MMIC based Doherty Amplifiers for High-Frequency Wireless Communications

by

Stefan Shopov, B.Eng.

A thesis submitted to the
Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of

Master of Applied Science in Electrical Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
Department of Electronics
Carleton University
Ottawa, Ontario
September, 2011

Your file Votre référence
ISBN: 978-0-494-83029-1
Our file Notre référence
ISBN: 978-0-494-83029-1

Abstract

This thesis explores the feasibility of integrating the Doherty power amplifier (PA) technique at frequencies of 10 GHz and above. Doherty circuits in two separate areas are improved by employing two integration techniques. A system-on-chip (SoC) approach is used at 71-76 GHz to improve the efficiency of mm-Wave CMOS Doherty PAs, while a system-on-package (SoP) approach is used at 10 GHz to circumvent the current-handling limitations and high costs of completely integrated high power Doherty PAs.

High-frequency device limitations are studied through a specifically developed model to explore their impact on the Doherty performance. From the analysis it is found that the output conductance of active devices is the main cause efficiency of degradation in the Doherty. The SoC implementation is made possible by the successful design of a 70-80 GHz CMOS branch-line coupler. Measurements of the coupler show best reported performance in terms of insertion loss, magnitude imbalance, and phase imbalance. By employing a similar coupler at 71-76 GHz a Doherty SoC is implemented; the Doherty SoC showed a 5.2% back-off efficiency improvement over a balanced amplifier. The high power SoP Doherty is implemented in the final design section of this thesis by integrating a GaN MMIC on an RT/duroid package. Compared to a balanced amplifier, the high power SoP implementation results in twice the back-off power added efficiency.

Acknowledgments

I would like to take this opportunity to express my profound gratitude to my thesis supervisors - Dr. Calvin Plett and Dr. John W. M. Rogers. They have introduced me to the field of RF electronics, continuously encouraged my creative ideas and work, and supported my constant pursuit of knowledge. Dr. Rony E. Amaya was no less important in the course of the two years as he has welcomed, supervised, and guided me at the Communications Research Center Canada (CRC). I am indebted to all three for the researcher I have become.

It would also have been impossible to continue my successful studies at Carleton University without the design and test facilities, financial support, and prototype fabrications provided by CRC, NSERC, and CMC, respectively.

Many thanks to Nagui Mikhail for always being there to help and for maintaining the high-end servers vital to my research, Adrian Momciu from CRC for his expertise and assistance with the testing equipment, and Dr. Leonard MacEachern for providing me with a productive work space.

The experience with the M.A.Sc. program would not have been the same without my fellow graduate students and friends from the Department of Electronics - Zahirul Alam, Kimia Ansari, Greg Brzezina, Peter Chyurlia, Ziad El-Khatib, Vahidin Jupic, Kiril Kidisyk, Yasser Soliman, Boris Spokoinyi, Octavian Stelescu, Zhan Xu, and Wenyao Zhai.

My parents, Tatiana Radkova-Shopova and Plamen Shopov, have played a huge role in my achievements and I am forever grateful for their unconditional love.

Finally I would like to thank my girlfriend and best friend Anastasia Smolentseva for being patient, understanding, and supportive during the course of this journey.

Contents

List of Tables

List of Figures

List of Abbreviations

AC	Alternating Current
$ACPR$	Adjacent Channel Power Ratio
$CMOS$	Complementary Metal Oxide Semiconductor
CPW	Coplanar Waveguide
$CPWG$	Coplanar Waveguide with Ground
CS	Common-Souce
DC	Direct Current
DPA	Doherty Power Amplifier
DSP	Digital Signal Processing
DUT	Device Under Test
EM	Electromagnetic
ERR	Envelope Elimination and Restoration
ET	Envelope Tracking
EVM	Error Vector Magnitude
FET	Field Effect Transistor
f_{MAX}	Maximum Stable Frequency
f_T	Transit Frequency
$GaAs$	Gallium Arsenide
GaN	Gallium Nitride
$HEMT$	High Electron Mobility Transistor
IC	Integrated Circuit
IM	Intermodulation
IP_3	Third-Order Intercept Point
ISM	Industrial, Scientific and Medical
$LINC$	LInear amplification using Nonlinear Components
$MESFET$	Metal Semiconductor Field Effect Transistor
MIM	Metal-Insulator-Metal

$MMIC$	Monolithic Microwave Integrated Circuit
mm-wave	Millimeter-Wave
MOM	Metal-Oxide-Metal
MSG	Maximum Stable Gain
NOF	Number of Fingers
$OFDM$	Orthogonal Frequency-Division Multiplexing
P_{1dB}	1-dB Compression Point
PA	Power Amplifier
PAE	Power Added Efficiency
$PAPR$	Peak-to-Average-Power Ratio
PCB	Printed Circuit Board
Q	Quality Factor
RF	Radio Frequency
SoC	System-on-Chip
$SOLT$	Short-Open-Load-Thru
SoP	System-on-Package
SRF	Self-Resonant Frequency
TE	Transverse Electric Mode
TEM	Transverse Electromagnetic Mode
TM	Transverse Magnetic Mode
TRL	Thru-Reflect-Load
UGW	Unit Gate Width
PNA	Performance Network Analyzer

Chapter 1

Introduction

Ever since the first wireless communication between Alexander Graham Bell and his associate Charles Tainter in 1880, there has been a continued effort to transfer more and more information "wirelessly". This comes with the challenge of implementing products that are simultaneously convenient for the user and are compatible with our existing wireless ecosystem. The consumer expectations put stringent requirements on power consumption and unit cost, while government regulations try to maximize spectrum reuse.

1.1 Motivation

Currently, the majority of consumer electronics are designed to operate in the first 10 GHz of the frequency spectrum making the frequency bands overcrowded and generally expensive to license. The countless wireless devices interfere with one another and limit the attainable data-rates. One method to satisfy the high data-rate requirements of modern *High Definition* (HD) applications is to move the communication to higher frequencies where more bandwidth is available and the spectrum is less crowded. However, to make proper use of the available bandwidth, advanced modulation schemes and wide-band transmission techniques such as OFDM [3] are necessary. Due to the large peak-to-average power ratio (PAPR) associated with these techniques, some form of dynamic power management is required.

With the power amplifier (PA) often being the most power hungry RF component, there is great interest in adapting some of the *classic* PA efficiency enhancement techniques [4] to frequencies above 10 GHz. One of the most promising techniques for these frequencies is the Doherty amplifier, originally proposed in 1936 [5]. The

Doherty topology can be scaled in frequency because it is purely analog and does not require any switching components or digital controls. Also, implementing the Doherty at high-frequencies, where the guided wavelengths are substantially shorter, allows to pursue cheap integrated solutions which are typically too bulky for lower frequencies. Therefore, successfully integrating high-frequency Doherty amplifiers will ultimately make the frequency leap feasible.

1.2 Previous Work

The original work proposed by Doherty in [5] was implemented with vacuum tubes and was intended for single sideband radio transmission. Since then the technique was almost forgotten until its reappearance with Raab's work in [6]. Now the recent trend is the monolithic integration of Doherty amplifiers. Table 1.1 below summarizes the performance of previously reported Doherty amplifiers that are based on monolithic microwave IC (MMIC) technologies.

Table 1.1: Previous work on MMIC based Doherty PAs.

Ref.	Frequency (GHz)	Technology	P_{1dB}/P_{SAT} (dBm)	Backoff (dB)	PAE (%) @ Backoff/P_{1dB}/P_{SAT}
[7]	1.7	0.13-μm CMOS	–/31.5	12	6/–/36
[8]	2.4	0.18-μm CMOS	21.5/–	7	10/14/–
[9]	2.4	90-nm CMOS	17.5/20.5	5	–/–/26.7
[10]	2.4	0.13-μm CMOS	22.7/23	5	35/60/64
[11]	3.65	90-nm CMOS	25.4/26.5	6	12.2/32/39
[12]	4-5	90-nm CMOS	–/25	6	8/–/22
[13]	17	GaAs HEMT	–/25	6	30/–/40
[14]	20	GaAs HEMT	–/23	6	25/–/32
[15]	38-46	GaAs HEMT	–/21.8	6	18/–/20
[16]	60	0.13-μm CMOS	7/7.8	6	1/–/3

The Doherty amplifiers reported in [7–12] operate at frequencies below 5 GHz and their output power ranges between 0.1-1.5 W. In order to save area they are all implemented using lumped elements. In [13] Campbell reported the first fully integrated pHEMT Doherty amplifier which is implemented at 17 GHz with a microstrip based

output combiner. In [14] McCarroll reported a 20 GHz pHEMT Doherty MMIC for satellite digital communications. The Doherty is implemented with the help of previously extracted wafer data. The first millimeter-wave (mm-wave) Doherty MMIC is reported by Tsai in [15]. The main stage of the Doherty is biased in class-AB instead of class-B in order to improve gain and linearity. In [16] Wicks reported the first fully integrated mm-wave CMOS Doherty. The input splitter and output combiner are implemented with microstrip transmission lines. However, the maximum achieved efficiency is only 3.0%.

Given the current state of the art Doherty MMICs, improvements can be made in the two following areas.

- Output power for the frequency range below 20 GHz; gallium-nitride (GaN) technology can be used since it is well established at those frequencies.
- Efficiency for the frequency range around 60 GHz; CMOS PAs have been reported with power-added-efficiencies above 10% [17, 18].

1.3 Thesis Objective

The research work in this thesis will be focused on improving the feasibility of MMIC based Doherty amplifiers at frequencies of 10 GHz and above. First, the difficulties of this task will be studied by analyzing the high-frequency device limitations of CMOS and GaN transistors on the Doherty system with the help of a custom model. Then, two Doherty amplifiers will be implemented by employing two separate integration techniques. A system-on-chip (SoC) approach will be used for a mm-wave CMOS implementation and a system-on-package (SoP) approach will be used for a high-power 10 GHz GaN implementation. To further improve the feasibility of SoC implementations, an in-depth experimental study of a CMOS branch-line coupler will also be done.

1.4 Thesis Outline

Some background information on PAs and their figures of merit will be given in Chapter 2. This chapter also covers the operation of conventional PA classes and the basics of PA efficiency enhancement techniques.

Chapter 3 will go into the analysis of the Doherty architecture. The design equations of the Doherty will be given and generalized beyond the classic Doherty implementation. Then the chapter will focus on the difficulties of implementing high-frequency Doherty systems. The high-frequency device limitations of CMOS and GaN processes will be discussed and incorporated into a custom model of the Doherty. The Doherty model will then be used to analyze the performance and to make suggestions on how to adapt the Doherty circuit.

Chapter 4 will discuss the design and characterization of passive and active components used in the Doherty implementations. Components such as coplanar waveguide (CPW) transmission lines, capacitors, inductors, pads, supply decoupling networks, and active devices will be explained. The chapter will end by going over the de-embedding techniques used to compare the measurement and simulation results.

Chapter 5 will begin with the experimental analysis of an important building block to the Doherty amplifier - the branch-line coupler. The design and measurements of a 70-80 GHz 130-nm CMOS branch-line coupler will be discussed. Then the chapter will focus on two separate Doherty implementations. The first system uses an SoC approach to integrate a 71-76 GHz Doherty PA in a 90-nm CMOS process, and the second system uses an SoP approach to integrate a high power 10 GHz GaN Doherty PA with an RT/duroid board.

Finally, Chapter 6 will summarize all contributions and will give direction for future work.

Chapter 2

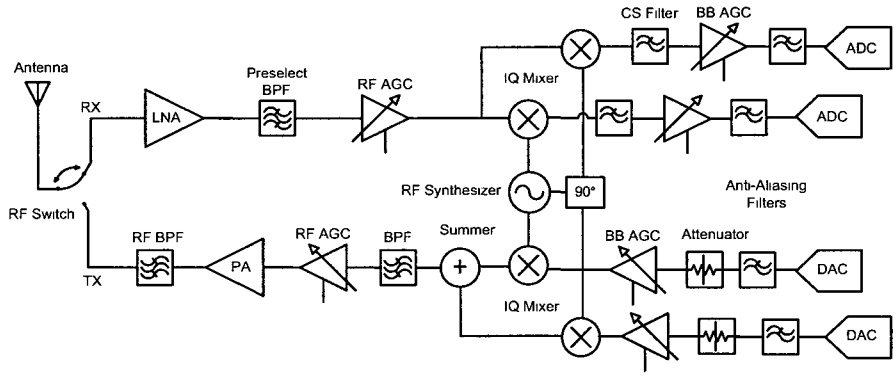
Power Amplifiers and Efficiency Enhancement

Wireless communication links can be made up of a few radio devices or thousands of them, while the information can be broadcasted by a single device or there could be a complex mesh of interactions. Regardless of the structure, every time there is a wireless exchange of information, there is a device that acts as a transmitter and a device that acts as a receiver. If the communicating device can both receive and transmit, then it is termed as a *transceiver*. There are two classic transceiver architectures which are shown in Figure 2.1 (a) and (b) [19]. Both of these architectures include a power amplifier (PA) component.

