

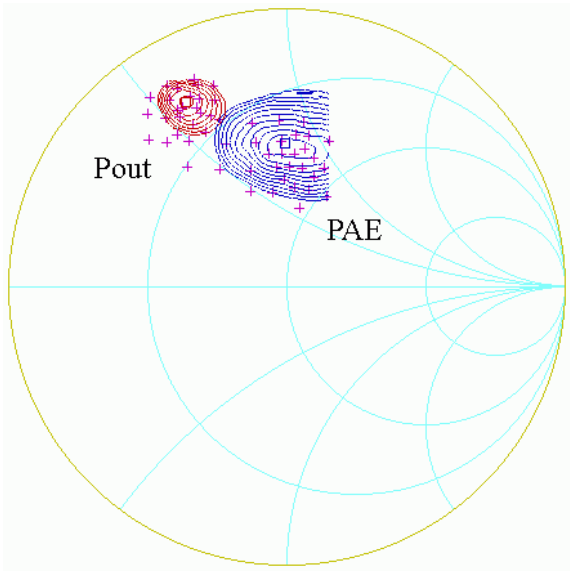
simplified to the one in Fig. 51 that uses BSIM3v3 model as a core. The BSIM3v3 has already included intrinsic parameters and junction capacitances C_{db} and C_{sb}. The substrate resistance can be simplified to a single R_{sub} that is calculated from (79). The advantage of using a single R_{sub} is the simplicity of the model where the accuracy should be well enough at the frequency below 10 GHz.

The impedance of C_{sb} is usually higher than that of R_{sb} at the frequency below 10 GHz. Therefore, when the source terminal is tied to bulk, C_{sb} can be omitted and the equivalent model in Fig. 31 becomes interchangeable to the simplified BSIM3v3 model (Fig. 51). The same extraction method can also be applied to get the R_{sub}.

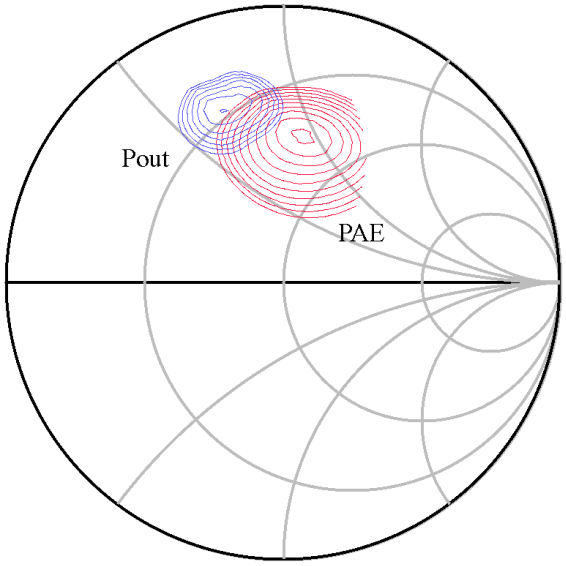
Large signal measurements such as load-pull and power sweep measurements are used to verify the model in Fig. 51. Load-pull measurements of 400-μm and 1200-μm devices are compared with the simulation and have shown a good match (Fig. 52 and 53). Power sweep results of the 1200-μm device and 4800-μm devices (illustrated in Fig. 54 and 55) show that the model can predict the large-signal characteristics such as gain and efficiency. The developed model can be used further in designing CMOS power amplifiers and other small- and large-signal circuits.

4.4 Summary

The substrate parameter scalability of RF MOSFETs having ring-shaped substrate contact surrounding the device has been analyzed and modeled using the newly proposed analytical expressions based on device geometry. The extracted substrate resistance and capacitance of devices up to 6 mm total gate width were used in the modeling procedure. The analysis shows that the distributed substrate coupling when using ring-shaped substrate contact can be modeled by simplifying it to substrate coupling in vertical and horizontal directions. With the use of the proposed method, the scalability of each substrate parameter, as well as the modeling of substrate parameters based on device geometry can be achieved with great accuracy. This scalable

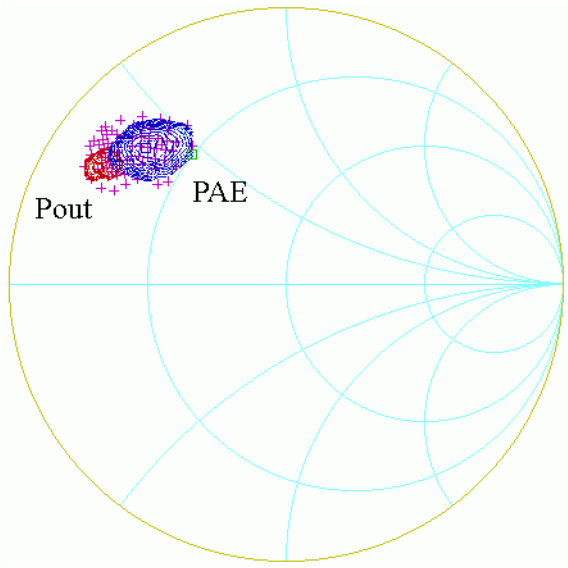


(a)

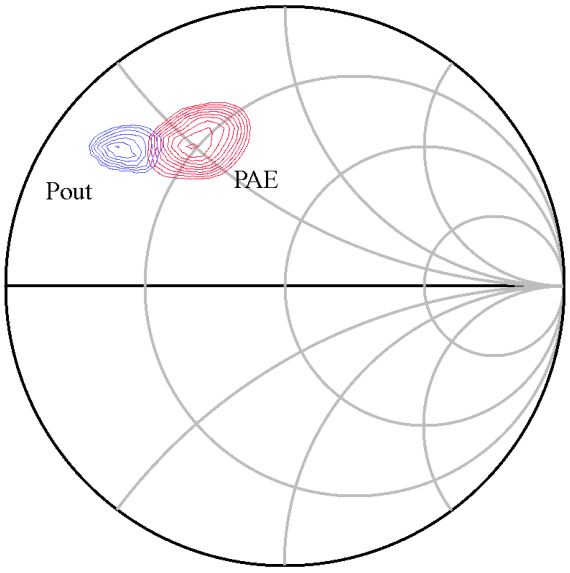


(b)

Figure 52: Load-pull results of the 400-μm NMOS device, (a) measurement, (b) simulation.



(a)



(b)

Figure 53: Load-pull results of the 1200-μm NMOS device, (a) measurement, (b) simulation.

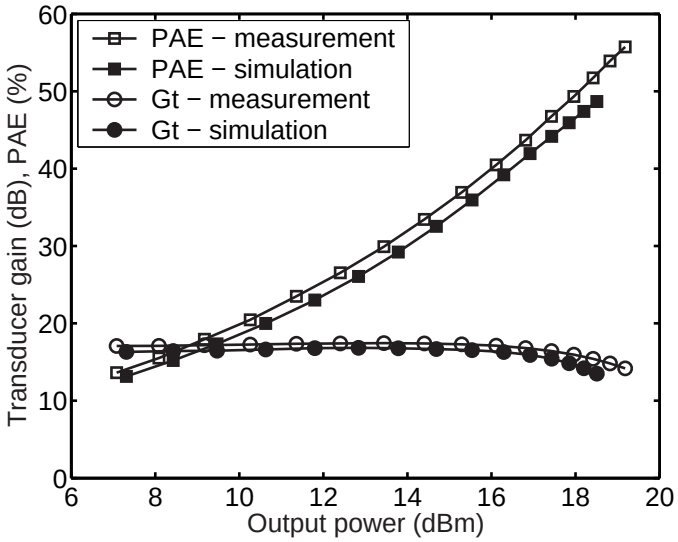


Figure 54: Transducer gain and PAE of the 1200-μm device. ($V_{ds} = 2.4V$; $V_{gs} = 0.65V$; $Z_{opt,in} = 5.5 + j28.0$ ohms; $Z_{opt,out} = 14.5 + j23.0$ ohms)

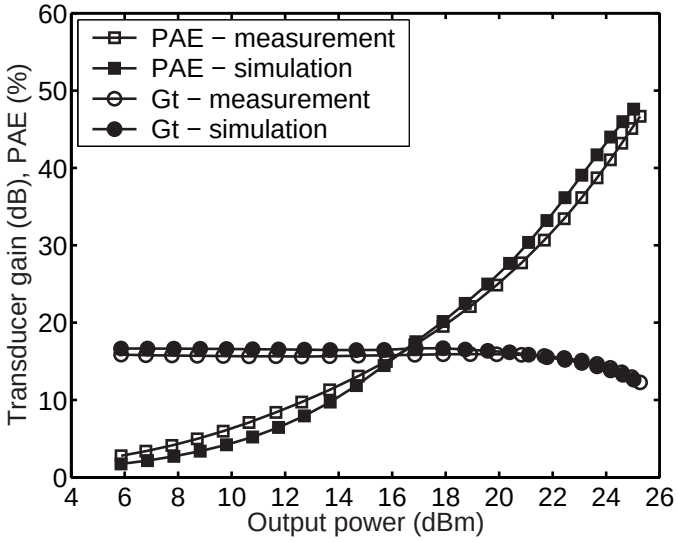


Figure 55: Transducer gain and PAE of the 4800-μm device. ($V_{ds} = 2.4V$; $V_{gs} = 0.79V$; $Z_{opt,in} = 2.3 + j6.0$ ohms; $Z_{opt,out} = 3.9 + j4.0$ ohms)

substrate model is verified by small-signal S-parameter measurements and has shown a good agreement. Furthermore, the gate resistance and substrate network model are incorporated with the BSIM3v3 model and verified by large-signal load-pull measurements. The results have ensured that the developed model can be used as a large-signal RF MOSFET model for RF power amplifier design and applications.

CHAPTER 5

THE DESIGN OF INTEGRATED RF POWER AMPLIFIERS WITH THE DEVICE PERIPHERY ADJUSTMENT TECHNIQUE

5.1 Motivation and Goal

Efficiency enhancement of linear RF power amplifiers is crucial for modern wireless communication systems. Until now, many techniques have been developed for efficiency enhancement in power amplifier design, such as envelope elimination and restoration (EER) and dynamic biasing technique [23, 53]. However, these techniques have not achieved the development of optimized compact power amplifier integrated circuits (ICs) since many additional controlling circuits and components are required.

Silicon-based RF power amplifiers have shown a strong potential for product-level development as it enables the design of system-on-a-chip (SOC). The total integration of RF building blocks can lower the cost and chip area, which makes the Si-based design a better option than the multiple-dies, III-V-based design. Recent SiGe HBT technology has shown excellent dc and RF electrical characteristics in terms of transition frequency (f_T) and dc current gain (β), which are comparable to those of GaAs HBT process [40]. The improved breakdown voltage (BV_{CEO}, as high as 5.5 V) of the high-voltage HBT device gives more headroom and is more attractive to RF power amplifier design. Even though these characteristics are still dominated by GaAs devices, the PNP or depletion-mode devices, which are not available in GaAs processes, are the main attractions of silicon-based technologies that allow more functionalities in the PA design.