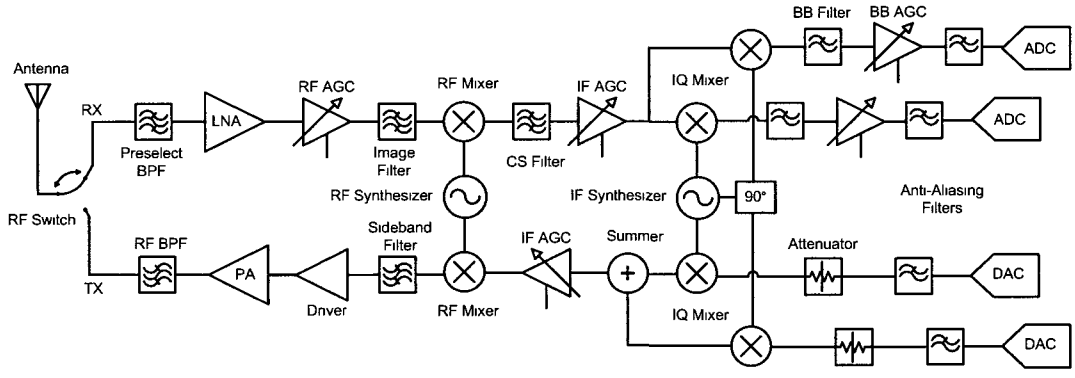
The PA is the topic of the work and it will be introduced in this chapter, followed by a discussion of PAs figures of merit and the operation of conventional PA classes. The second part of this chapter will concentrate on complex efficiency enhancement PA systems.

2.1 Power Amplifiers (PAs)

The power amplifier (PA) of the radio is located on the transmitter side of the system and its role is to efficiently energize signals so that they have enough power to be successfully transmitted. The applications in which wireless PAs are needed range from high power RF base-stations to efficient millimeter-wave portable electronics. The common ground between the various applications is that PAs need to be simultaneously efficient and linear. Efficiency is needed to maintain good battery life in portable applications and to ensure low thermal heating for high power applications



(a) Direct-down conversion transceiver.



(b) Superheterodyne transceiver.

Figure 2.1: Transceiver architectures.

while linearity is needed to prevent interference from nearby users and to conserve a certain signal quality. The main predicament when designing a PA is the inherent trade-off between the two. This trade-off occurs due to the physical limitations of transistors and the circuit topology.

2.1.1 Efficiency

The efficiency (η) of a PA measures how well power is converted from DC to RF and can be calculated by taking the ratio of the output RF power to the consumed DC power with equation (2.1) [19].

$$\eta = \frac{\mathrm{P}_{rf}}{P_{DC}} \tag{2.1}$$

For a FET device this type of efficiency is often termed *drain efficiency* and is independent of gain The drain efficiency is not the best measure of PA usefulness because the amplifier could have a high drain efficiency and almost no signal gain. Therefore to quantify the effectiveness of a PA one must find the power-added-efficiency (PAE) through (2.2).

$$PAE = \frac{\mathrm{P}_{rf} - P_{in}}{P_{DC}} = \eta\left(1 - \frac{1}{G}\right) \tag{2.2}$$

PAs are most efficient when they produce maximum drain voltage and current swing. But as the signal swing increases non-linear effects begin to appear.

2.1.2 Linearity

The non-linearities of PAs are generated from the intrinsic properties of active devices, the circuit topology, and the limited supplies in the circuit. To measure the linearity of a circuit either the 1-dB compression point (P_{1dB}) or the third-order intercept point (IP_3) can be used [19]. The P_{1dB} is a single tone measure of the output power at which the fundamental signal compresses by one decibel. In single carrier transmitter systems the P_{1dB} specification is the main measure of linearity. However, modern wide-band systems are increasingly employing multi-tone data transmission techniques such as orthogonal frequency-division multiplexing (OFDM) to avoid the difficulties with equalization, impulsive noise, and interferers [3]. In these multi-tone

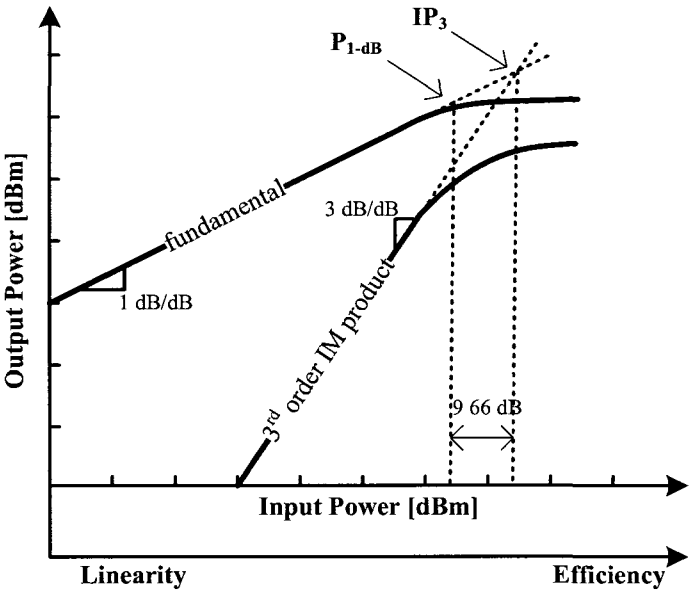


Figure 2.2: Output power of fundamental and IM3 product versus input power.

systems, the main cause of signal degradation is no longer caused by signal compression but instead it is caused by intermodulation (IM) distortion between the carriers. Therefore, for a multi-carrier system the IP_3 point becomes much more relevant since it is a direct measure of the third-order IM products. The IP_3 can be measured for a given non-linear system by applying two tones of equal power at an offset, and finding the fictional point at which the extrapolated third-order IM product intercepts the power of the extrapolated fundamental. The P_{1dB} and IP_3 are shown in Figure 2.2 and although they quantify two separate distortion effects, they are both derived from a power series expansion and thus can be related with (2.3) [19].

$$P_{1dB} = IP_3 - 9.66dB \tag{2.3}$$

From a transceiver system point of view, the signal produced by the PA must meet the error vector magnitude (EVM) and the adjacent channel power ratio (ACPR) requirements for the intended application. The EVM measures the in-band distortion by quantifying the amount by which a signal deviates from its ideal and the ACPR measures the ratio of the total RF power that is dumped in adjacent channels. More details on both can be found in [19,20], but from a PA perspective the important part

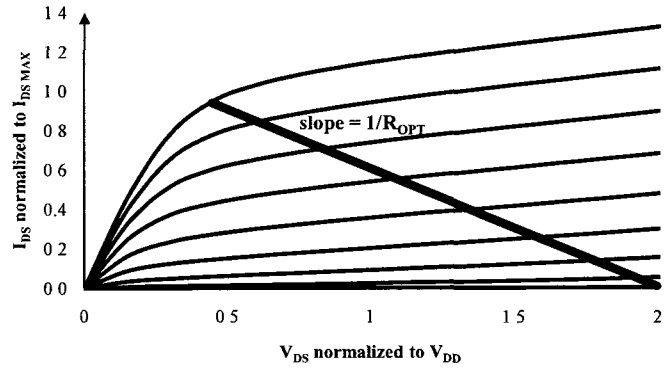


Figure 2.3: Drain current and voltage swing of PA.

is that the EVM and ACPR are related to the P_{1dB}/IP_3. The exact relation can either be quantified using analytical equations for simple modulation schemes, or through empirical equations and detailed system level simulations for complex modulation schemes. Regardless of the modulation scheme and the exact relationship between EVM, ACPR, P_{1dB}, and IP_3, the linearity of a PA can be improved with one of the three methods: 1) by operating the PA away from its maximum output swing (at a power back-off); 2) by using a more linear class of amplifier; or 3) by employing external linearization circuits [4]. All three methods affect the efficiency at which power is produced so the linearity and efficiency trade-off can never be avoided.

2.1.3 Output Power and Matching Considerations

Until now the power levels produced by the PA were left out to demonstrate the relationship between linearity and efficiency. However, in practical designs one would most likely be designing in accordance with some specifications so the output power is indeed needed. Assuming that a technology is pre-selected to handle the PA specifications and the maximum supply voltage is fixed by the physical limits of that technology, the output power can be found in terms of the load (R_L) with (2.4). The current and voltage excursions on the I-V curves are both limited by the load. In order to maximize the output power, R_L is selected such that it maximizes the drain current and voltage swing. The load producing maximum power for a given device is termed as the optimum load (R_{OPT}) and is shown in Figure 2.3. The R_{OPT} value is device dependent, so for a small device R_{OPT} is large and for a large device R_{OPT} is

small For all sizes, the device R_{OPT} maximizes swing and produces a roughly constant ratio of RF power to DC power Therefore, the linearity and efficiency trade-off is theoretically independent of output power Although it is more difficult to design higher power PAs because of the thermal issues and the extreme impedances to which they need to be matched After selecting the R_{OPT} value with equation (2 4) to which the PA delivers the required power, the PA transistor is scaled such that it can supply the required current swing

$$P_{out} = \frac{(2 \ V_{DD} - V_{knee}) \ I_{DS\max}}{8} \cong \frac{V_{DD}^2}{2 \ R_L} \tag{2 4}$$

The output matching network is then designed to present R_{OPT} at the output of the PA transistor while simultaneously resonating any of the reactive parasitic components This type of matching is termed *optimum power matching* and is different from the *small-signal conjugate matching* used in gain amplifiers since it provides maximum large-signal power delivery to the load instead of maximum gain The output large-signal matching network is one reason why PAs have lower power gains than small-signal amplifiers and often require additional driving stages The input network is typically designed by conjugately matching the source to the reflection coefficient (S_{11}) of the PA transistor Figure 2 4 compares the matching strategy and the resulting output power of a gain amplifier and a power amplifier A few output matching networks that can be employed in PA design are suggested in Figure 2 5 These networks should be valid across technologies and frequencies since in most cases PAs are driving low impedance loads that contain an inductive reactance to resonate the capacitive parasitics of the device For low frequencies the lumped element matching method is recommended since it is more compact As the frequency of operation increases and all elements become distributed, a hybrid method which combines lumped and distributed elements may provide higher quality matching networks It is important to note that the bulky RF choke and $\lambda/4$ bias feeds are often absorbed in the matching networks in the form of smaller elements for layout saving reasons

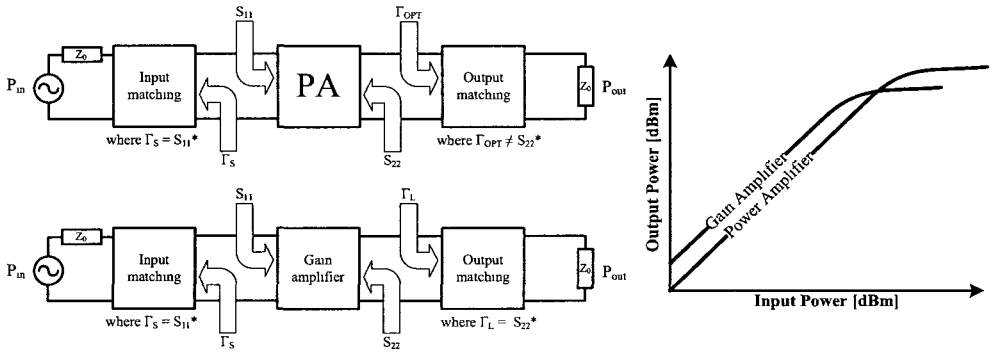
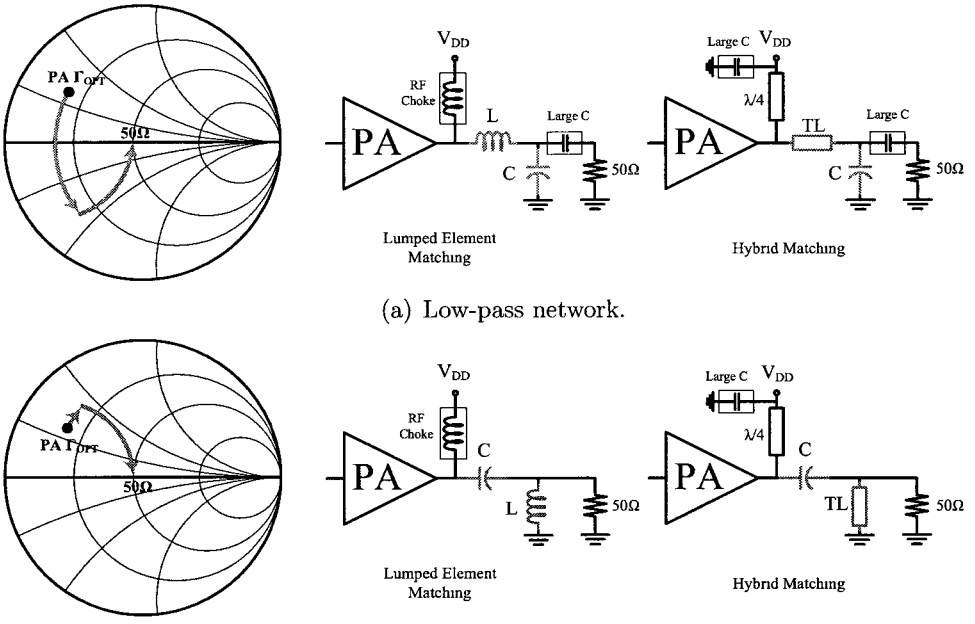


Figure 2.4: PA matching compared to small-signal gain matching.