Additionally, advancements in CMOS technology for integrated high-power amplification in the gigahertz range have been reported [2]. It has shown that, with

improved design techniques, CMOS can become a prospected candidate for RF PA design. Thermal conductivity of silicon (1.5 W/cm-C) is three times higher than that of GaAs (0.49 W/cm-C), showing an advantage of less current collapsing issues. Still, new methods to overcome problems such as low efficiency and poor linearity in linear RF CMOS and SiGe power amplifiers are necessary and need more investigations.

In current mobile communication standards such as CDMA or W-CDMA, the output power of a handset must be reduced when the interference level to the nearby channels is high. This is adjusted by reducing the input signal generated from the variable-gain amplifier (VGA) in the prior stage before power amplification. The efficiency of power amplifier at this low input level is smaller than that in high power level. Therefore, power amplifier must be designed to be able to control the dc consumption to improve the conversion efficiency.

To date, there are a few techniques suitable for improving low-power efficiency in single-chip power amplifier IC design. The switching of collector or drain quiescent current level is widely used in commercial power amplifier MMICs in order to reduce dc power consumption while keeping the linearity level within the system requirements [20]. This can increase the average power-added efficiency (PAE) of the power amplifier and prolong the battery usage time for a mobile handset with trade-offs in performances such as gain and linearity. Nevertheless, limited PAE improvement has been observed in class-AB amplifiers where the quiescent current in low-power mode is sufficiently small. This technique will be more effective with class-A amplifiers where the quiescent current is usually high (see Fig. 56).

Alternatively, power amplifier design with stage bypassing technique was developed to improve the efficiency when operating in low power level [64, 65]. This technique uses a switch to bypass the output stage in low-power mode, as shown in Fig. 57(a), and shuts down the output transistor, which decreases the overall dc power consumption and increases the PAE. Limitations of this approach are the use

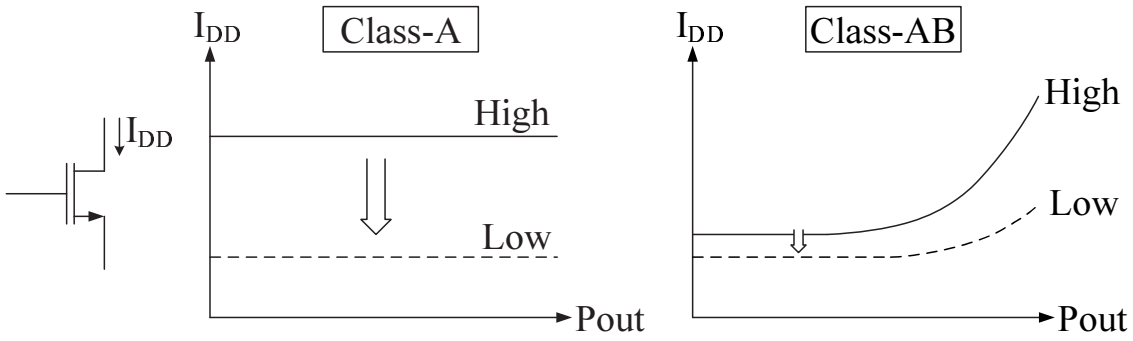


Figure 56: Current level switching method to improve efficiency in low-power transmission mode.

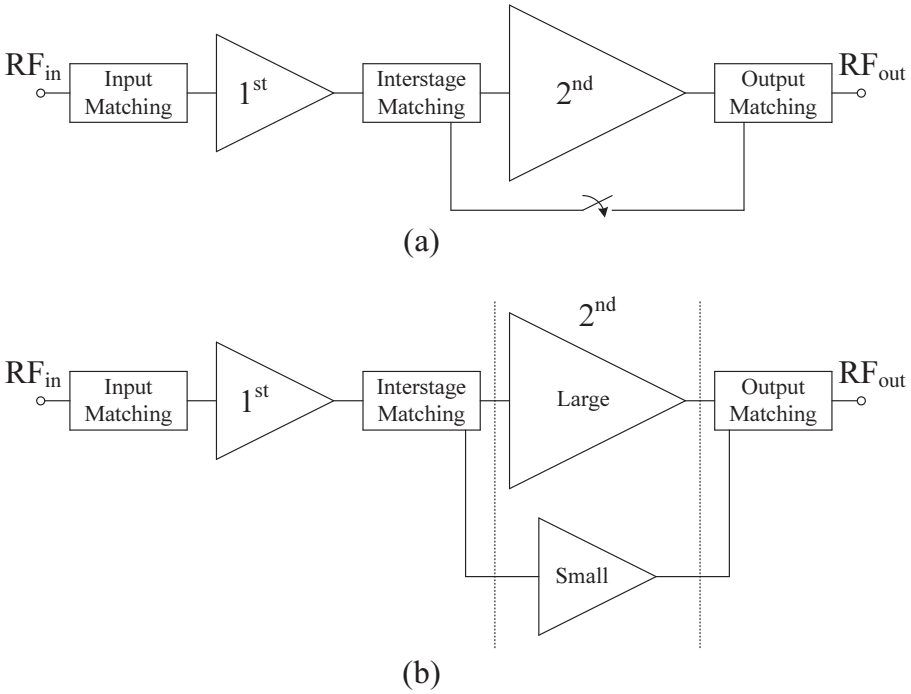


Figure 57: Different topologies for efficiency enhancement of single-chip linear power amplifier integrated circuits, (a) stage bypassing technique, (b) active area adjustment technique.

of switch that introduces loss and the gain difference between modes of operation. Even though an external low-loss switch can be used, the extra component adds cost and increases area, which could become a disadvantage of this approach.

An approach to improve efficiency by adjusting the active area as illustrated in Fig. 57(b) has been reported recently [29, 30, 44]. The improvement in PAE can be achieved at the cost of increased complexity in matching network design. Reported results indicate that this approach provides maximum improvement in PAE while meeting current wireless communication standards such as IS-95 and W-CDMA.

In this chapter, the technique of adjusting the active device area is implemented for the first time in silicon-based technology including standard digital CMOS process and SiGe HBT process. Furthermore, a linearization method using the cancellation of third-order intermodulation terms of parallel FETs known as derivative superposition [72] is combined to enhance the linearity of CMOS power amplifier. Altogether, these two approaches will demonstrate a better solution for advanced linear and high-efficiency RF power amplifier IC design.

5.2 Analysis of the Device Periphery Adjustment Technique

The main consideration of this approach is the adjustment of device periphery to reduce dc consumption while maintaining maximum device performances. In small-signal circuits, figures of merit such as the transition frequency (f_T) – frequency where short-circuit current gain equals one – or the maximum oscillating frequency (f_{max}) – the frequency where conjugate-matched power gain (maximum available power gain) equals one – are used to justify the performance of a transistor. However, it is difficult to find a figure of merit for power amplifier design since the power amplifier is matched for maximum output power. Roughly, f_{max} may be used in the driver stage design as

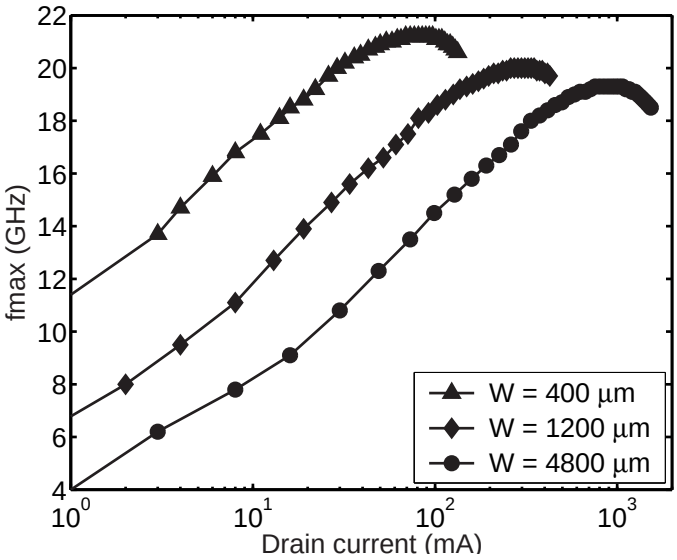


Figure 58: f_{max} of various MOS devices with different total gate width. These results include effects of pad and interconnect parasitics. (L = 0.4 μm)

the signal level is not too high and the matching design usually assumes conjugate-matched condition. In the output stage, load-pull measurements or simulations are used in practice to find a suitable device for a particular output power and bias level. For this design, f_{max} is used initially as a figure of merit and load-pull measurements will be used to determined the device size for a desired output power and efficiency.

f_{max} of a MOS transistor is given by [35]

$$f_{max} \approx \sqrt{\frac{f_T}{2\pi R_g C_{gd}}} \tag{87}$$

where f_T can be found from the following relationship

$$f_T = \frac{g_m}{2\pi C_{gs} C_{gd}} \tag{88}$$

Usually f_{max} is larger than f_T in many cases since R_g can be reduced greatly by careful layout techniques and silicided gate. f_{max} profiles of MOS and SiGe HBT devices used in this work are shown in Fig. 58 and 59. For bipolar transistors, base resistance (R_b) and base-collector capacitance (C_{bc}) are used instead of R_b and

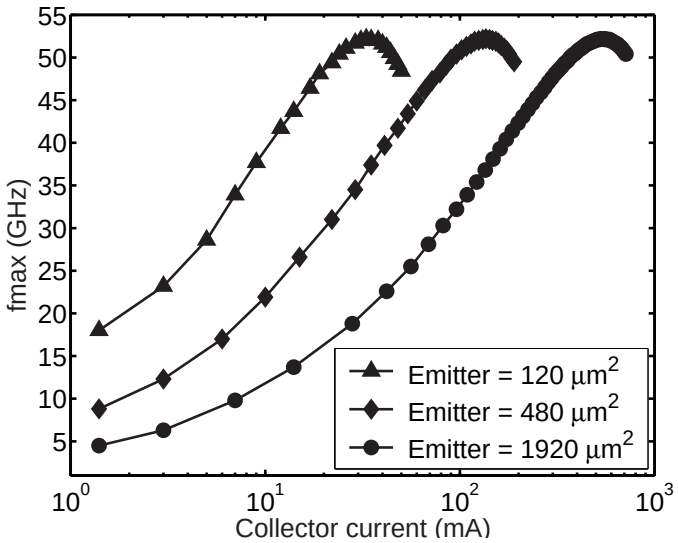


Figure 59: f_{max} of various SiGe HBT devices with different total emitter area (pad and interconnect parasitics excluded).