(a) Low-pass network.

(b) High-pass network.

Figure 2.5: Comparison between lumped element matching and hybrid matching.

2.2 RF PA Classes

A single type of PA cannot be simultaneously optimized for output power, efficiency, linearity, and gain. Therefore, to accommodate each individual application there are different classes of PAs. The discussion of the various PA classes begins with the simplest and most intuitive family of PAs - the class-A, -AB, -B, and -C.

2.2.1 Class-A, -AB, -B, and -C

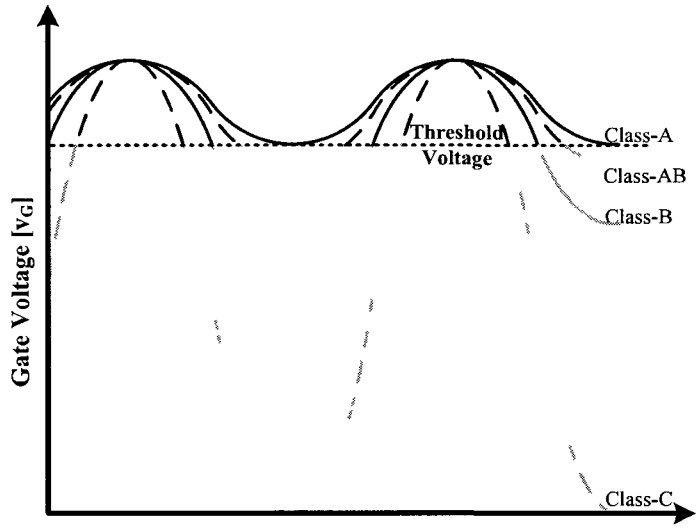
The main difference between these classes of PAs is the conduction angle (α) at which the PA transistors are biased. The conduction angle is defined as shown in Figure 2.6 and represents the portion of the full RF cycle under which PA transistor is conducting current.

To understand the linearity and efficiency trade-offs between these classes, a Fourier analysis of the drain current waveforms from Figure 2.6 (b) is done as described in [4]. Until now the only assumption made regarding the drain current is that the PA behaves as an ideal transconductor. The transconductor alone does not generate any non-linearities, but because of the fixed threshold voltage of the transistor the output current is clipped. With the Fourier analysis, the amount of current converted to the harmonic frequencies with the clipping mechanism can be calculated. Thus, the resulting DC component of the drain current is represented by equation (2.5), and the resulting fundamental signal (n = 1) and harmonics (n > 1) are represented with equation (2.6). By solving for each component, the DC and harmonics of the drain current are plotted as a function of the conduction angle in Figure 2.7.

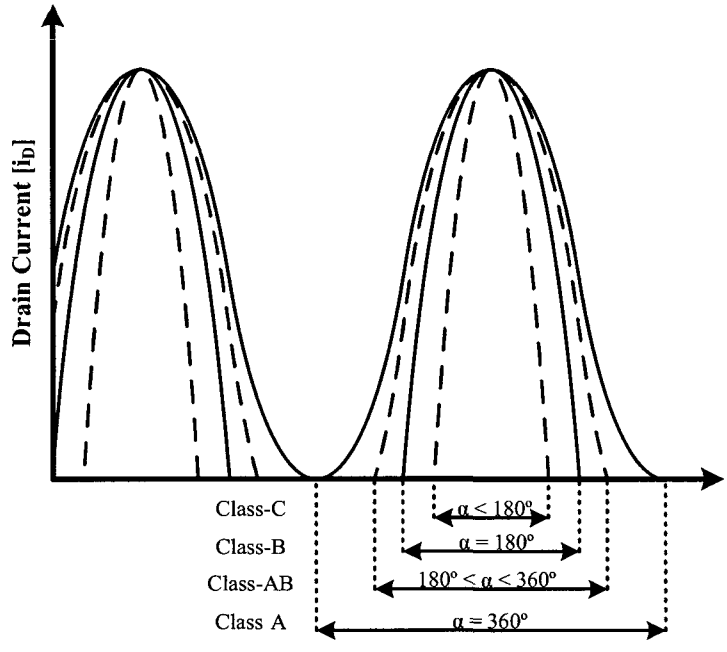
$$I_{DC} = \frac{1}{2\pi} \cdot \int_{-\alpha/2}^{\alpha/2} \frac{I_{\max}}{1 - \cos(\alpha/2)} \cdot (\cos\theta - \cos(\alpha/2)) \cdot d\theta \tag{2.5}$$

$$I_n = \frac{1}{\pi} \cdot \int_{-\alpha/2}^{\alpha/2} \frac{I_{\max}}{1 - \cos(\alpha/2)} \cdot (\cos\theta - \cos(\alpha/2)) \cdot \cos(n\theta) \cdot d\theta \tag{2.6}$$

In order to translate the drain current waveforms into efficiency curves and output power, an additional assumption needs to be made: this is that all harmonics are



(a) Input signal applied at the gate



(b) Resulting drain current

Figure 2.6: Waveforms for reduced conduction angle PAs

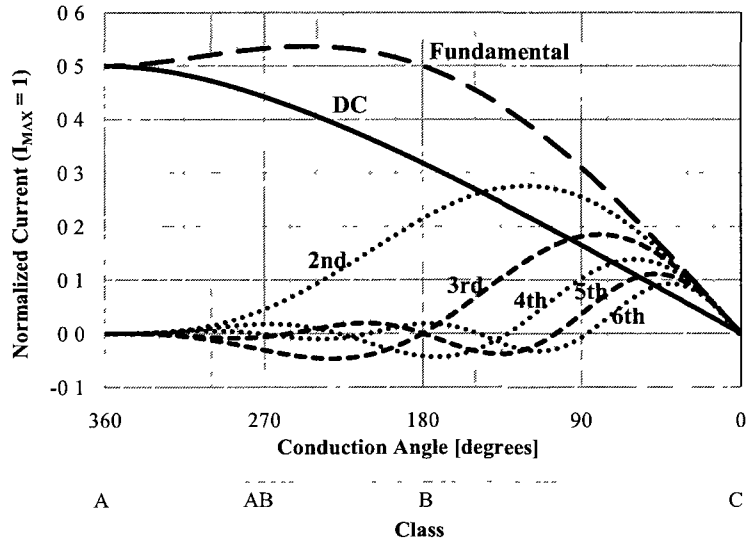


Figure 2.7: Fourier analysis of ideal transconductor.

shorted through the output matching network and generate no voltage at the load [4]. Although this assumption results in an idealistic case, important conclusions can be drawn about the operation of reduced conduction angle amplifiers without the heavy non-linear analysis. The resulting output power is found with (2.7) by ignoring the harmonics and by assuming that the PA is matched to achieve full voltage swing. The DC power consumption is found with its standard definition as shown in (2.8). The efficiency and output power (normalized to class-A) are plotted in Figure 2.8 using equations (2.1) and (2.7).

$$P_{fundamental} = \frac{V_{DC} \cdot I_{fundamental}}{2} \tag{2.7}$$

$$P_{DC} = V_{DC} \cdot I_{DC} \tag{2.8}$$

By observing Figures 2.7 and 2.8, it can be noticed that the output power of the fundamental is roughly constant for conduction angles from classes-A through -B while the drain efficiency increases from 50% to 78.5%. Also, it can be noticed that class-AB amplifiers provide a very good efficiency to linearity trade-off because they are practically as linear as a class-A but are around 10% more efficient. Finally,

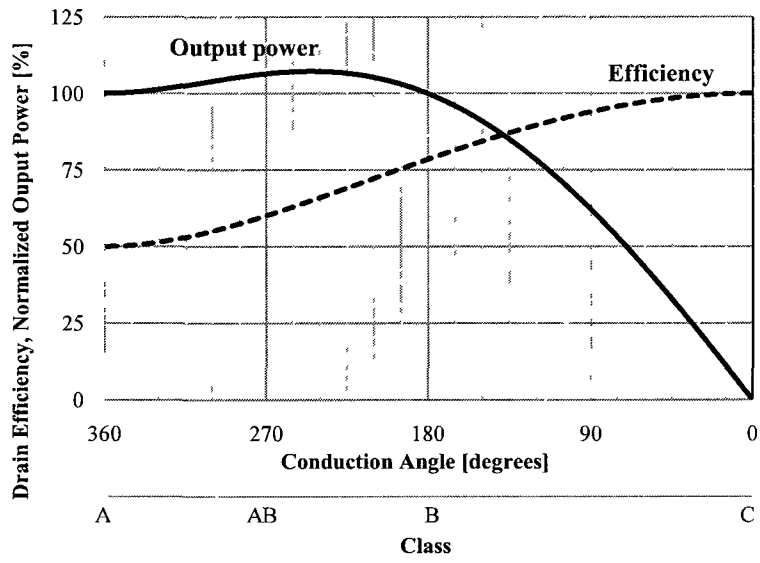
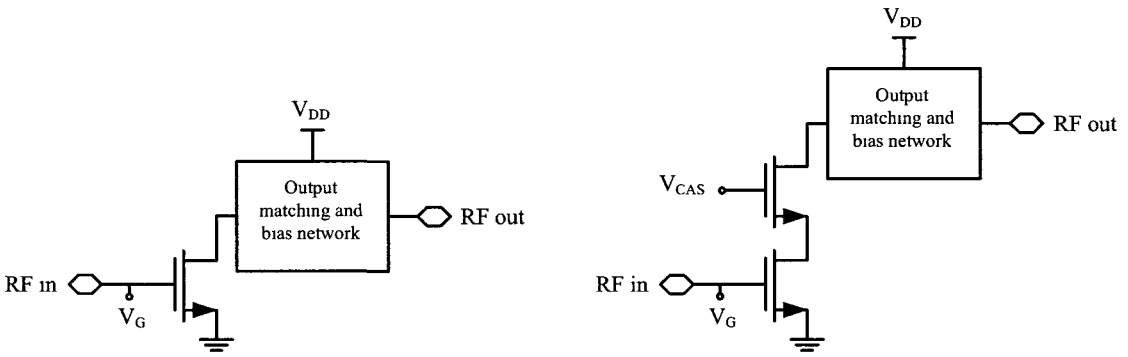


Figure 2.8: Output power and efficiency from Fourier analysis.

class-C amplifiers are not very well suited for non-constant envelope signals since the output power is a strong function of the conduction angle. Even though gain is not directly part of this analysis, it can be observed from Figure 2.8 that low conduction angle PAs produce low power while requiring large input signal swings which implies that they have very low gains. Therefore, the usefulness of class-C amplifiers becomes questionable as soon as they stop providing gain.

From a design perspective, PA classes-A through -C are topologically identical and at RF frequencies they are typically designed with either a common-source (CS) or a cascode circuit as shown in Figure 2.9. The advantage of using a single transistor CS amplifier is the lower dependence on power-supply and process variations [21]. On the other hand, cascode structures offer higher output impedances, better linearity, and are less prone to breaking down, but their drawbacks include a pole at approximately $f_T/2$ formed by the parasitics of the cascode device that reduces the frequency of operation.



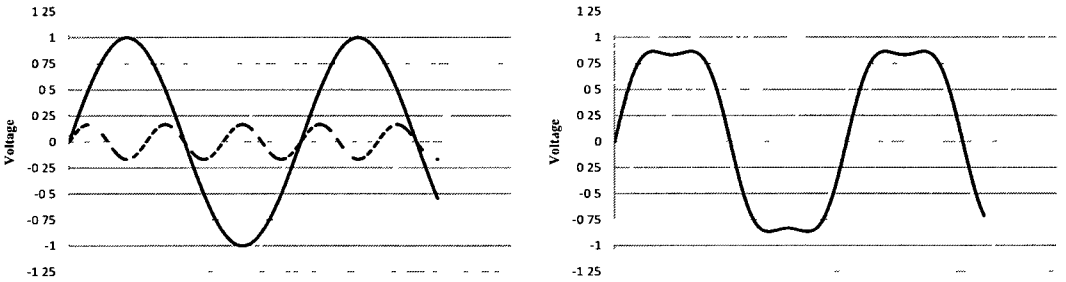
(a) Common-source. (b) Cascode.