C_{gd}. It is shown that a transistor exhibits a peak in f_{max} only in a small range of biasing current (or current density per area). It is advisable to maintain the device current density at the optimum (peak f_{max} range) to maximize device performance. Therefore, to reduce the dc current for power amplifiers operated in low-power mode, lowering the device periphery is better than adjusting the dc current of the same device.

The next step is to realize this concept in the MOS and SiGe HBT power amplifier designs. In the MOS case, the design has two separate FET devices with different W/L ratio connected in parallel at the output stage as shown in Fig. 60. The device with smaller W/L ratio is intended for an operation in low-power mode, where the larger device is operated together with the small device in the high-power mode. This can reduce dc consumption without reducing device performance such as g_m and f_{max} since the current densities of devices in both cases are unaltered. Each transistor is terminated with an optimum output impedance termination for maximizing efficiency

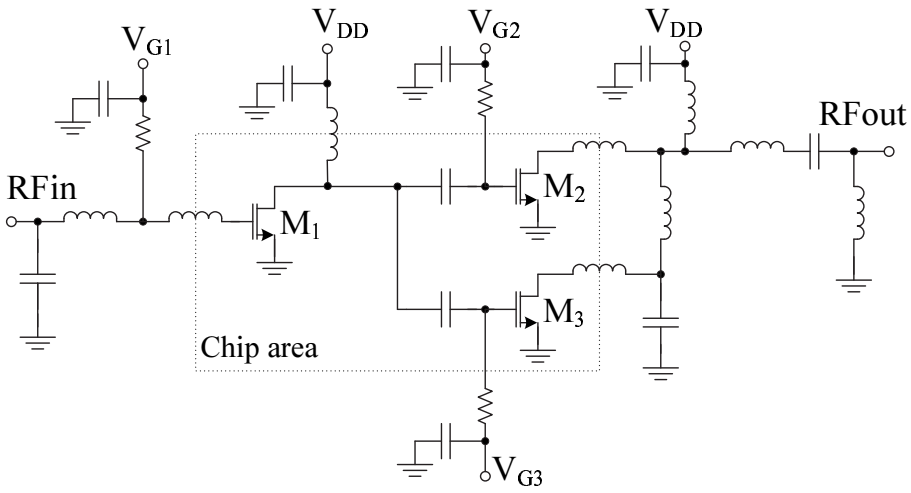


Figure 60: Simplified schematic diagram of the proposed CMOS power amplifier with 2-step power control. (W/L: M_1 = 1200 μm/0.4 μm; M_2 = 4800 μm/0.4 μm; M_3 = 1200 μm/0.4 μm)

at the desired output power. To account for the effects of pad parasitics and bond-wires, load-pull measurements are performed to obtain realistic matching conditions. The same design method is also used with the SiGe HBT case.

A few practical concerns are the phase difference between two signal paths (one path through the large transistor at the output, another path from the smaller transistor connected in parallel), signal loss due to the loading of devices in parallel, and stability issues. The interstage matching and the output matching need to transform optimum impedances to each transistor as well as minimize phase difference that causes distortion or signal cancellation. The output impedance of the output-stage devices must be taken into account when designing output matching network as it becomes a load to another transistor. Also, the circuit must be properly grounded and bypassed since it is prone to oscillation due to the parasitics from additional parallel devices.

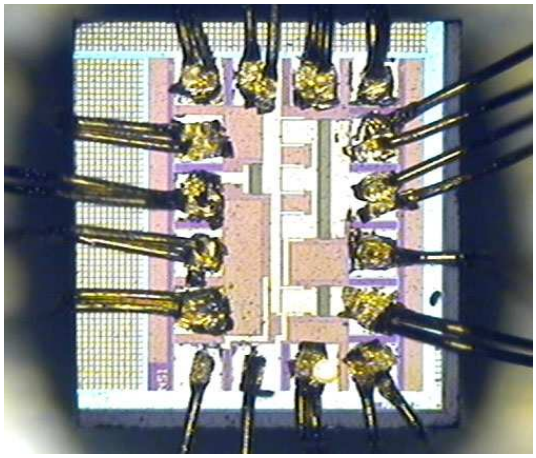


Figure 61: Chip photograph of the 1.9 GHz CMOS power amplifier using the proposed technique.

5.3 1.9 GHz CMOS Power Amplifier Design

The prototype power amplifier in Fig. 60 is designed and implemented at 1.9 GHz using standard CMOS-9 0.18-μm process. The thick-oxide devices (L = 0.4 μm) are used to improve drain-source breakdown voltage. The drain voltage (V_{DD}) of 2.4 V is used in both stages. The driver-stage transistor is designed with total gate width of 1200 μm (M_1) where the output-stage transistors have the size of 4800 μm (M_2) and 1200 μm (M_3). The total die area is approximately 0.55 x 0.7 mm^2 including pads with the chip photograph shown in Fig. 61 and the board layout shown in Fig. 62. The matching networks are designed externally except the interstage-matching capacitors which are on-chip, poly-poly capacitors. Load-pull measurements are used to find optimum matching impedances of each transistor. All inductors are realized using bondwires, trace lines, and off-chip SMT inductors to minimize loss. The final schematic including the bias networks and component values is shown in Fig. 63.

The bias network must be carefully designed to avoid the parasitic oscillation. Silicon-based RF power amplifiers are easier to oscillate because of the lower substrate resistivity, which increases the reactance of the parasitic capacitance, compared to

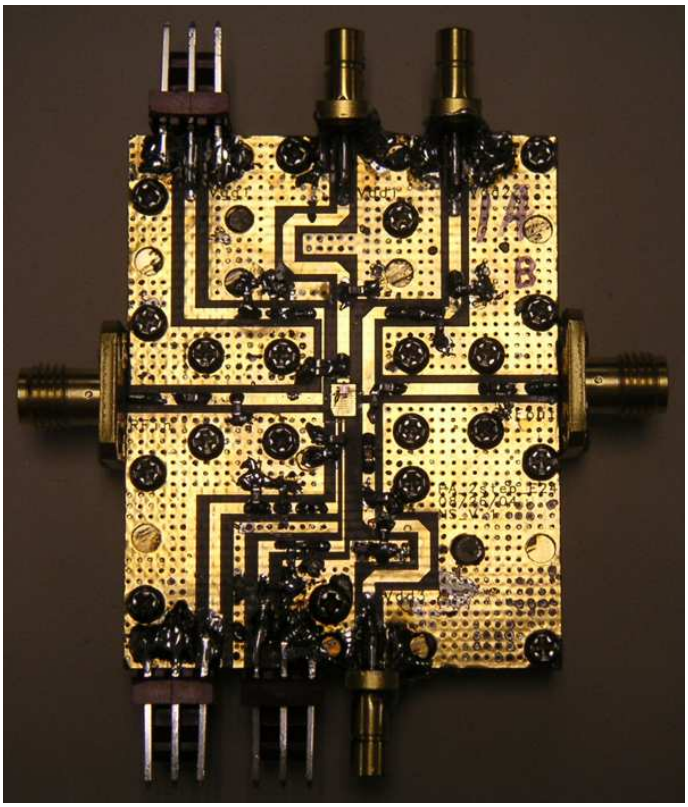


Figure 62: PCB layout of the prototype 1.9 GHz CMOS power amplifier using the proposed technique.

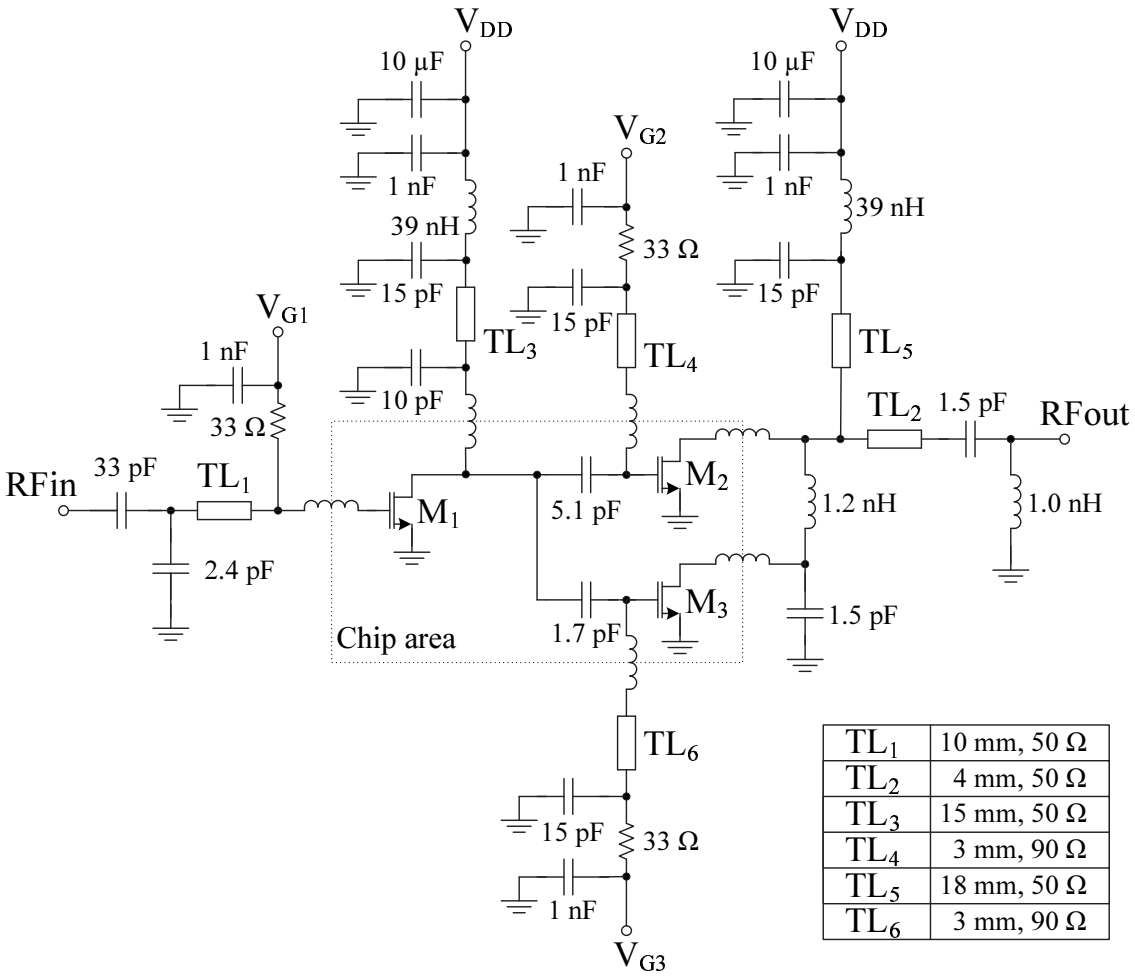


Figure 63: Detailed schematic diagram of the prototype 1.9 GHz CMOS power amplifier. (W/L: $M_1 = 1200\ \mu m/0.4\ \mu m$; $M_2 = 4800\ \mu m/0.4\ \mu m$; $M_3 = 1200\ \mu m/0.4\ \mu m$)

other technologies such as GaAs process. Parasitic oscillation normally happens in low-frequency region (kHz range to a few hundred MHz) where the gain of the circuit is high. Therefore, it is important to avoid the forming of a low-frequency tank circuit, which normally occurs around the RF chokes. In this design, low value RF inductors (realized by transmission lines TL_2 and TL_3) are connected to the drain of the transistors and then bypassed by small capacitors (22 pF). These inductors present an impedance of around 5-8 times higher than the impedance transferred from matching networks, which is enough to block RF signal. This inductance is low enough in low-frequency range and will not resonate with the parasitic capacitance at the drain node. Then, higher value of RF chokes can be connected to provide more isolation between RF and dc. Large capacitors (1 nF and 10 μF) are for bypassing low frequency signals from external circuits and creating low baseband-impedance (reactance) paths looking back from the PA circuit to the biasing networks, which help minimizing the IMD asymmetry [16, 45].