Figure 2.9: Class-A, -AB, -B, and -C circuit topologies.

2.2.2 Class-F

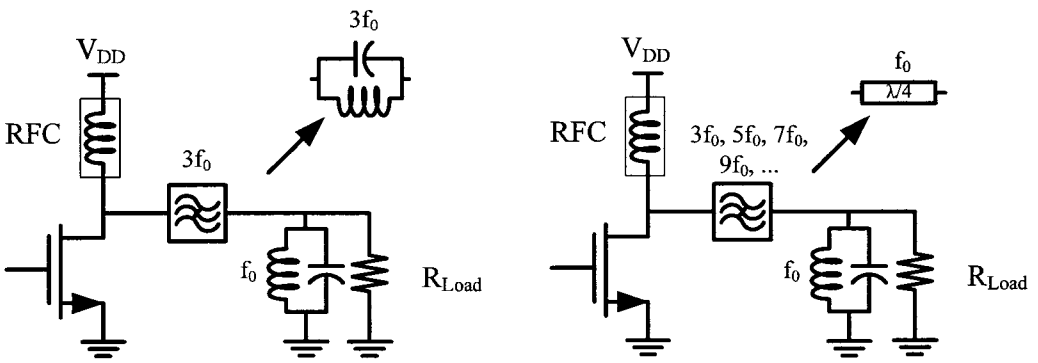
Class-F PAs are similar to class-B but with specifically introduced harmonics to increase the output power beyond class-B levels. With the introduction of harmonics, the total voltage waveform at the drain of the PA can be made to appear smaller than the voltage waveform of the fundamental alone. This occurs because the harmonic signals reduce the peaks of the fundamental signal, thus the circuit supports a larger fundamental signal. The *classic* class-F amplifier only uses the third order harmonic to shape its voltage waveform and achieves a theoretical efficiency of about 90%. It is shown in [4] that for the *classical* class-F implementation, best efficiencies are achieved when the third-order harmonic has an amplitude of 1/6 the amplitude of the fundamental. The waveforms for this case are shown in Figure 2.10. To achieve even better efficiencies (theoretically close to 100%), all higher order odd harmonic terms (i.e. 3rd, 5th, 7th, etc.) can be introduced to make the drain voltage waveform appear as an ideal square [19].

The implementation of a class-F PA requires a band-stop filter that prevents the harmonics at the drain node from being dampened by the load. For the *classic* class-F circuit where only the third harmonic is needed, the band-stop can be achieved by placing a single tank circuit between drain and the load as shown in Figure 2.11 (a), and for the case that uses all odd-order harmonic terms, this is done with a quarter-wave transformer and a shunt tank circuit as shown in Figure 2.11 (b). The advantage of using a class-F PA is that it produces about 0.5 dB more power with



(a) $Sin(\theta)$, $\frac{1}{6}Sin(3\theta)$ (b) $Sin(\theta) + \frac{1}{6}Sin(3\theta)$

Figure 2.10: Voltage waveform of fundamental (a) without third-order harmonic and (b) with third-order harmonic.



(a) Classic class-F circuit. (b) Transmission line in class-F circuit.

Figure 2.11: Class-F circuit topologies.

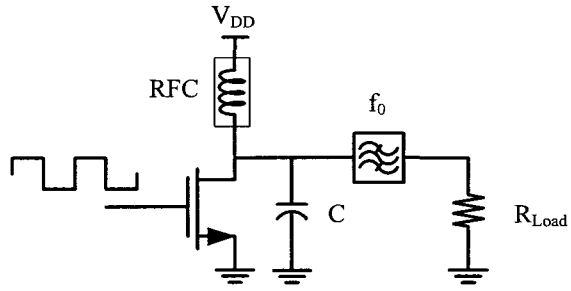


Figure 2.12: Class-E circuit topology.

the same DC power consumption compared to a class-B [4], however this requires an additional output resonating circuit which contributes some additional losses.

2.2.3 Switched Mode: Class-E

Class-E amplifiers which were first introduced by Sokal in [22], consist of a transistor operating as an on-off switch and a load network that provides proper transient response. The purpose of the load network is to shape the current and voltage waveforms such that they have minimum overlap thus achieving close to 100% efficiency. A class-E amplifier circuit is shown in Figure 2.12; it works by storing energy in the RFC during the *on-cycle* of the switch and then delivering this energy to the load during the *off-cycle* of the switch. To ensure lossless operation, the shunt capacitor and filtering network are designed such that the internal node of the switch meets the conditions in (2.9) at the time of switching. An additional output filtering network is often necessary to the clean harmonic content because without it the loaded quality factor (Q) of the output does not provide enough harmonic suppression [23].

$$\begin{aligned} v_C(\theta) &= 0 \\ \frac{dv_C(\theta)}{d\theta} &= 0 \end{aligned} \tag{2.9}$$

This PA class is made feasible in modern technologies by absorbing the transistor's output parasitic capacitance into the shunt capacitor of the circuit. The main drawback of class-E amplifiers is that they do not have a flat gain response [4]. Therefore, they are not suited to amplify amplitude modulated signals or signals that contain a high peak-to-average-power-ratio (PAPR). Also, class-E amplifiers are limited to

a lower output power than class-AB amplifiers because they exhibit a larger peak voltage swing, so to prevent device breakdown their supplies are often reduced [19].

2.2.4 Balanced Amplifiers

Two PAs of the same class can be combined in parallel through a pair of 3-dB quadrature couplers to form a balanced PA. The resulting balanced architecture is shown in Figure 2.13 and not only that it outperforms its single ended counter-part, it also makes the circuit much more manufacturable. Balanced amplifiers have key advantages when it comes to: 1) matching, because mismatches appear in anti-phase and cancel, 2) stability, since they are unconditionally stable in-band and often have improved stability out of band, and 3) reliability, because there is a level of redundancy [4]. From a PA perspective, balanced PAs also offer twice the output power compared to single-ended PAs. The drawbacks of using balanced amplifiers include: 1) larger area due to the additional amplifier and the pair of couplers (although some of this area can be spared by employing simpler matching and stabilizing networks), 2) lower gains since the couplers have an insertion loss, and 3) a limited bandwidth because of the couplers.

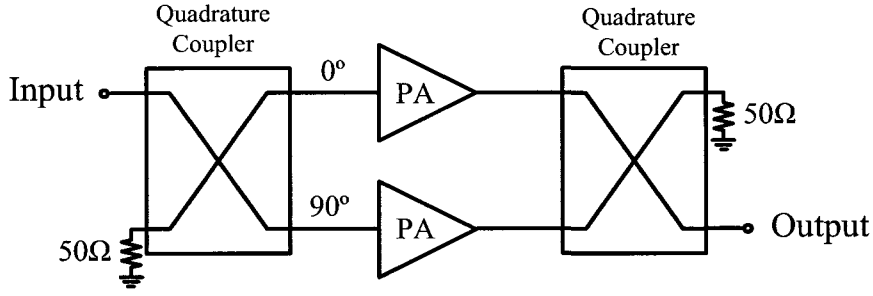


Figure 2.13: Balanced PA architecture.

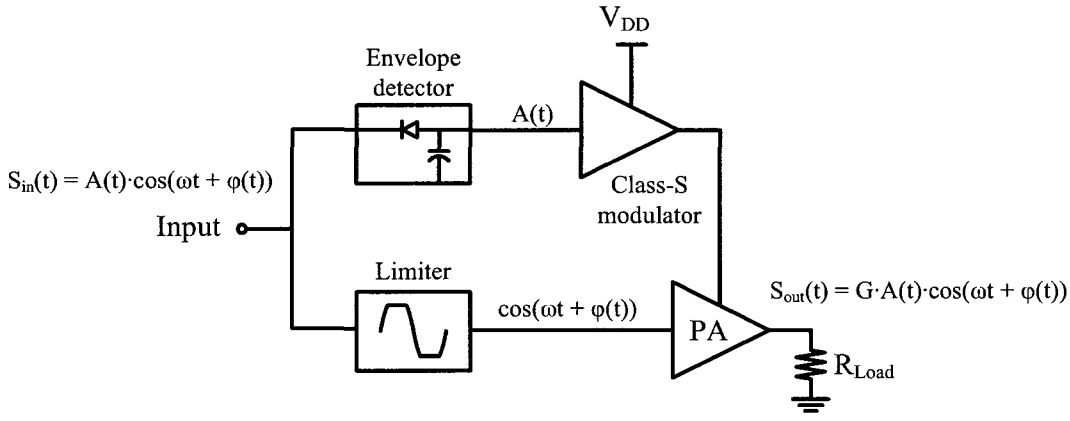
The quadrature coupler can be implemented with either an interdigitated Lange coupler, a branch-line coupler, or with a Wilkinson coupler combined with a 90° phase shift [4]. The coupler type is selected based on the bandwidth and area restrictions set by the specification, as well as on the selection of layers available in the technology used for implementation.

2.3 PA Efficiency Enhancement Techniques

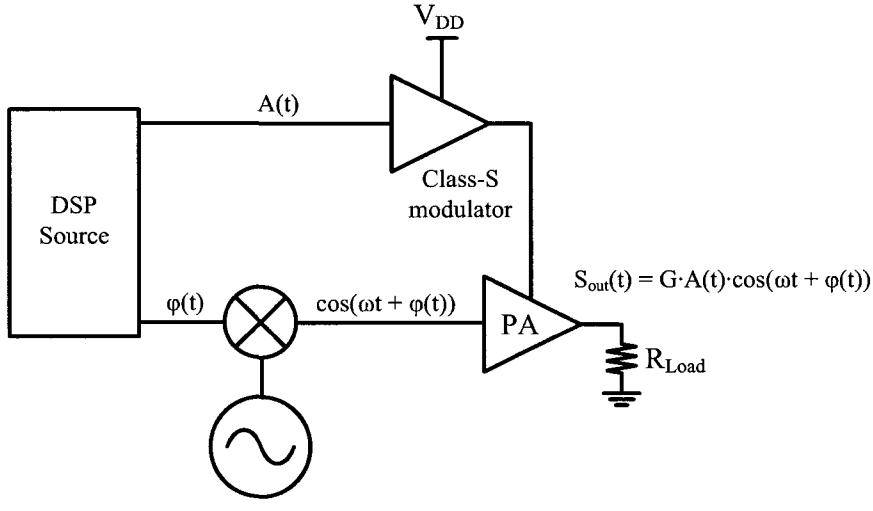
The trade-off between linearity and efficiency has further implications when amplifying modulated signals. Modulated signals have a non-constant envelope because they contain information in both their phase and their magnitude. This results in a constantly changing PA efficiency and linearity based on the properties of the signal at each instant. The conventional PAs discussed in Section 2.2 can be optimized to meet certain linearity and efficiency specifications at a single power level. Complex PA systems are used to extend the peak linearity or peak efficiency over a broader range of power levels. The techniques used to extend the maximum linearity region are termed *linearization enhancement* and the techniques used to extend the maximum efficiency region are termed *efficiency enhancement* [4]. The primary focus of this work is extending the PA range under which maximum efficiency is achieved, and consequently the following subsections describe the main efficiency enhancement systems. For the interested reader, further information on linearization techniques can be found in [4, 24, 25].

2.3.1 Envelope Elimination and Restoration (EER)

The classic envelope elimination and restoration (EER) technique, which was first introduced by Kahn in [26], combines a highly efficient and nonlinear RF PA with a low-frequency envelope amplifier. The nonlinear RF PA is used to amplify the phase-modulated portion of the signal with a high efficiency class-C, -E, or -F power amplifier, while the envelope amplifier, typically implemented as class-S, amplifies the envelope signal at ideal efficiencies approaching 100% [27]. After amplification the envelope of the signal is recombined with the phase-modulated carrier by modulating the supply of the RF PA. In the classic implementation shown in Figure 2.14 (a), the envelope and phase-modulated carrier are separated from the original signal with an envelope detector and limiter circuit. In modern implementations, the envelope detector and limiter can be omitted and both the envelope and phase modulation can be generated with a DSP unit as shown in Figure 2.14 (b) [27]. The main difficulty in implementing EER systems is that the high efficiency supply modulator must handle a large dynamic range as well as a switching frequency six times the RF bandwidth [27].



(a) Classic EER system.



(b) Modern Polar PA system.

Figure 2.14: Envelope elimination and restoration (EER) systems.

2.3.2 Envelope Tracking

The envelope tracking (ET) technique is similar to the ERR technique, but instead of separating the envelope and phase-modulation, the ET system amplifies the original non-constant envelope signal and uses its envelope to dynamically adjust the bias of the PA. The main difference in implementation of ET systems is that the RF PA must be relatively linear because it amplifies a non-constant envelope signal. Therefore, the RF PA is typically implemented with either a class-A, -AB, or -B [25].

Ideally, this technique would allow the bias to exactly follow the signal. However, typical ET systems are implemented with a DC-DC converter which has discrete supply levels resulting in stepped efficiency characteristics. Similar to the EER technique, the limitations of ET systems are implementing the supply modulation.

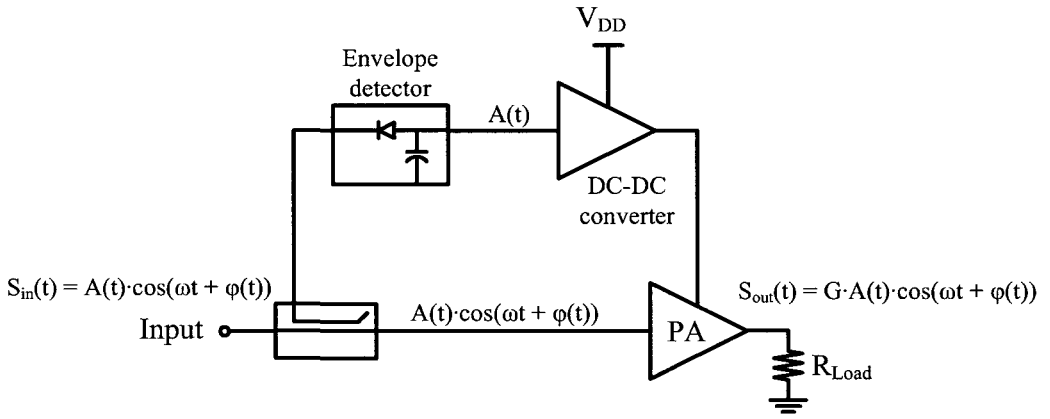


Figure 2.15: Envelope tracking (ET) system.

2.3.3 The Doherty Amplifier

The Doherty technique was first developed in [5] and consists of two parallel PA stages, a main amplifier, and an auxiliary (peak) amplifier. In the *classical* implementation shown in Figure 2.16 the main amplifier is biased in class-B and the auxiliary amplifier is biased in class-C [4]. The main PA is always amplifying; meanwhile, due to its bias point, the auxiliary amplifier only turns on when high output power is needed. The purpose of having the auxiliary amplifier is to modulate the load seen by the main amplifier thus keeping it in saturation. Since PAs are most efficient when they are saturated, the high efficiency region which is typically only seen at peak powers is extended. Ideally, the Doherty technique will produce the famous two peak efficiency response shown in Figure 2.17. Both peaks are located at 78.5% which is the maximum efficiency of a class-B amplifier. Between the peaks there is a slight drop in efficiency because the auxiliary amplifier is conducting but not yet at its optimum efficiency. In the classic circuit, the high power regime occurs when the input signal is at 50% of the full rail-to-rail voltage swing thus resulting in a 6-dB high-efficiency region.

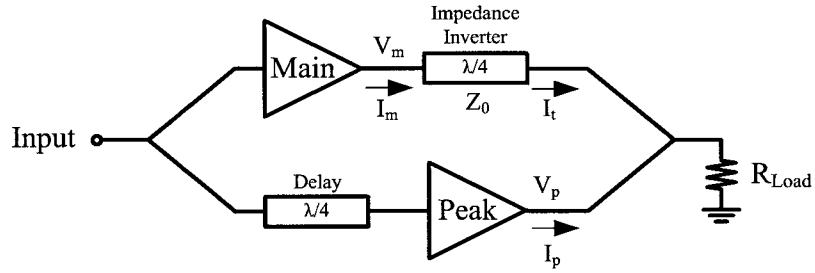


Figure 2.16: Doherty power amplifier system.

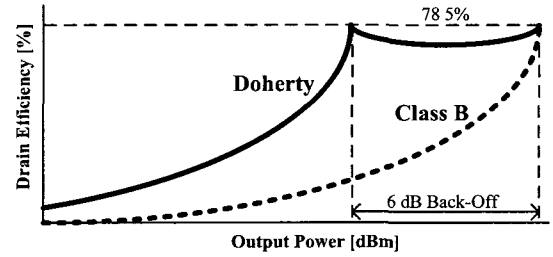


Figure 2.17: Classic Doherty efficiency curves.

The extended efficiency region is achieved through an impedance inversion of the load along with the current and voltage characteristics shown in Figure 2.18. In the low-power regime the main amplifier current (I_m) and voltage (V_m) rise linearly. As the input power increases and V_m begins to saturate, the peaking amplifier starts to conduct producing I_p. The resulting I_p contributes to the current flow into the load which makes the effective load impedance as seen from the quarter-wave transformer (Z_t) appear larger. With the quarter-wave transformer which behaves as an impedance inverter, the effective load seen by the drain of the main amplifier (Z_m) is reduced. Therefore, the load modulation effect allows the current of the main amplifier to continue to increase linearly. To demonstrate the load modulation mathematically, the load impedances as seen by the main amplifier (Z_m), the peak amplifier (Z_p), and the impedance transformer (Z_t) are expressed in terms of the current contributions with equations (2.10)-(2.13) [4].

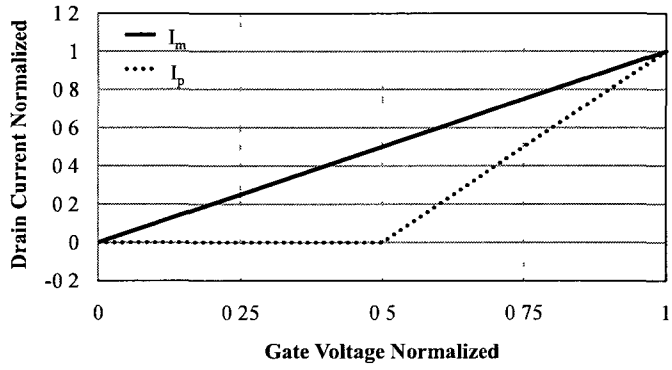
$$Z_p = R_{Load}\left(1 + \frac{I_p}{I_t}\right) \tag{2.10}$$

$$Z_t = R_{Load}\left(1 + \frac{I_t}{I_p}\right) \tag{2.11}$$

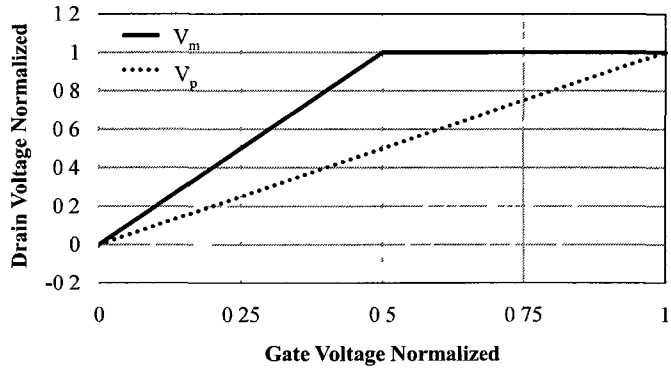
$$Z_m = \frac{Z_0^2}{Z_t} \tag{2.12}$$

$$\Rightarrow Z_m = \frac{Z_0^2}{R_{Load}}\left(\frac{I_t}{I_t + I_p}\right) \tag{2.13}$$

The main difficulty with the Doherty technique is to implement the ideal device characteristics with actual FETs [4].



(a)



(b)

Figure 2.18: Ideal Doherty characteristics.

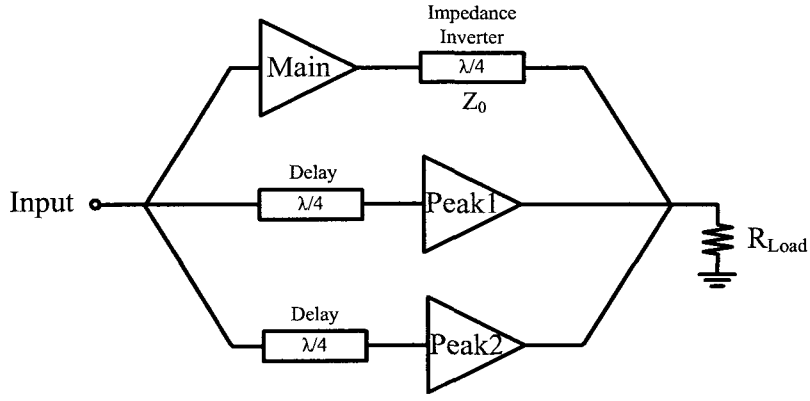


Figure 2.19: Three-way Doherty PA.

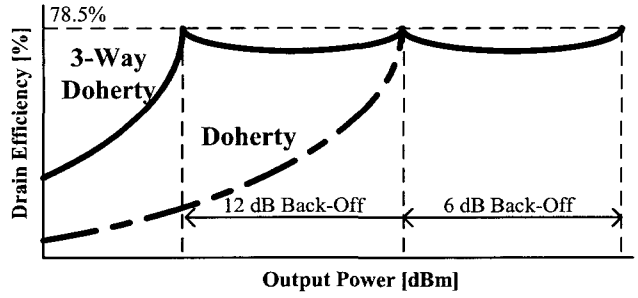


Figure 2.20: Three-way Doherty efficiency characteristics.

The Doherty technique can also be extended to three or more stages [28]; the idea is to increase the number of peak stages such that they can modulate the load seen by the main amplifier for an extended back-off region. A three-way Doherty circuit is shown in Figure 2.19 and its efficiency characteristics are compared to the classic case in Figure 2.20. The analysis for the three-way Doherty can be found in [6]. Theoretically it is very elegant but its practical advantage has yet to be proven.

2.3.4 Outphasing (LINC)

The outphasing technique or often referred to as *LINC*, which is an acronym for *LInear amplification using Nonlinear Components* [4], was first developed by Chireix in [29]. It employs a rather unconventional method to solve the PA efficiency problem. Instead of adapting the system to handle instantaneous signal levels (similarly to the

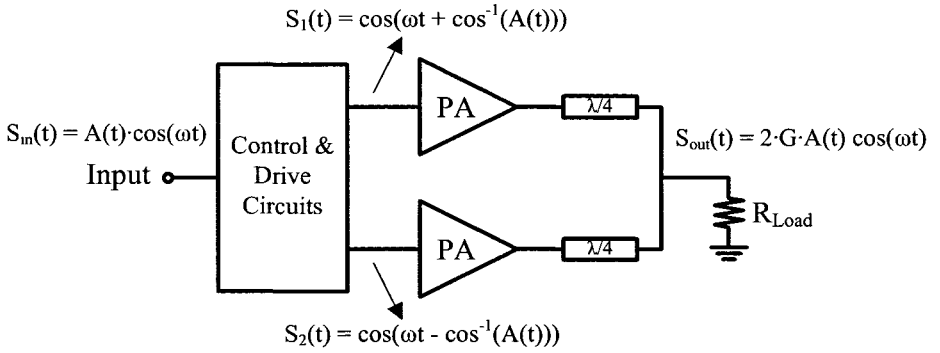


Figure 2.21: LINC system.

techniques described in Sections 2.3.2 - 2.3.3), it mathematically breaks down the amplitude information of a signal into two constant envelope signals that encode the amplitude as a differential phase-shift [4]. This transformation of amplitude modulation into phase modulation can be done through the properties of the trigonometric identity in equation (2.14).

$$\cos(A) + \cos(B) = 2 \cdot \cos\left(\frac{A+B}{2}\right) \cdot \cos\left(\frac{A-B}{2}\right) \tag{2.14}$$

The two separate constant envelope signals are then amplified through a pair of nonlinear PAs. After being amplified they are combined to reconstruct the original modulation. The circuit topology of an outphasing LINC system is shown in Figure 2.21. Similarly to the Doherty architecture, the LINC technique combines two separate signals at the output and this results in a load modulation effect. However, unlike the Doherty the combined signals are not in-phase so the modulated load contains a reactive by-product. This reactive term is related to the phase difference between S_1 and S_2 and must be removed through shunt reactances to preserve the optimum efficiency characteristics. Therefore, one of the main design issues is the proper selection of shunt reactances to shape the efficiency curves. The relationship between the signal characteristics and the shunt reactances has been tabulated in [30]. Another difficulty in implementing the system is generating the input drive circuit [4].

Chapter 3

Doherty Analysis and Design Challenges

The work in all following chapters is focused on the Doherty architecture because out of the efficiency enhancement techniques it is the only one that is purely RF and can be designed to be operationally independent from any baseband control signals. Unlike the rest of the techniques, it does not require any switching components that are typically difficult to scale in frequency. Moreover, the Doherty system is stand-alone and does not require complicated DSP signal conditioning.