The output matching network transforms the 50-ohm load down to 5 - j5 ohms, which is the optimum load for the 4800 μm device (M_2) to achieve 23 dBm output power at 1-dB compression (P_{1dB}). An additional L-network is connected from the main output matching to create an optimum load of 8 - j10 ohms for M_3, which is designed to deliver 17 dBm of output power at the P_{1dB}. Practically, the output impedance of each transistor in the output stage becomes an undesirable load to another transistor and reduces the output power to the 50-ohm load. The output impedance of M_2 is found to be 7 - j8 ohms from the device model and greatly affects the matching network design for M_3. Nevertheless, the output impedance of M_3 has a much higher value than the optimum load for M_2 that results in less loading effect.

The CMOS power amplifier is designed at 1.9 GHz to have a P_{1dB} of 23 dBm. The gain of the amplifier in low-power and high-power modes are approximately 16 dB and 21 dB, respectively, as shown in Fig. 64. PAE in the low-power region, shown

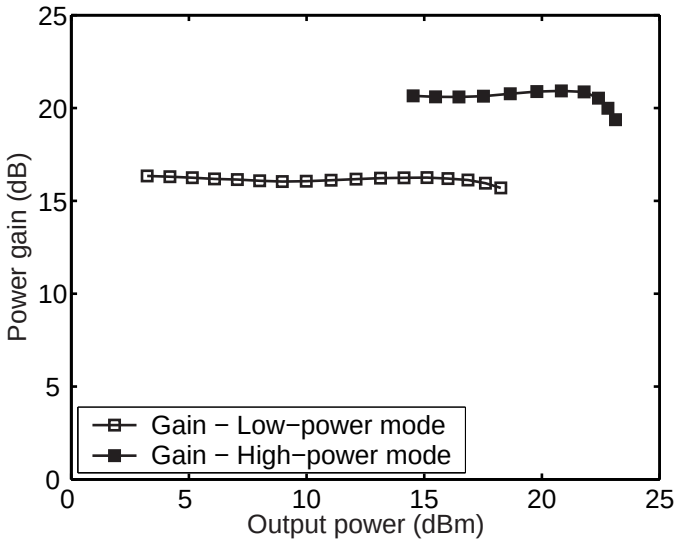


Figure 64: Power gain of the CMOS power amplifier at 1.9 GHz.

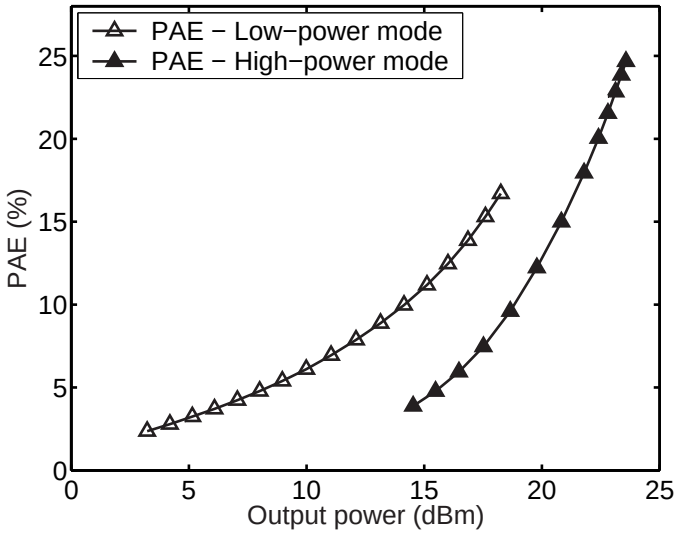


Figure 65: PAE of the CMOS power amplifier at 1.9 GHz.

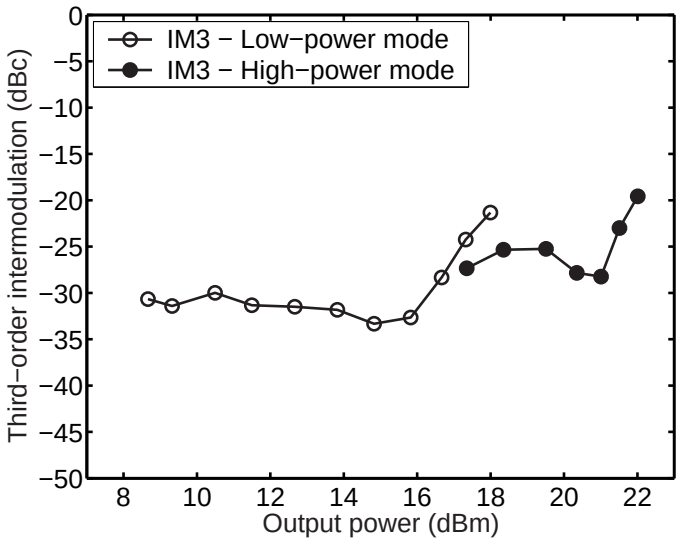


Figure 66: Third-order intermodulation distortion levels of the CMOS power amplifier with two-tone input signals at 1.88 GHz and 1.8801 GHz (100 kHz signal separation.)

in Fig. 65, is substantially improved as a result of reducing the dc consumption while providing good match at low power level. The PAE in low-power mode at the 6-dB back-off level is improved to 15% which is about a factor of 2 improvement compared to a normal class-AB amplifier (by operating all transistors in the design as in the high-power mode), while the PAE improvement is greater than a factor of 2.5 in lower power levels. The IM3 of power amplifier in low and high power levels is shown in Fig. 66.

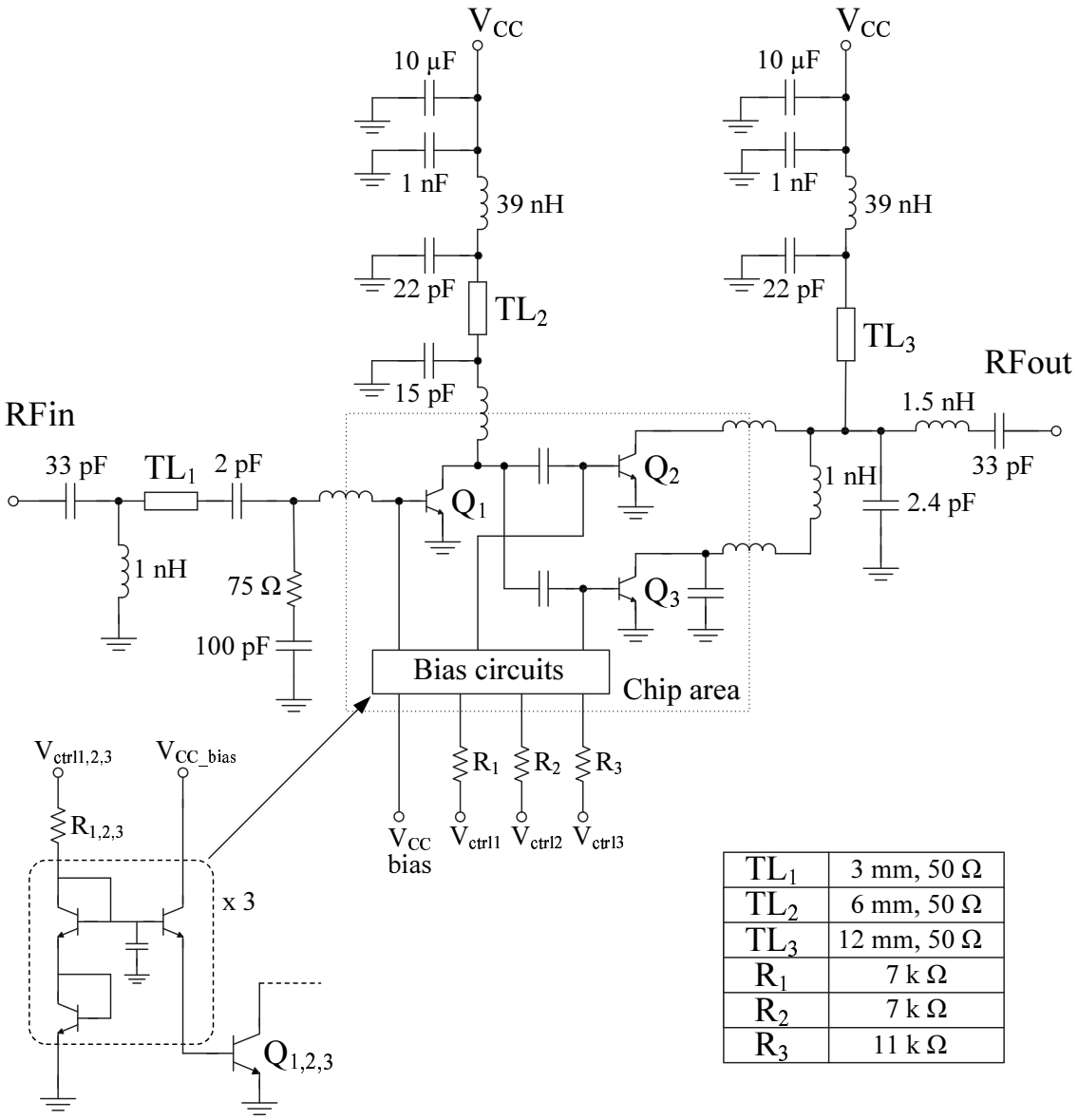
5.4 1.9 GHz CDMA/PCS SiGe HBT Power Amplifier

A 1.9 GHz SiGe HBT power amplifier for CDMA/PCS application is also designed based on the same concept. The design consists of two stages with on-chip bias networks. The emitter area of the first stage is 480 μm^2. The second stage has two transistors with emitter areas of 480 μm^2 and 1920 μm^2. The input and output matching networks are mostly designed off-chip using lumped components and

transmission lines. Components such as interstage-matching capacitors and output-matching capacitors of the small device in the second stage are realized on the chip by metal-insulator-metal (MIM) capacitors. The detailed schematic is shown in Fig. 67. The total die size is about 1 mm^2 and the chip package size is 4 mm $\times$ 4 mm using the 16-pin leadless package. The die photograph and prototype board layout are shown in Fig. 68 and Fig. 69.

The input matching is designed to provide an optimum match of 4.0 - j12.6 ohms for the first stage. From the schematic in Fig. 67, the 75-ohms resistor and 100-pF capacitor connected in series at the input serve as a decoupling network to reduce low-frequency gain and prevent low-frequency parasitic oscillation. The output matching network is designed to create an optimum matching impedance of 9.8 - j12.2 ohms for the large transistor and an additional L-section (the 1-nH inductor plus bondwire and the on-chip capacitor) is inserted to match the small transistor which requires 15.4 + j14.8 ohms optimum impedance at 18-dBm output power. Interstage matching is design with L-section topology where the inductor is realized by the bondwire that also serves as dc bias path. The on-chip bias circuits are designed using two diode-connected transistors to bias the V_{be} of the power transistor and the transistor that supplies base current for the power transistor. This topology is suitable for class-AB bias where base current changes in a wide range and provides high impedance for blocking RF frequency to enter the bias circuit.

The prototype SiGe HBT PA is tested at 1.9 GHz for CDMA/PCS IS-95 standard. The design achieves 1-dB output compression at 27 dBm with a maximum PAE of 27% in high-power mode. The power amplifier is switched to low-power mode when output power is smaller than 18 dBm. In low-power mode, the maximum PAE is 15%, which exhibits about twice the performance of a design that has single transistor in the output stage (see Fig. 70). The design provides a power gain of 20 dB in high-power mode and 18 dB in low-power mode as shown in Fig. 71. The linearity is measured



TL_1	3 mm, 50 Ω
TL_2	6 mm, 50 Ω
TL_3	12 mm, 50 Ω
R_1	7 k Ω
R_2	7 k Ω
R_3	11 k Ω

Figure 67: Detailed schematic diagram of the prototype 1.9 GHz SiGe CDMA power amplifier. (Emitter area: $Q_1 = 480\ \mu m^2$; $Q_2 = 1920\ \mu m^2$; $Q_3 = 480\ \mu m^2$)

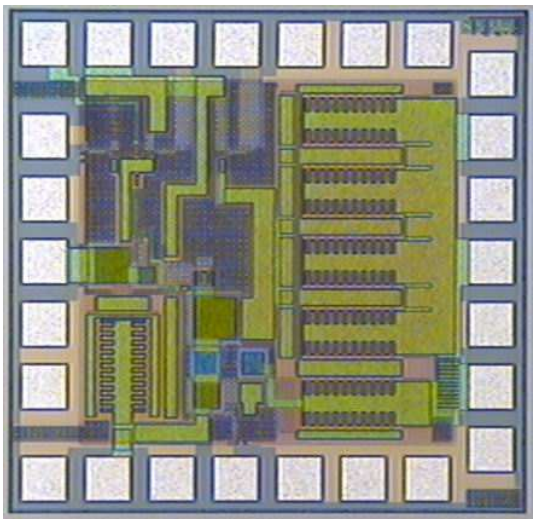


Figure 68: Chip photograph of the 1.9 GHz SiGe CDMA power amplifier using the proposed technique.

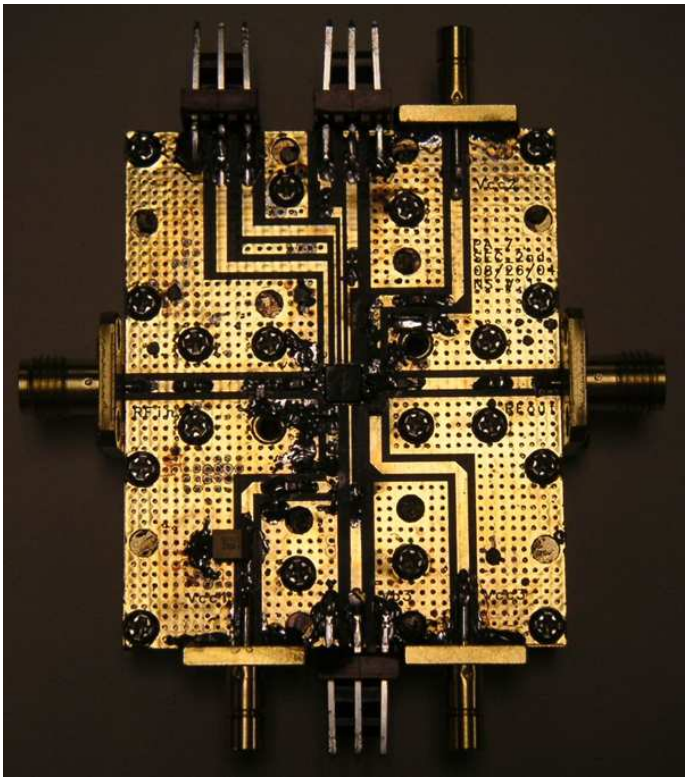


Figure 69: PCB layout of the prototype 1.9 GHz SiGe CDMA power amplifier using the proposed technique.

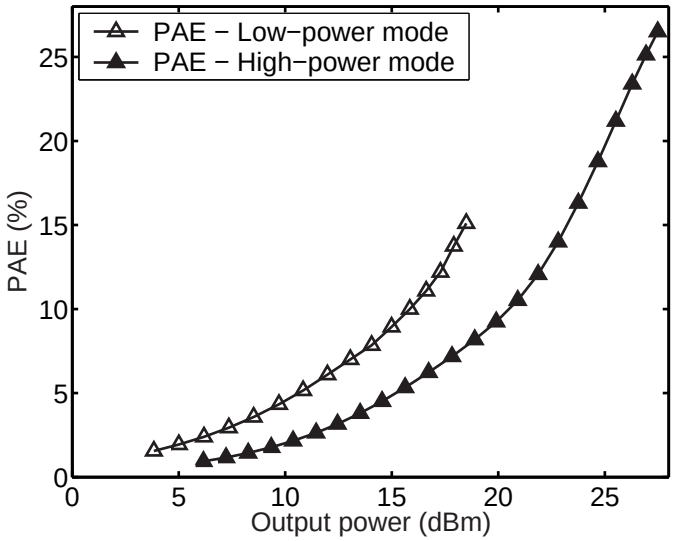


Figure 70: PAE of the SiGe power amplifier at 1.9 GHz.

in terms of the adjacent and the alternate channel power leakage ratios (ACPR1 and ACPR2) where the frequency offset from the carrier is 1.25 MHz for the adjacent channel and 2.25 MHz for the alternate channel. The specification limits these levels to be lower than 44 dBc and 53 dBc for the ACPR1 and the ACPR2, respectively. The measurement results in Fig. 72 show that the design meets the specification up to 16-dBm output power in low-power mode and 25-dBm output power in high-power mode. The linearity level can be improved in higher output power by redesigning the circuit to have larger device sizes in the second stage. However, this current design shows that the proposed design topology that uses a smaller transistor to operate in the low-power region can improve PAE while meeting the linearity requirements.

5.5 Nonlinear Terms Cancellation for Linearity Improvement

In addition to the efficiency improvement, the FET amplifier can be linearized by the cancellation of nonlinear terms. If the biases of parallel devices are carefully optimized, the nonlinear products will have opposite polarities that result in the

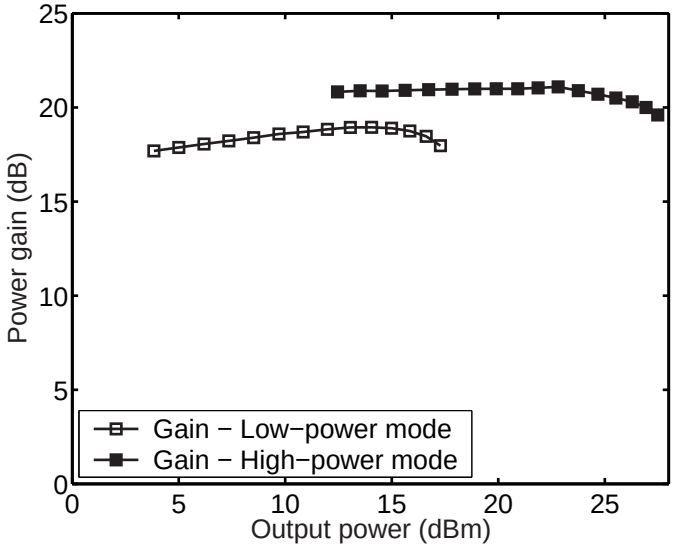


Figure 71: Power gain of the SiGe power amplifier at 1.9 GHz.

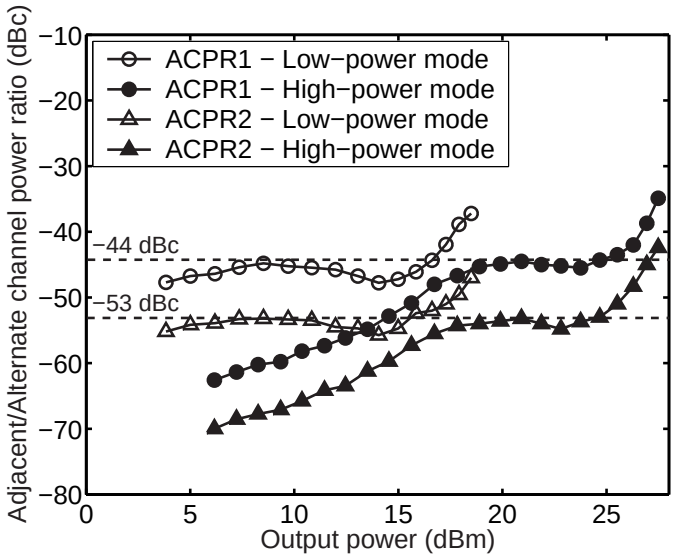


Figure 72: Adjacent and alternate channel power leakage ratios of the SiGe CDMA power amplifier tested for IS-95 standard at 1.9 GHz.