In Section 3.1, the Doherty design equations are given and are generalized beyond the classic implementation. Then the focus is shifted towards the difficulties of implementing high-frequency Doherty systems. This is done by first discussing the high-frequency device limitations in Section 3.2, and then by incorporating them into a circuit level model of the Doherty in Section 3.3. Finally, in Section 3.4 the Doherty model is used to quantify the efficiency degradation and to suggest improvements for operation at high-frequencies.

3.1 Design Equations of the Doherty

3.1.1 Classic Doherty

As explained in Chapter 2, the *classic* Doherty is obtained when the high-efficiency region extends from peak power to a 6-dB back-off. For optimum power combining in this classic case, the load and transformer impedances from Figure 2.16 must satisfy the relationships in equations (3.1) and (3.2) [4],

$$Z_0 = R_{OPT} \tag{3.1}$$

$$R_{Load} = \frac{R_{OPT}}{2} \tag{3.2}$$

where R_{OPT} is the optimum load seen by each stage at maximum power.

To achieve the ideal performance described in Section 2.3.3, the peaking PA device must be biased such that it turns on at half of the full voltage swing and its current must rise at twice the rate of the main PA device. However, achieving both of these current characteristics with same periphery devices is not possible since the class-C FET used to hold off the turn-on point of the peak amplifier cannot possibly achieve the same drain current as the main FET. This can be confirmed with the Fourier analysis in Section 2.2.1 where it was shown that reduced conduction angle PAs (beyond class-B) exhibit a lower drain current fundamental component.

Fortunately, there are three solutions to the non-ideal current characteristics of the peak device [4]. The first and simplest solution involves increasing the periphery of the peak transistor to achieve the same output power from the class-C stage as the main class-B stage. The scaling factor varies depending on how deep into class-C operation the peak device is biased. From the previous designs in [11, 12] and from the discussion in [4], the peak transistor should be scaled to about twice the periphery of main transistor. The two other alternatives shown in Figures 3.1 and 3.2 are not as elegant since they require additional adaptive components. The system in Figure 3.1 uses two identical PA stages biased in class-B along with an input attenuator to shape the gain and cut-off characteristics of the peak amplifier, whereas the system in Figure 3.2 uses a class-B for the main PA and a equal periphery device for the peak PA that is adaptively biased. The peak device in Figure 3.2 initially starts off as a class-C to meet the cut-off characteristics, then as the power increases, the adaptive bias turns it into a class-B to also meet the peak power requirements.

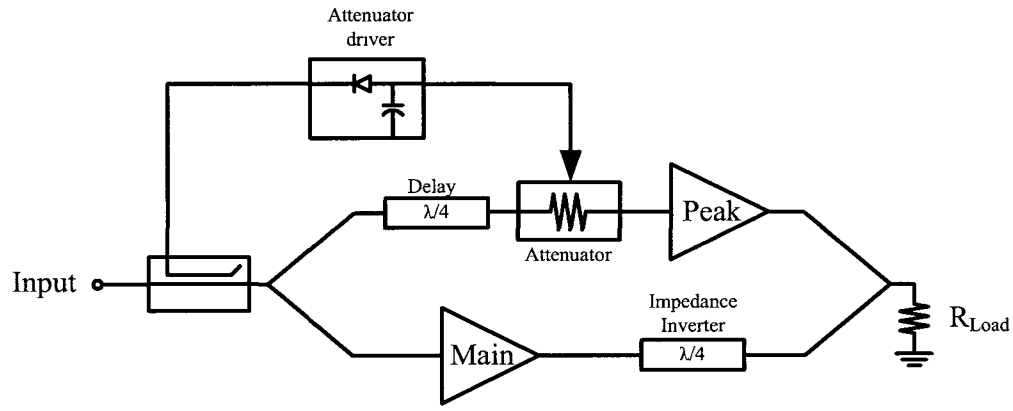


Figure 3.1: Doherty system with attenuator.

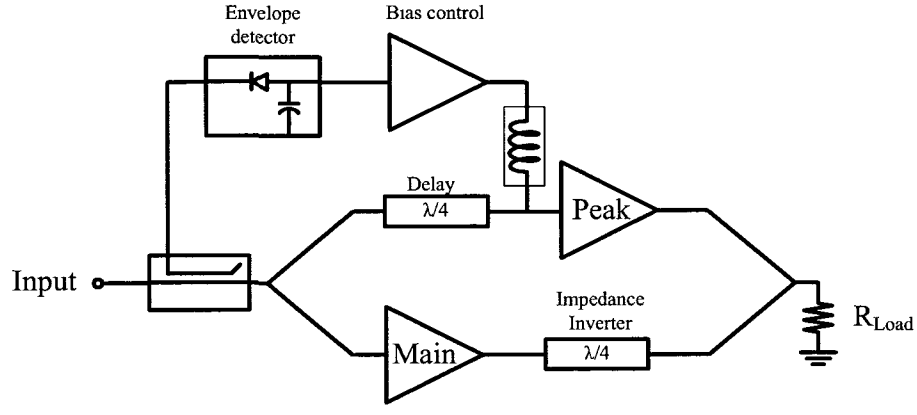


Figure 3.2: Doherty system with adaptive bias.

3.1.2 Generalizing the Doherty

The Doherty concept can be extended to handle back-off efficiencies other than the 6-dB in the classic circuit. This generalization is either termed *asymmetrical* Doherty [24], or *extended* Doherty [31], and is useful in applications that consist of modulated signals with PAPRs that are not exactly 6-dB. The generalized circuit consists of the same architecture as the classic case (see Figure 2.16), but the turn-on point and the maximum current of the peak stage are modified to start the load modulation effect at a different point. As analyzed by Iwamoto et al. in [31], the turn-on point of the peak PA ($V_{g\_critical}$) can be written as a function of the design parameter γ in (3.3). Since the load modulation effect needs to generate optimum efficiency during the high power regime, the maximum current of the peak amplifier needs to satisfy (3.4) which is now a function of γ.

$$V_{g\_critical} = \frac{1}{\gamma} V_{g\_\max} \tag{3.3}$$

$$I_{p\_\max} = (\gamma - 1)\, I_{m\_\max} \tag{3.4}$$

Therefore, the impedances in Figure 2.16 that provide optimum power combining are also functions of γ and are shown in equations (3.5) and (3.6) [31]. The optimum loads seen by the main PA (R_{m_OPT}) and the peak PA (R_{p_OPT}) at maximum power are no longer the same since both stages have different peak current values. Thus, the two PA stages must be matched such that they satisfy equation (3.7).

$$Z_0 = R_{m\_OPT} \tag{3.5}$$

$$R_{Load} = \frac{R_{m\_OPT}}{\gamma} \tag{3.6}$$

$$R_{p\_OPT} = \frac{R_{m\_OPT}}{\gamma - 1} \tag{3.7}$$

To demonstrate how this generalized case reduces to the classic Doherty, the design parameter γ can be replaced with a value of 2. Then the design equations (3.6)-(3.7) will reduce to (3.1)-(3.2). Also, in a case where a PA needs to efficiently handle a PAPR of 12-dB, equation (3.3) can be used to find that a γ of 4 is needed. Thus, the resulting current and voltage characteristics for this 12-dB high efficiency regime

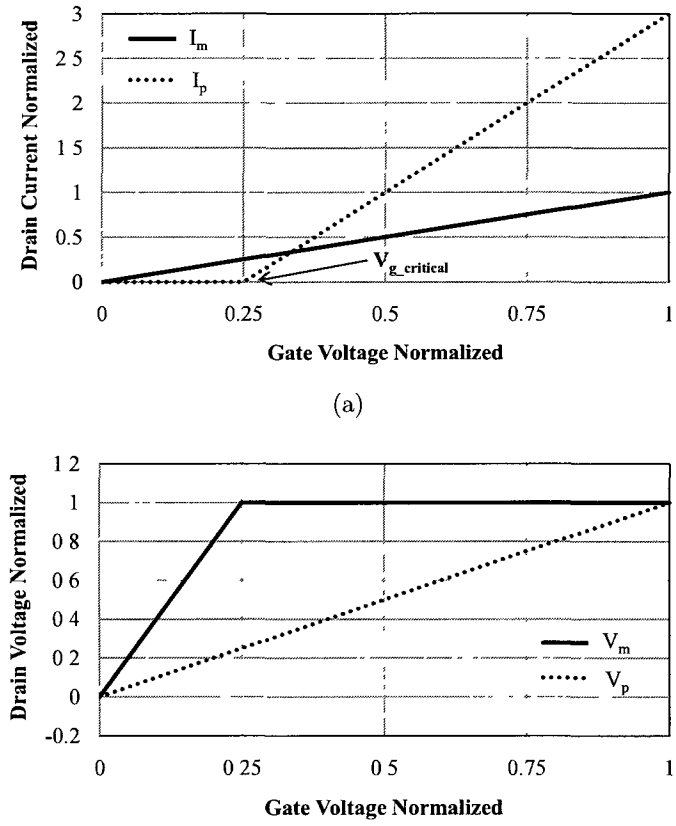


Figure 3.3: Ideal extended Doherty characteristics ($\gamma = 4$).

Doherty are shown in Figure 3.3. The efficiency advantage of the extended Doherty is compared to the classic Doherty in Figure 3.4.

3.2 High-Frequency Challenges

As it was shown by the literature review in Table 1.1, many Doherty PAs have been reported in the RF cellular bands, but there has been a lack of implementations at 10 GHz and above into the millimeter-wave range. This is mainly due to limitations of the active devices as it will be shown in this section. The processes available for this work consist of several deep sub-micron CMOS technological nodes and a 0.5-μm

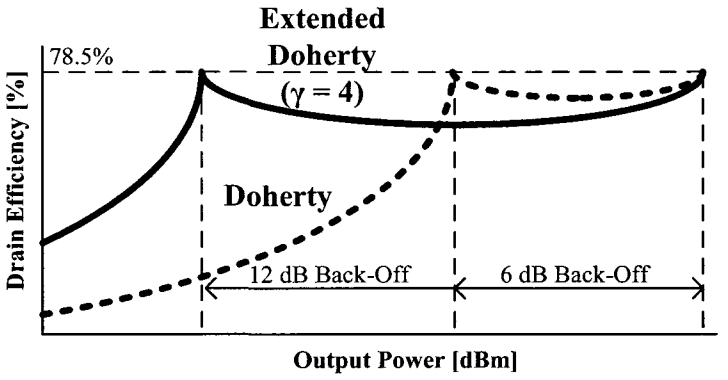


Figure 3.4: Extended Doherty efficiency characteristics ($\gamma = 4$).

GaN technology. Before introducing the high-frequency Doherty model in Section 3.3, the high-frequency process limitations of these technologies are discussed.

3.2.1 Deep Sub-micron CMOS Processes

The high f_T CMOS devices necessary to implement PAs beyond the standard RF bands have fundamental limitations in comparison to dedicated power devices. The most obvious drawback is the reduction of supply voltages (V_{DD}) in lower technology nodes due to the thinning of gate oxides. The supply voltage limits the upper bound of the output signal swing to $2V_{DD}$. Therefore, to generate the same output power with smaller feature size transistors, an increase in current is needed such that (3.8) remains constant. This larger current has two implications: 1) it requires larger width devices which could exhibit worse performance than smaller periphery devices and 2) it reduces the optimum matching impedances since the impedance is inversely proportional to current; consequently, higher Q and/or more complicated matching networks are needed to deliver the same power to a 50-Ω load. The two other major drawbacks to using deep sub-micron devices, as demonstrated through the I-V curves in Figure 3.5, are the large knee voltage and the finite output impedance.

$$P_{outmax} = \frac{(V_{DSmax} - V_{knee}) \cdot I_{DSmax}}{8} \tag{3.8}$$

The knee voltage is the lower bound to the voltage swing and in advanced CMOS processes it can impact as much as 40% of the total output voltage swing. Since it limits the voltage swing, the knee voltage will also result in a reduction of output

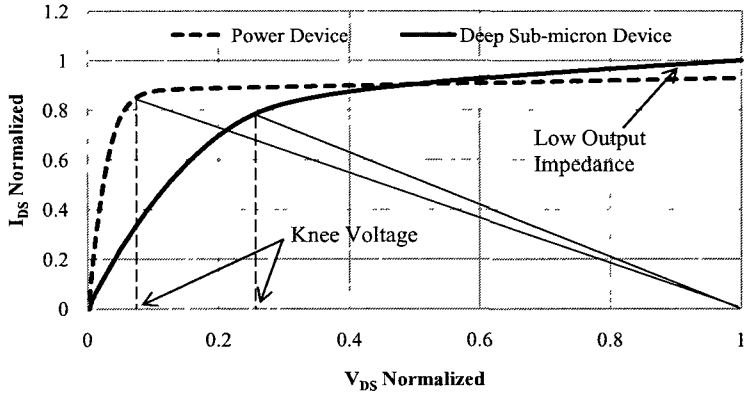


Figure 3.5: I-V curve for a deep sub-micron device and a power device [1].

power. Nonetheless, the reduction is less than predicted by (3.8) because deep sub-micron devices generate a significant portion of their output power in the triode region [32].