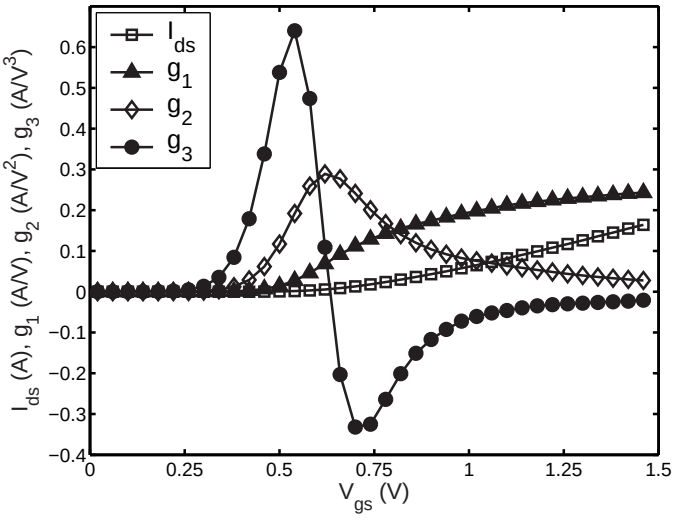


Figure 73: Small-signal components of drain current (i_d) including transconductance (g_1), and higher order terms (g_2 and g_3) of N-MOSFET device.

reduction of output intermodulation distortion (IMD). This can be explained from the Taylor series expansion of the drain current, given by

$$\begin{aligned} i_d &= \left.\frac{dI_d}{dV_g}\right|_{V=V_{g,0}} v_g + \frac{1}{2}\left.\frac{d^2I_d}{dV_g^2}\right|_{V=V_{g,0}} v_g^2 + \frac{1}{6}\left.\frac{d^3I_d}{dV_g^3}\right|_{V=V_{g,0}} v_g^3 + \ldots \\ &= g_1 v_g + g_2 v_g^2 + g_3 v_g^3 + \ldots \end{aligned} \tag{89}$$

At different gate bias levels, g_3 may have opposite signs; a positive sign below threshold voltage (V_t) and a negative beyond that point (see Fig. 73). Therefore, by offsetting the gate bias voltage of output-stage transistors, their third-order intermodulation (IM3) products will be cancelled. This cancellation, however, must occur with minimal phase difference between two signal paths.

The adjustment of gate bias voltage of the small transistor (M_3) in the output stage has shown improvement of IM3, as illustrated in Fig. 74. It is found experimentally that the V_{G3} of 0.55 V provides the most improvement in IM3 in the high-power mode at the average of 8 dB improvement from the normal design where the small transistor is biased into class-AB (with V_{G3} of 0.90 V). The results are obtained from two-tone measurements with two carrier signals at 1.88 GHz and 1.8801 GHz (100 kHz signal

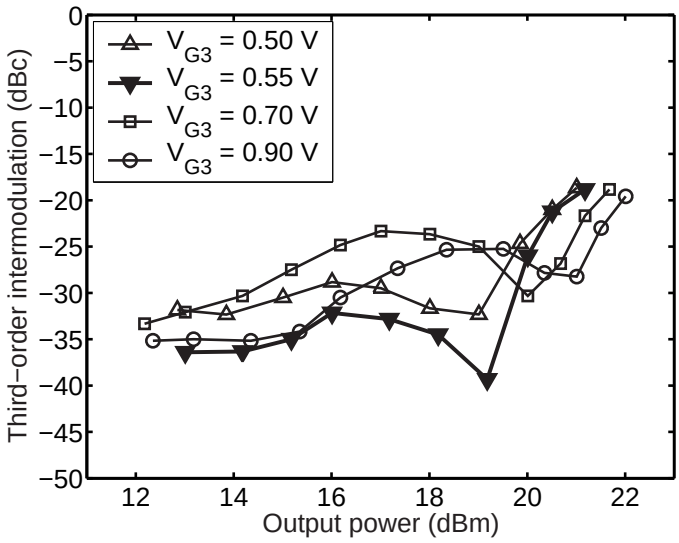


Figure 74: Third-order intermodulation distortion levels of the CMOS power amplifier with two-tone input signals at 1.88 GHz and 1.8801 GHz in high-power mode. The gate bias of M_3 is varied from 0.50 V to 0.90 V.

separation). The output spectrums of two-tone measurements with V_{G3} of 0.9 V and V_{G3} of 0.55 V are illustrated in Fig. 75 and Fig. 76, respectively, and have shown an IM3 improvement of 8 dB (from -25.17 dBc to -33.17 dBc) at the output power of 19.5 dBm. These results have proven the success of implementing the combination of PAE and linearity improvement techniques in the FET power amplifier IC design.

5.6 Summary

The device periphery adjustment technique for efficiency enhancement in linear amplifier is proposed and demonstrated. The 1.9 GHz CMOS and SiGe HBT power amplifiers having two transistors in parallel at the output stage are designed to improve efficiency in the low-power region. The designed power amplifiers using two-step efficiency improvement control achieve the PAE improvement by a factor of two or higher at both 6-dB back-off level (CMOS PA) and 9-dB back-off level (SiGe HBT PA), as compared to normal class-AB power amplifiers. In addition, linearity of the CMOS power amplifier can be improved by offsetting the gate bias voltages of the

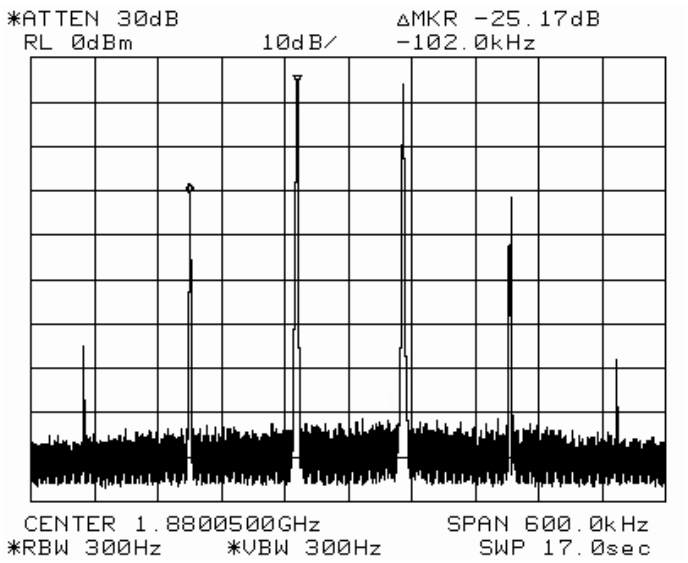


Figure 75: Spectrum of output signal at the output power of 19.5 dBm showing carrier signals, 3rd, and 5th order intermodulation products in high-power mode when V_{G3} is 0.90 V.

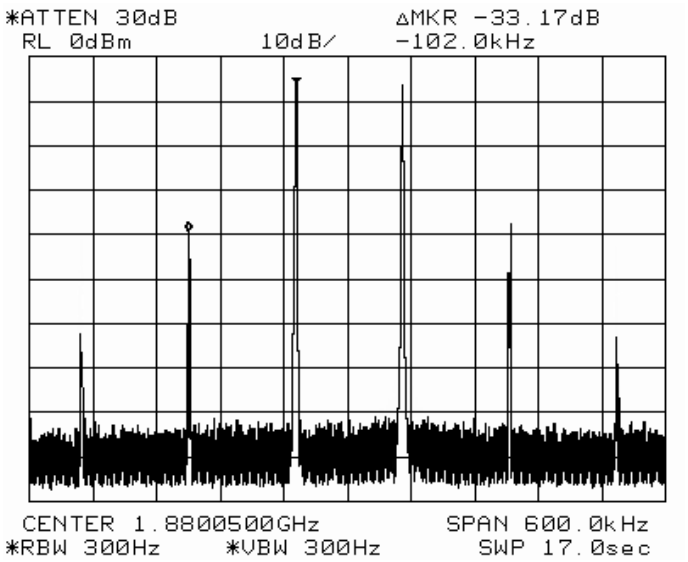


Figure 76: Spectrum of output signal at the output power of 19.5 dBm showing carrier signals, 3rd, and 5th order intermodulation products in high-power mode when V_{G3} is changed from 0.90 V to 0.55 V, resulting in 8-dB reduction in IM3.

parallel output-stage transistors to cancel their nonlinear products. This proposed method has been proven through two-tone measurements and has shown to reduce 8 dB of IM3 levels on average at high output power. This successful design method can be readily implemented on any other integrated FET power amplifiers. The device periphery adjustment technique does not need transmission lines or additional controlling circuits, thus making it ideal for power amplifier MMIC designs in the production level where size and cost are among the most important concerns.

CHAPTER 6

CONCLUSION

6.1 Contributions of the Dissertation

The work demonstrated in this thesis focuses on the efficiency improvement for linear RF power amplifiers, which is one of the well-known problems in RF circuits. In the past, many circuit design approaches were suggested and implemented in several power amplifier systems. The multistage Doherty amplifier is developed based on the concept of conventional Doherty amplifier and is newly proposed in this thesis as a better solution for high-efficiency linear RF power amplifier design. In the multistage Doherty amplifier, a number of amplifier stages are used to optimize for dc power consumption and the amount of output power. A three-stage Doherty power amplifier is designed using GaAs MESFET devices on a FR-4 board prototype to enhance the efficiency of a WCDMA power amplifier with the output power of 33 dBm. The design achieves two times and seven times improvement in efficiency at 6-dB and 12-dB back-off levels, respectively, while satisfying the adjacent and the alternate channel power leakage ratios of -33 dBm and -43 dBm. This performance shows significant improvement and proves the success when implementing the technique on current mobile communication standard. The mathematical approach presented in this work to understand the multistage Doherty amplifier concept is useful in achieving performances as close to the ideal as possible. Linearity improvement by optimal device periphery design and the bias adaptation, used for avoiding the AM/AM distortion, are also provided to further improve the design.

This dissertation also includes the design and implementation of the device periphery adjustment technique that is used for improving efficiency of RF power amplifier

ICs operated in the low-power mode. This new technique advances the previously established approaches in that the device periphery of the amplifier is adjusted for optimum transistor performances (device f_{max}, dc consumption, and output power) that, in turns, contributes to the overall efficiency improvement. The CMOS and SiGe power amplifiers operated at 1.9 GHz range are designed and have shown twice higher efficiency than the normal single-transistor amplifier designs, while meeting the linearity requirements. In addition, the use of multiple, parallel FET transistors enables the cancellation of the nonlinear terms generated from devices biased at different V_{gs} ranges. This provides a simpler linearization scheme to reduce the IM3 of a FET amplifier operated in high-power mode, in addition to the efficiency enhancement scheme in the same circuit. The implementation of this concept in the designed CMOS power amplifier has shown an improvement in IM3 at high-power mode by 8 dB on average. Moreover, the practical design guidelines of matching circuits and bias networks are made available and helpful for future research. Thus, this dissertation shows a complete detail in implementing the concept of device periphery adjustment, which is expected to be an initiative for the future development of compact, high-efficiency, high-linearity power amplifier integrated circuits.

The MOSFET model proposed in this dissertation provides a new approach in developing accurate scalable substrate network model, which can improve the accuracy of the small-signal equivalent circuit model and the BSIM3v3 model in RF and microwave frequencies. This dissertation provides guidelines in parameter extraction and a novel approach in modeling the substrate coupling effects of MOSFET devices with surrounding substrate contacts. The proposed model simplifies the complexity of substrate effect modeling and enables the development of scalable substrate network model with great accuracy. Moreover, the proposed substrate network model is geometry dependent and can be generalized for different finger width, length, and different structures of substrate ring. The scalable model in this work is developed

based on NMOS devices ranging from 200 μm to 6 mm total gate width. The results show good agreement in model scalability and S-parameter verification up to 20 GHz. Moreover, the modified BSIM3v3 model with the proposed substrate resistance network and the gate resistance performs well in comparison to the large-signal load-pull measurements. Therefore, this dissertation provides a good approach for future MOS model development, which is a significant part to the success in RF CMOS circuit design.

6.2 Future Work

The work described in this dissertation represents starting points for several future researches and developments. First, the implementation of the multistage Doherty amplifier at higher power levels is still a challenge. Further study and possibly some modifications in the output transformer are required to understand the dominating effect of the output impedance on the overall performance. In addition, the integration of multistage Doherty amplifiers for IC-level design requires the miniaturization of transmission lines and transformers. The use of equivalent π- or T-network to replace the quarter-wave transmission lines is possible when accurate and high-Q passive components are available. In addition, the implementation of bias control concept to improve the linearity performances is attractive and worthwhile for future studies.

The modeling of MOS transistor in this work is based on a set of CMOS devices in 0.18-μm process. It will become a greater challenge when the channel length starts to decrease as the IC world is moving towards the sub 0.1-μm regime. The behavior of transistor parameters can change significantly, compared to that of larger channel length process and may require new approaches in parameter extractions. Also, the accurate modeling of substrate coupling at higher frequencies (greater than 40 GHz) demands modifications for the equivalent silicon substrate network (parallel R-C structure.)

For RF power amplifier IC design, there are many challenges that are still open for future research. The ultimate goal in RF power amplifier design is to provide a solution for compact integration and cost effective design with no compromise in performances. The integration of RF power amplifier incorporating the proposed device periphery adjustment technique can reduce the overall size of the design and will enhance efficiency and linearity performances. For future design integration, an accurate and high-Q passive library in IC process is the first requirement. The implementation of high-frequency and high-power amplifier IC is still a challenging task. Modifications in device technology (to reduce output and feedback capacitances, and increase f_T, f_{max}, and breakdown voltages), as well as new circuit design approaches (transistor stacking, power combining techniques, new transmitter architectures, etc.) are necessary and will become the focus of future research in this area.

6.3 List of Publication

The work in this dissertation has been published in several IEEE conferences and journal papers as listed below.

1. N. Srirattana, P. Sen, H.-M. Park, C.-H. Lee, P. E. Allen, and J. Laskar, "Linear RF CMOS power amplifier with improved efficiency and linearity in wide power levels," to be presented at *2005 IEEE Radio Frequency Integrated Circuits Symp.*

2. N. Srirattana, A. Raghavan, D. Heo, P. E. Allen, and J. Laskar, "Analysis and design of a high-efficiency multistage Doherty power amplifier for wireless communications," *IEEE Trans. Microwave Theory Tech.*, vol. 53, no. 3, Mar. 2005.

3. N. Srirattana, D. Heo, H.-M. Park, A. Raghavan, P. E. Allen, and J. Laskar, "A new analytical scalable substrate network model for RF MOSFETs," in *2004 IEEE*

MTT-S Int. Microwave Symp. Dig., pp. 699-702.

4. N. Srirattana, D. Heo, A. Raghavan, K. Lim, P. E. Allen, and J. Laskar, "A scalable small-signal model for MOSFET including substrate network," in *Proc. Asia-Pacific Microwave Conf.*, Seoul, South Korea, Nov. 2003, pp.1203-1206.

5. N. Srirattana, A. Raghavan, D. Heo, P. E. Allen, and J. Laskar, "Analysis and design of a high-efficiency multistage Doherty power amplifier for WCDMA," in *Proc. IEEE Eur. Microwave Conf.*, Munich, Germany, Oct. 2003, pp. 1337-1340.

6. N. Srirattana, A. Raghavan, D. Heo, P. E. Allen, and J. Laskar, "Development of the multistage Doherty power amplifier for wireless communications," in *Proc. IEEE Topical Workshop on Power Amplifier for Wireless Communications*, San Diego, CA, Sep. 2003.

7. N. Srirattana, A. Raghavan, D. Heo, P. E. Allen, and J. Laskar, "A high-efficiency multistage Doherty power amplifier for WCDMA," in *Proc. IEEE Radio and Wireless Conf.*, Boston, MA, Aug. 2003, pp. 397 - 400.

8. N. Srirattana, M. S. Qureshi, A. Aude, V. Krishnamurthy, D. Heo, P. E. Allen, and J. Lasker, "SiGe HBT power amplifier for IS-95 CDMA using a novel process, voltage, and temperature insensitive biasing scheme," in emphProc. IEEE Int. Symp. Circuits Syst. (ISCAS'03), Bangkok, Thailand, May 2003, vol. I, pp. 437 - 440.

REFERENCES

[1] Angelov, I., Choumei, K., and Inoue, A., "An empirical HBT large signal model for CAD," in *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 1237–1240, June 2002.

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

[3] Bae, S., Kim, J., Nam, I., and Kwon, Y., "Bias-switching quasi-Doherty-type amplifier for CDMA handset applications," in *IEEE Radio Frequency Integrated Circuits Symp. Dig.*, pp. 131–140, 2003.

[4] Bagheri, M. and Tsividis, Y., "A small signal dc-to-high-frequency nonquasistatic model for the four-terminal MOSFET valid in all regions of operation," *IEEE Trans. Electron Devices*, vol. 32, pp. 2383–2391, November 1985.

[5] Bousnina, S. and Ghannouchi, F. M., "Analysis and experimental study of an L-band new topology Doherty amplifier," in *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 935–938, 2001.

[6] Campbell, C. F., "A fully integrated *Ku*-band Doherty amplifier MMIC," *IEEE Microwave Guided Wave Lett.*, vol. 9, pp. 114–116, March 1999.

[7] Cha, J., Yang, Y., Shin, B., and Kim, B., "An adaptive bias controlled power amplifier with a load-modulated combining scheme for high efficiency and linearity," in *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 81–84, 2003.

[8] Chai, K.-W. and Paulos, J. J., "Unified nonquasi-static modeling of the long-channel four-terminal MOSFET for large- and small-signal analyses in all operating ranges," *IEEE Trans. Electron Devices*, vol. 36, pp. 2513–2520, November 1989.

[9] Chan, Y. J., Huang, C. H., Weng, C. C., and Liew, B. K., "Characteristics of deep-submicrometer MOSFET and its emprical nonlinear RF model," *IEEE Trans. Microwave Theory Tech.*, vol. 46, pp. 611–615, May 1998.

[10] Chang, R. T., Yang, M.-T., Ho, P. P. C., Wang, Y.-J., Chia, Y.-T., Liew, B.-K., Yue, C. P., , and Wong, S. S., "Modeling and optimization of substrate resistance for RF-CMOS," *IEEE Trans. Electron Devices*, vol. 51, pp. 421–426, March 2004.

[11] Cheng, Y. and Matloubian, M., "On the high-frequency characteristics of substrate resistance in RF MOSFETs," *IEEE Electron Device Lett.*, vol. 21, pp. 604–606, December 2000.

[12] Cheng, Y. and Matloubian, M., "Parameter extraction of accurate and scalable substrate resistance components in RF MOSFETs," *IEEE Electron Device Lett.*, vol. 23, pp. 221–223, April 2002.

[13] Chireix, H., "High power outphasing modulation," *Proc. IRE*, vol. 23, no. 11, pp. 1370–1392, 1935.

[14] Cripps, S. C., *RF Power Amplifiers for Wireless Communications*. Norwood, MA: Artech House, 1999.

[15] Cripps, S. C., *Advanced Techniques in RF Power Amplifier Design*. Norwood, MA: Artech House, 2002.

[16] de Carvalho, N. B. and Pedro, J. C., "Large- and small-signal IMD behavior of microwave power amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 47, pp. 2364–2374, December 1999.

[17] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, *BSIM3 Version 3.0 Manual*, 1996.

[18] Doherty, W. H., "A new high efficiency power amplifier for modulated waves," *Proc. IRE*, vol. 24, pp. 1163–1182, September 1936.

[19] Enz, C. C. and Cheng, Y., "MOS transistor modeling for RF IC design," *IEEE J. Solid-State Circuits*, vol. 35, pp. 186–201, February 2000.

[20] Fowler, T., Burger, K., Cheng, N.-S., Samelis, A., Enobakhare, E., and Rohlfing, S., "Efficiency improvement techniques at low power levels for linear CDMA and WCDMA power amplifiers," in *IEEE Radio Frequency Integrated Circuits Symp. Dig.*, pp. 41–44, 2002.

[21] Han, J. and Shin, H., "A scalable model for the substrate resistance in multi-finger RF MOSFETs," in *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 3, pp. 2105–2108, 2003.

[22] Hanington, G., Chen, P. F., and Asbeck, P. M., "A 10 mhz dc to dc converter for microwave power amplifier efficiency improvement," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 1998.

[23] Hanington, G., Chen, P.-F., Asbeck, P. M., and Larson, L. E., "High-efficiency power amplifier using dynamic power-supply voltage for CDMA applications," *IEEE Trans. Microwave Theory Tech.*, vol. 47, pp. 1471–1476, August 1999.

[24] Ho, M. C., Green, K., Culbertson, R., Yang, J. Y., Ladwig, D., and Ehnis, P., "A physical large signal Si MOSFET model for RF circuit design," in *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 391–394, June 1997.

[25] Iwamoto, M., Williams, A., Chen, P.-F., Metzger, A., Larson, L., and Asbeck, P., "An extended Doherty amplifier with high efficiency over a wide power range," *IEEE Trans. Microwave Theory Tech.*, vol. 49, pp. 2472–2479, December 2001.

[26] Jen, S. H.-M., Enz, C. C., Pehlke, D. R., Schroter, M., , and Sheu, B. J., "Accurate modeling and parameter extraction for MOS transistors valid up to 10 GHz," *IEEE Trans. Electron Devices*, vol. 46, pp. 2217–2227, November 1999.

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

[28] Kenington, P. B., *High-Linearity RF Amplifier Design*. Norwood, MA: Artech House, 2000.

[29] Kim, J. H., Kim, J. H., Noh, Y. S., Kim, Y. S., Kim, S. G., and Park, C. S., "An MMIC smart power amplifier of 21% PAE at 16 dBm power level for W-CDMA mobile communication terminals," in *Proc. IEEE GaAs IC Symp.*, pp. 181–184, 2002.

[30] Kim, J. H., Kim, K. Y., Choi, Y. H., and Park, C. S., "A power efficient W-CDMA smart power amplifier with emitter area adjusted for output power levels," in *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 1165–1168, 2004.