The output impedance (Z_{out}) is an important parameter for power amplifier designs because it shunts power away from the output load which affects the gain and the drain efficiency. At low frequencies it is dominated by the output conductance of the transistor (g_{ds}) [33]. The output conductance depends on the bias current and on the channel modulation parameter (λ) as shown by the square-law equations in (3.9) and (3.10).

$$i_{DS} = \frac{\mu C_{OX}}{2}\left(\frac{W}{L}\right)(v_{GS} - V_T)^2(1 + \lambda v_{DS}) \tag{3.9}$$

$$g_{ds} = \frac{\partial i_{DS}}{\partial v_{DS}} = I_{DS}\lambda \tag{3.10}$$

When transistors with a fixed width are implemented in lower process nodes, the output resistance scales due to the change in current but also due to the second order effects related to λ. Hypothetically, if the second order effects were not present, the optimum load and the output impedance would scale in the same manner resulting in a constant drain efficiency across device sizes. However, the channel modulation effect does exist and its impact becomes more severe as the channel length is reduced [19]. To understand how efficiency is affected across technological nodes (i.e. gate

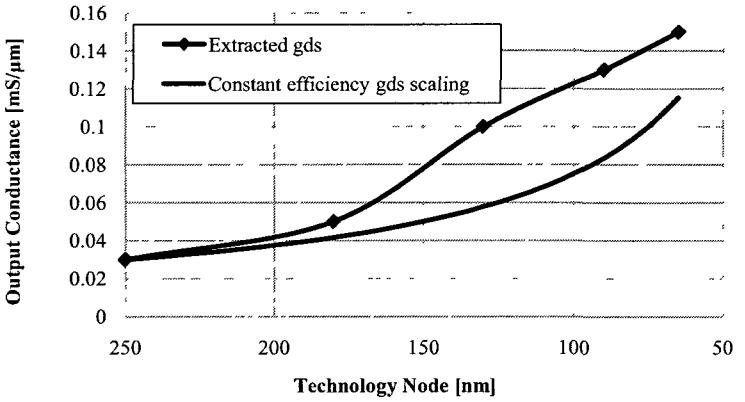


Figure 3.6: Output conductance across deep sub-micron technology nodes.

lengths), Figure 3.6 is used to compare the extracted g_{ds} of technological nodes 250-nm through 65-nm from the work in [34] to the hypothetical constant efficiency g_{ds}. The deviation of the extracted g_{ds} curve from constant efficiency g_{ds} curve results in efficiency degradation. Therefore, small feature size transistors perform worse in terms of efficiency and should be avoided if enough gain can be achieved with larger feature sizes.

The g_{ds} covers one dimension of the problem since it dominates the output impedance at low frequencies. At higher frequencies a more complete parasitic model of the transistor output should be considered. This equivalent output model as shown in Figure 3.7 [33] is similar to the one employed in the design kit BSIM4 models of the 90-nm process used in this work. The equivalent model can be made to fit the simulated Z_{out} of the 90-nm design kit transistors across all frequencies. To demonstrate the accuracy of the model, the output impedance curves for an arbitrary transistor are fitted and shown in Figure 3.8. The fit is achieved with the parameter values in Table 3.1. It can be noticed that the output resistance begins to drop-off in the gigahertz range and by the time it reaches millimeter-wave frequencies it is significantly lower than the original DC value. Therefore, correct modeling of the output parasitic network is absolutely necessary for circuit design at high-frequencies [33].

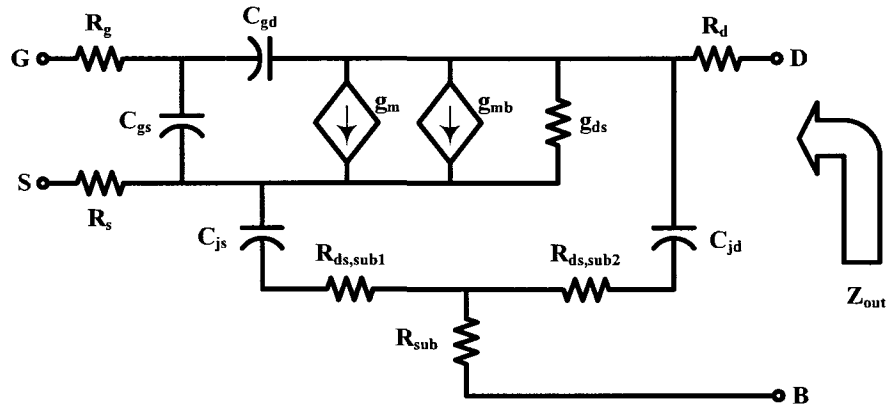
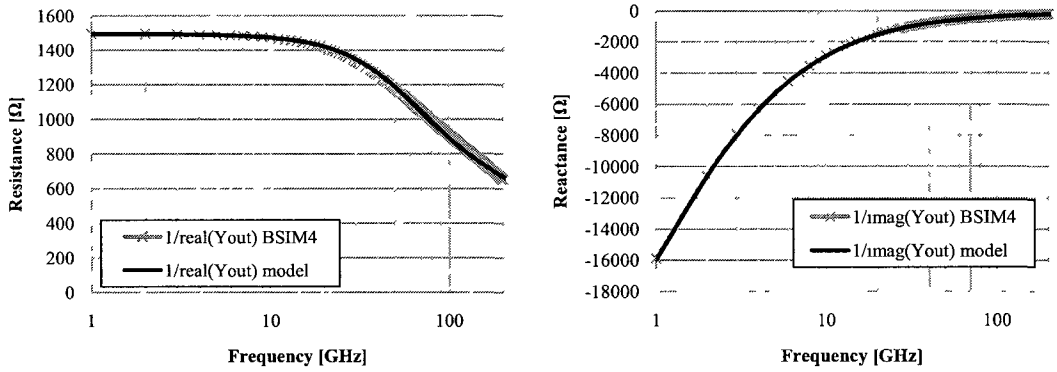


Figure 3.7: High-frequency output impedance model for deep sub-micron CMOS devices.

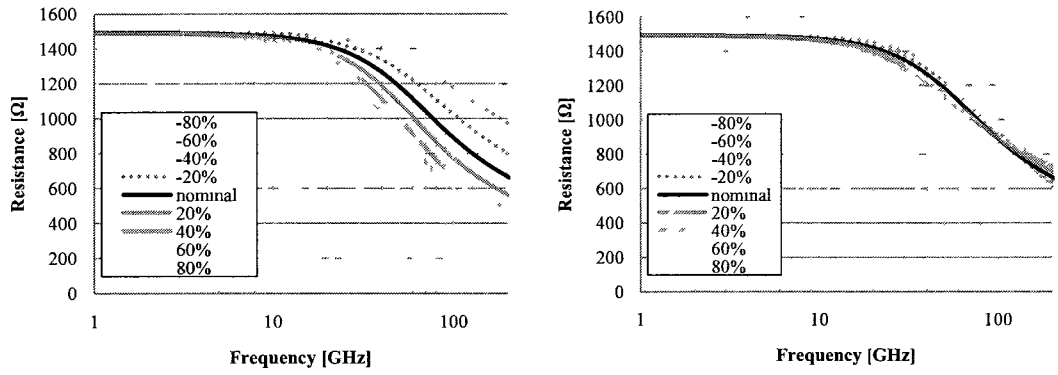


(a) $1/Re\{Y_{out}\}$. (b) $1/Im\{Y_{out}\}$.

Figure 3.8: Output impedance of 90-nm CMOS fitted with model.

Table 3.1: Parameter values for the output impedance network model fitting. The device used has UGW = 2 μm, NOF = 4.

Parameter	C_{jd}	C_{js}	g_{ds}	R_{sub}	R_d	R_s
Value	10.01 fF	26.40 fF	0.67 mS	80.27 Ω	2.80 Ω	3.95 Ω



(a) C_{jd}. (b) R_{sub}.

Figure 3.9: Change in output resistance with (a) C_{jd} and with (b) R_{sub}.

The output network model can now be analyzed to gain insight on how the transistor layout affects the output resistance. In Figure 3.9 the junction capacitance at the drain (C_{jd}) and the substrate resistance (R_{sub}) parameters from Table 3.1 are swept individually with 20% steps to quantify their effects on R_{out}. It can be noticed that C_{jd} is the dominant factor that affects the output resistance roll-off. As a circuit designer, not much can be done to reduce C_{jd} since it depends on the size of the drain junction which is limited by the process lithography. However, it can be expected that the finer lithography steps applied to reduce the channel length in lower technology nodes will also be used to reduce the C_{jd}. Therefore, future technology nodes could provide better drain efficiencies (and power-added-efficiencies) at millimeter-wave frequencies. The R_{sub} on the other hand can be affected by the layout configuration since R_{sub} depends on the distance between the substrate contact and the active region of the transistor. Although it may seem counter intuitive, increasing R_{sub} does in fact improve the output impedance at millimeter-wave frequencies. This occurs because the output network can be approximated with two poles and a zero; by increasing R_{sub}, the zero shifts to a lower frequency. It should be noted that increasing R_{sub} could potentially lead to latch-up problems. Therefore, when altering R_{sub} the latch-up design rules should always be verified.

The lower supply, the higher knee voltages, and the higher output conductances are all degradation mechanisms that directly impact the performance of high-frequency power amplifiers. These effects need to addressed on an architectural

and topological level. Other device effects such as mobility degradation, geometry-dependent stress, drain-induced barrier lowering (DIBL), and threshold variations become important in the sub-100-nm range [34]. Fortunately, these type of effects can be mitigated by employing constant-current-density design techniques [35].

3.2.2 GaN Processes

Although GaN is a relatively new semiconductor technology and its frequency of operation has only recently reached the millimeter-wave range, it promises performance that is unthinkable from mature technologies such as CMOS and GaAs. The wide bandgap nature of the GaN material allows for high output power densities, operation from high supply voltages, and operation in extreme environments [36]. Unlike the deep sub-micron CMOS technologies discussed in Section 3.2.1, GaN technologies have attributes that are very desirable from a power amplifier point of view.

In terms of performance limiting effects, the GaN high electron mobility transistors (HEMTs) are also limited by a knee voltage, but due to the large supplies it is much less of a concern. The main drawback to be investigated is the finite output impedance. Similarly to the CMOS transistors, the output resistance in GaN HEMTs rolls off at higher frequencies but the mechanism differs since the GaN substrate is non-conductive. The output conductance dispersion along with other effects such as transconductive dispersion, DC current collapse, and transient lags have been linked to various trapping mechanisms [37]. These trapping mechanisms influence performance by forming surface and bulk quasi-static charge distributions [37]. The empirical single trap model developed in [38] is extended in Figure 3.10 to multiple traps with different time constants.

Figure 3.11 shows the simulated output impedance curves of an arbitrary device from 0.5-μm GaN design kit fitted with the extended trap model. The fitting is done while attempting to preserve the non-trap related small-signal parameters. It can be noticed that the extended model with parameter values in Table 3.2 captures the frequency dispersion effects quite well. The trapping center occurring at a frequency of 100 kHz coincides with the trapping observed in [38]. On the other hand, the trapping center occurring around 1 GHz does not coincide with previously published work and could be specific to the process used.

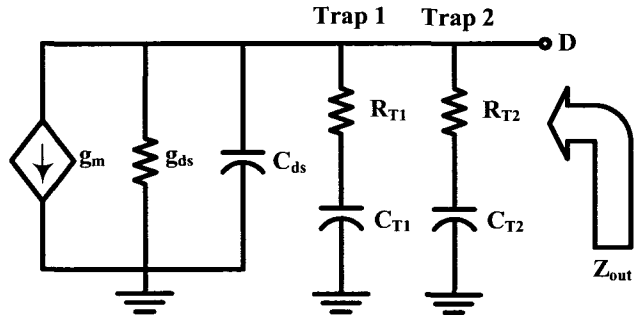
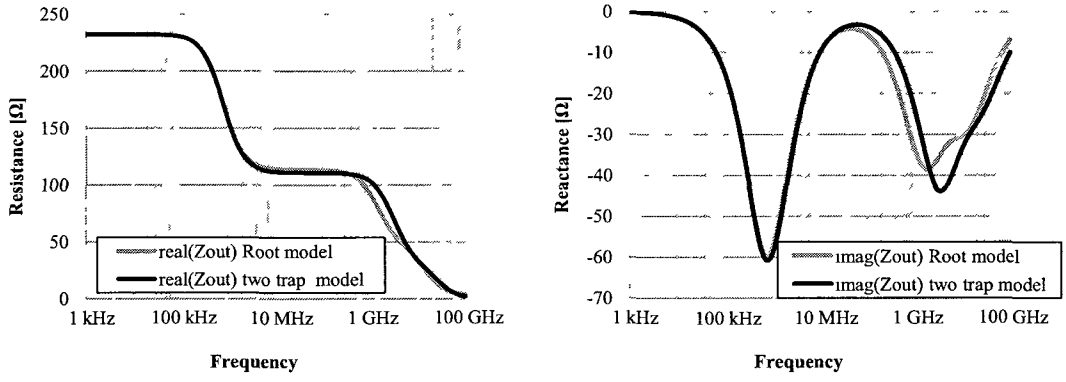


Figure 3.10: Two trap conductance dispersion model for 0.5-μm GaN HEMT devices.