[31] Kobayashi, K. W., Oki, A. K., Guitierrez-Aitken, A., Chin, P., Yang, L., Kaneshiro, E., Grossman, P. C., Sato, K., Block, T. R., Yen, H. C., and Streit, D. C., "An 18–21 Ghz InP DHBT linear microwave Doherty amplifier," in *IEEE Radio Frequency Integrated Circuits Symp. Dig.*, pp. 179–182, 2000.

[32] Kraus, H. L., Bostian, C. W., and Raab, F. H., *Solid State Radio Engineering*. New York: John Wiley, 1980.

[33] Lee, S., "A small-signal rf model and its parameter extraction for substrate effects in RF MOSFETs," *IEEE Trans. Electron Devices*, vol. 48, pp. 1374–1379, July 2001.

[34] Lee, S., "Direct extraction of substrate parameters for the small-signal model of a RF MOSFET," in *Proc. Asia- Pacific Microwave Conf.*, vol. 1, pp. 346–349, November 2003.

[35] Lee, T. H., *The design of CMOS radio-frequency integrated circuits*. New York: Cambridge University Press, 1998.

[36] Lees, J., Goss, M., Benedikt, J., and Tasker, P. J., "Single-tone optimisation of an adaptive-bias Doherty structure," in *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 2213–2216, 2003.

[37] Liu, W., "RF MOSFET modeling accounting for distributed substrate and channel resistances with emphasis on the BSIM3v3 SPICE model," in *IEDM Tech Dig.*, pp. 309–312, 1997.

[38] McCarroll, C. P., Alley, G. D., Yates, S., and Matreci, R., "A 20 Ghz Doherty power amplifier MMIC with high efficiency and low distortion designed for broad band digital communication systems," in *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 537–540, June 2000.

[39] McMorrow, R. J., Upton, D. M., , and Maloney, P. R., "The microwave Doherty amplifier," in *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 1653–1656, June 1994.

[40] Nellis, K. and Zampardi, P. J., "A comparison of linear handset power amplifiers in different bipolar technologies," *IEEE J. Solid-State Circuits*, vol. 39, pp. 1746–1754, October 2004.

[41] Ou, J.-J., Jin, X., Ma, I., Hu, C., and Gray, P., "CMOS RF modeling for GHz communication ICs," in *VLSI Technology Symp.*, pp. 94–95, June 1998.

[42] Park, H. J., Ko, P. K., and Hu, C., "A nonquasi-static MOSFET model for SPICE-transient analysis," *IEEE Trans. Electron Devices*, vol. 36, pp. 561–576, March 1989.

[43] Park, H. J., Ko, P. K., and Hu, C., "A charge conserving nonquasi-static (NQS) model for SPICE transient analysis," *IEEE Trans. Computer-Aided Design*, vol. 10, pp. 629–642, May 1991.

[44] Park, H.-M., Cheon, S.-H., Park, J.-W., and Hong, S., "Demonstration of on-chip appended power amplifier for improved efficiency at low power region," in *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 2, pp. 691–694, June 2003.

[45] Park, H.-M. and Hong, S., "Characterization and modeling of intermodulation distortion asymmetry in HBT using large-signal model," in *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 773–776, 2003.

[46] Pehlke, D., "High frequency application of MOS compact models and their development for scalable RF model libraries," in *Proc. IEEE Custom Integrated Circuit Conf.*, pp. 219–222, MAY 1998.

[47] Philips, *MOS MODEL 9 documentation.* `http://www.semiconductors.philips.com/Philips_Models/mos_models/model9/`.

[48] Pozar, D. M., *Microwave Engineering.* New Jersey: John Wiley & Son, 1999.

[49] Raab, F. H., "Idealized operation of the class E tuned power amplifier," *IEEE Trans. Circuits and Systems*, vol. CAS-24, December 1977.

[50] Raab, F. H., "Efficiency of outphasing RF power amplifier systems," *IEEE Trans. Commun.*, vol. 33, pp. 1094–1099, October 1985.

[51] Raab, F. H., "Efficiency of Doherty RF power-amplifier systems," *IEEE trans. Broadcast.*, vol. BC-33, pp. 77–83, September 1987.

[52] Raab, F. H., "Class-F power amplifiers with maximally flat waveforms," *IEEE Trans. Microwave Theory Tech.*, vol. 45, pp. 2007–2011, November 1997.

[53] Raab, F. H., Asbeck, P., Cripps, S., Kenington, P. B., Popovic, Z. B., Pothecary, N., Sevic, J. F., and Sokal, N. O., "Power amplifiers and transmitters for RF and microwave," *IEEE Trans. Microwave Theory Tech.*, vol. 50, pp. 814–826, March 2002.

[54] Raab, F. H., Sigmon, B. E., Myers, R. G., and Jackson, R. M., "L-band transmitter using kahn EER technique," *IEEE Trans. Microwave Theory Tech.*, vol. 46, pp. 2220–2225, December 1998.

[55] Razavi, B., Yan, R. H., and Lee, K. F., "Impact of distributed gate resistance on the performance of MOS devices," *IEEE Trans. Circuits Syst. I*, vol. 41, pp. 750–754, November 1994.

[56] Rollett, J. M., "Stability and power-gain invariants of linear twoports," *IRE Trans. on Circuit Theory*, pp. 29–32, March 1962.

[57] Rudolph, M., Doerner, R., Beilenhoff, K., and Heymann, P., "Scalable GaInP/GaAs HBT large-signal model," *IEEE Trans. Microwave Theory Tech.*, vol. 48, December 2000.

[58] Singh, R., Juge, A., Joly, R., and Morin, G., "An investigation into the nonquasi-static effects in MOS devices with on-wafer S-parameter techniques," in *Proc. IEEE ICMTS*, pp. 21–25, 1993.

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

[60] Srirattana, N., Heo, D., Park, H.-M., Raghavan, A., Allen, P. E., and Laskar, J., "A new analytical scalable substrate network model for RF MOSFETs," in *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 699–702, 2004.

[61] Srirattana, N., Heo, D., Raghavan, A., Lim, K., Allen, P. E., and Laskar, J., "A scalable small-signal model for MOSFET including substrate network," in *Proc. Asia-Pacific Microwave Conf.*, vol. 2, pp. 1203–1206, November 2003.

[62] Srirattana, N., Qureshi, M. S., Aude, A., Krishnamurthy, V., Heo, D., Allen, P. E., and Lasker, J., "Sige HBT power amplifier for IS-95 CDMA using a novel process, voltage, and temperature insensitive biasing scheme," in *IEEE Int. Symp. Circuits Syst.*, vol. I, pp. 437–440, May 2003.

[63] Srirattana, N., Raghavan, A., Heo, D., Allen, P. E., and Laskar, J., "Analysis and design of a high-efficiency multistage Doherty power amplifier for WCDMA," in *Proc. IEEE Eur. Microwave Conf.*, pp. 1337–1340, October 2003.

[64] Staudinger, J., "Applying switched gain stage concepts to improve efficiency and linearity for mobile CDMA power amplification," *Microwave J.*, vol. 43, pp. 152–162, September 2000.

[65] Staudinger, J., "Stage bypassing in multi-stage PAs," presented at *IEEE MTT-S Symp. Workshop on Efficiency & Linearity Enhancement Methods for Portable RF/MW Power Amplifier*, June 2000.

[66] Technology Modeling Associates, *Medici*, 1997.

[67] Tin, S. F., Osman, A. A., Mayaram, K., and Hu, C., "A simple subcircuit extension of the BSIM3v3 model for CMOS RF design," *IEEE J. Solid-State Circuits*, vol. 35, pp. 612–624, April 2000.

[68] Tsividis, Y. P., *Operation and modeling of the MOS transistor*. New York: McGraw-Hill Book Company, 1987.

[69] Turchetti, C., Mancini, P., and Masetti, G., "A CAD-oriented nonquasistatic approach for transient analysis of MOS ICs," *IEEE J. Solid-State Circuits*, vol. SC-21, pp. 827–835, October 1986.

[70] Upton, D. M., "A new circuit topology to realize high efficiency, high linearity, and high power microwave amplifiers," in *Proc. Radio and Wireless Conf.*, pp. 317–320, 1998.

[71] Vandeloo, P. J. V. and Sansen, W. M. C., "Modeling of the MOS transistor for high frequency analog design," *IEEE Trans. Computer-Aided Design*, vol. 8, pp. 713–723, July 1989.

[72] Webster, D. R., Ataei, G., and Haigh, D. G., "Low-distortion MMIC power amplifier using a new form of derivative superposition," *IEEE Trans. Microwave Theory Tech.*, vol. 49, pp. 328–332, February 2001.

[73] Yang, Y., Cha, J., Shin, B., and Kim, B., "A fully matched N-way Doherty amplifier with optimized linearity," *IEEE Trans. Microwave Theory Tech.*, vol. 51, pp. 986–993, March 2003.

[74] Yang, Y., Cha, J., Shin, B., and Kim, B., "A microwave Doherty amplifier employing envelope tracking technique for high efficiency and linearity," *IEEE Microwave and Wireless Comp. Lett.*, vol. 13, pp. 370–372, September 2003.

[75] Yang, Y., Yi, J., Woo, Y. Y., and Kim, B., "Experimental investigation on efficiency and linearity of microwave Doherty amplifier," in *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 1367–1370, 2001.

[76] Yang, Y., Yi, J., Woo, Y. Y., and Kim, B., "Optimum design for linearity and efficiency of microwave Doherty amplifier using a new load matching technique," *Microwave J.*, vol. 44, pp. 20–36, December 2001.

[77] Zhang, X., Larson, L. E., and Asbeck, P. M., *Design of Linear RF Outphasing Power Amplifiers*. Norwood, MA: Artech House, 2003.

[78] Zhao, Y., Iwamoto, M., Larson, L. E., and Asbeck, P. M., "Doherty amplifier with DSP control to improve performance in CDMA operation," in *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 687–690, 2003.

VITA

Nuttapong Srirattana was born in Bangkok, Thailand, in 1979. He received the B. Eng. degree with first class honors in Telecommunication Engineering from the King Mongkut's Institute of Technology, Ladkrabang, Bangkok, Thailand, in 1999 and the M.S. degree in Electrical Engineering from the Georgia Institute of Technology, Atlanta, Georgia, in 2002. He is currently working toward the Ph.D. degree at the Georgia Institute of Technology in the area of silicon-based transistor modeling and its applications to RF power amplifiers, and novel architectures for high-efficiency linear RF power amplifiers.

During 2001 to 2002 he was a part-time design engineer at the National Semiconductor Corporation in Atlanta and involved in the development of SiGe HBT power amplifier for GSM handset. Since 2001, he has worked as a research assistant in the Analog Integrated Circuit and Signal Processing Laboratory and the Microwave Application Group at the Georgia Institute of Technology. His research interests include microwave/RF device characterization and modeling, linearization techniques, and high-efficiency linear RF power amplifier design.