(a) $Re\{Z_{out}\}$. (b) $Im\{Z_{out}\}$.

Figure 3.11: Output impedance of 0.5-μm GaN fitted with model.

Table 3.2: Parameter values for the two trap model fitting. The device used has UGW = 200 μm, NOF = 4.

Parameter	g_{ds}	C_{ds}	C_{T1}	r_{T1}	C_{T2}	r_{T2}
Value	4.3 mS	151.2 fF	452.5 pF	211.7 Ω	195.5 fF	94.29 Ω

3.3 High-Frequency Doherty Model

The high-frequency transistor effects discussed in Section 3.2 are now translated into a Doherty system model. The purpose of this model is to provide intuitive understanding of the Doherty system by introducing an intermediate step between the ideal mathematics and the schematic simulations. The main requirement for this

model is to capture the large and small-signal effects while being able to isolate them such that they can be quantified individually. Since the Doherty is an efficiency enhancement technique, this analysis concentrates on the parameters that affect the efficiency. Therefore, the non-linear characteristics of the transconductors and parasitic capacitors are almost irrelevant. This simplification is valid because the efficiency characteristics in many of the published Doherty systems [8–12,15,16] are determined only by a single tone analysis. For analysis with a modulated signal, this model would not capture all the compression mechanisms of the fundamental thus resulting in some inaccuracies.

The critical parameters for accurate single tone efficiency analysis of the Doherty are the knee voltage, the supply voltage limit, the output resistance, and proper current characteristics for the peak and main amplifiers. The current characteristics can be represented with equation (3.11) which is a simplified version of the Curtice-Ettenberg model [39]. Even though the Curtice-Ettenberg model is traditionally used for GaAsFETs, it is selected because of its simple representation of the knee voltage. The Curtice-Ettenberg model implementation part of the *Agilent ADS* environment [40] is used where additional parasitic effects such as output impedance and parasitic capacitances can be included. The simplicity of the model allows for isolation of non-ideal transistor effects so the significance of each can be evaluated individually. The threshold voltage, transconductance, knee voltage, and output impedance are controlled with the A_0, A_1, γ, and R_{rf} parameters, respectively. Figure 3.12 replicates the theoretical current curves from Figure 2.18 (a) with the Curtice-Ettenberg model. The A_0 and A_1 parameters of the main and peaking amplifiers are set such that load modulation effect begins when the input reaches a 50 % swing (6-dB power back-off).

$$I_{DS} = (A_0 + A_1 \cdot V_{GS} + A_2 \cdot V_{GS}^2 + A_3 \cdot V_{GS}^3) \cdot \tanh(\gamma \cdot V_{DS}) \tag{3.11}$$

This Curtice-Ettenberg model along with an ideal 90° hybrid coupler and two quarter-wave transformers for the output combiner are used to create the Doherty system model shown in Figure 3.13. Since ideal circuits are prone to convergence problems, a 1 MΩ drain-to-gate feedback resistor is included on both stages. Additionally, a 50Ω resistor is added at the gate of each stage to ensure perfect input matching at all frequencies.

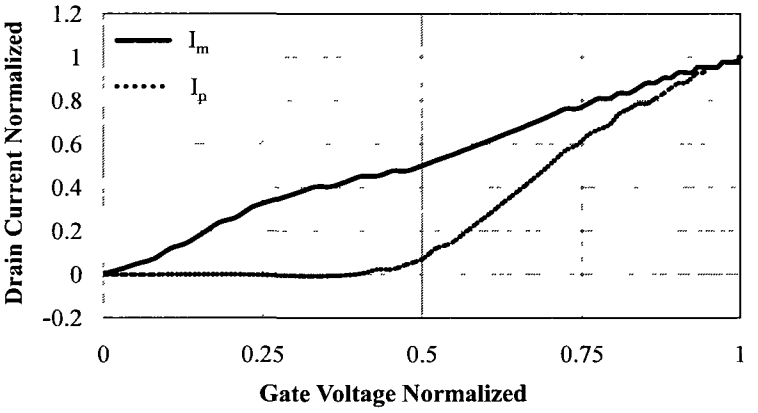


Figure 3.12: Doherty current characteristics with Curtice-Ettenberg [1].

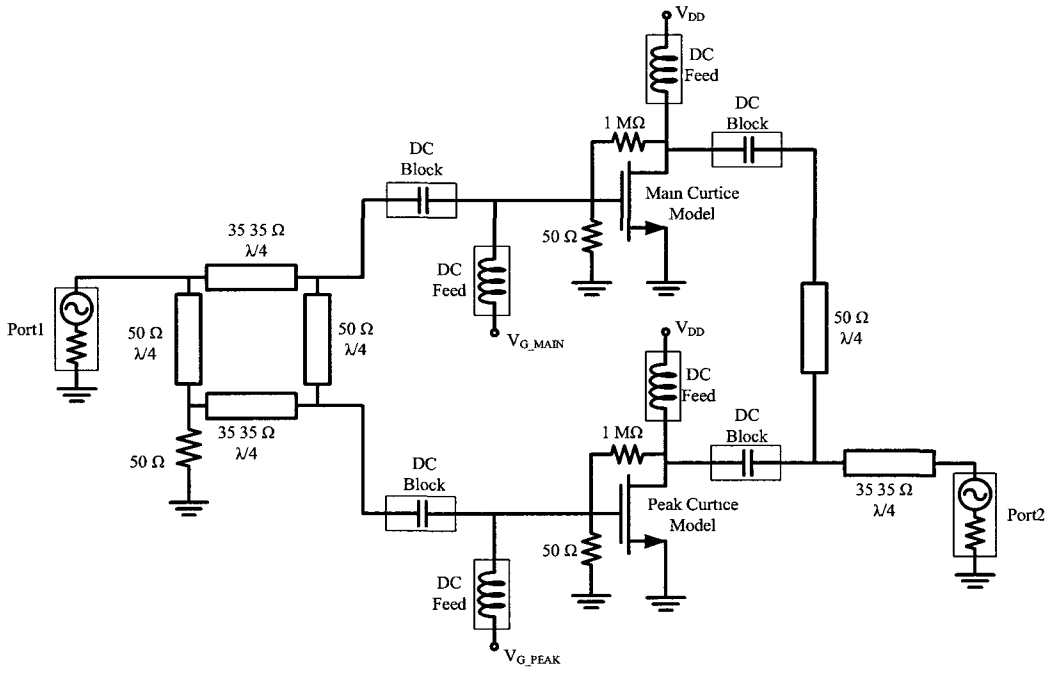


Figure 3.13: Curtice based Doherty system model.

3.4 High-Frequency Degradation on Doherty

In this section the efficiency of the Doherty is evaluated with the model from Figure 3.13. Then, suggestions are made on how to adapt the circuit for the non-ideal effects in order to minimize the degradation in performance. The first effect analyzed with the model is the output resistance.

3.4.1 Output Resistance

The r_{DS}/R_{OPT} ratio is used in this analysis to normalize the output resistance across device sizes. This ratio can also be used as an initial figure of merit for the efficiency of a certain process since for a fixed gate length the device width scales the output conductance and the drain current by the same factor. Therefore, the ratio of the output resistance (r_{DS}) to the optimum load resistance (R_{OPT}) remains constant and no advantage can be gained in terms of drain efficiency This is valid for low frequencies where the dominant effect on the output resistance is $1/g_{ds}$. At high frequencies, the r_{DS}/R_{OPT} ratio can vary across devices of the same technological node since the substrate resistance begins to have an effect on the output impedance. Regardless of how constant r_{DS}/R_{OPT} remains across different device sizes of the same node, it is clear that r_{DS} shunts power away from the load. The effect on the power added efficiency (PAE) as r_{DS} becomes comparable to R_{OPT} is shown in Figure 3.14.

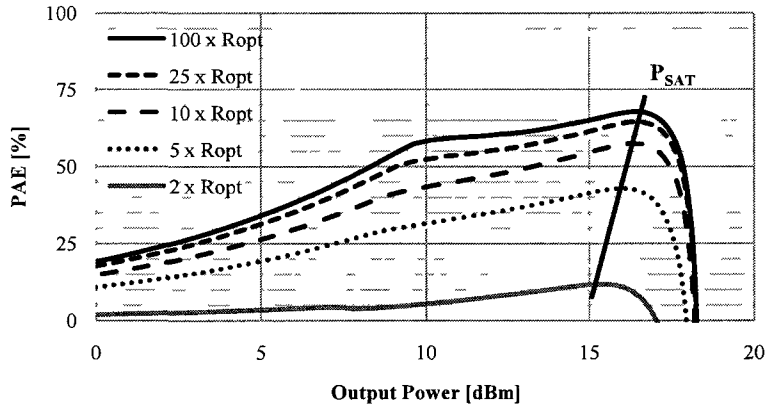


Figure 3.14: Effects of output resistance (r_{DS}) on efficiency [1].

In Figure 3.14 it can be observed that for a very high output impedance the classic two peak efficiency response is present (shown by the dark line). As the output impedance becomes lower, not only does the peak efficiency worsens but also the back-off peak becomes more subtle. This means that Doherty efficiency characteristics are more dependent on r_{DS} than conventional classes of PAs. Also, as r_{DS} decreases, the output saturates at a lower power which reduces the flatness of the gain versus input power response; this effect is caused by a larger portion of the RF power being dissipated internally.

To improve the gain linearity, the auxiliary amplifier must be biased to turn-on at a lower input power similarly to the extended Doherty from Section 3.1.2. However, this moves the auxiliary amplifier closer to class-B and reduces the efficiency even further. Furthermore, a low r_{DS} reduces the effectiveness of the active load-pull because less current makes it to the load. The effectiveness of the load-pull can be improved by increasing the width of the peak device.

3.4.2 Breakdown Voltage

Material breakdown occurs in both Si and GaN devices, but due to the lower bandgap of the material it imposes a much stronger limit on the Si devices. Furthermore, the gates of CMOS transistors are made up of a very thin insulating material that should not exceed electric fields greater than 1 V/nm (for silicon dioxide) [41]. Therefore, due to the stringent breakdown requirements in CMOS, the supply levels in this analysis are based on CMOS circuits. In this analysis all the other effects such as the knee voltage and the finite output impedance are removed thus Figure 3.15 only shows the effects of the supply voltage. To maintain maximum output power, the output matching network should be adjusted to present the lower R_{OPT} caused by the reduction in supply.

The analysis in Figure 3.15 can also be applied to GaN transistors as long as the supply voltages are reduced by the same relative values.

3.4.3 Knee Voltage

For this analysis, the knee voltage is defined as the percentage of the total V_{DS} voltage at which the transistor is not in saturation. The degradation in efficiency as the knee voltage increases is shown in Figure 3.16. It can be noticed that efficiency is only

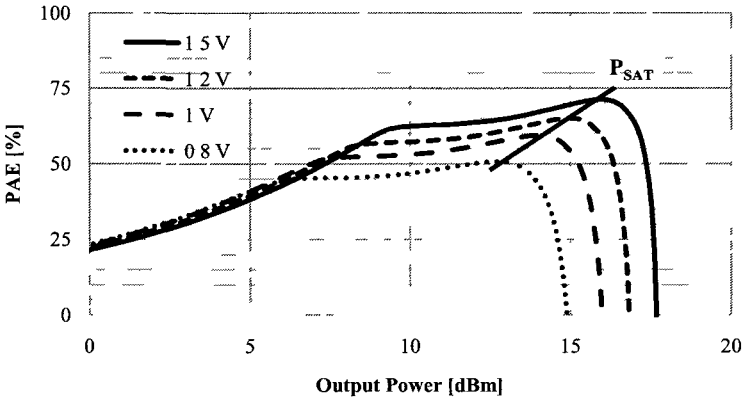


Figure 3.15: Effects of supply voltage on efficiency [1].

lost in the high power regime when the knee causes early compression A larger knee voltage also produces an earlier and much more gradual transition between the triode and saturation regions. This reduces the linearity of the circuit but it can be alleviated by widening the auxiliary device to provide more current. For example, a knee voltage that is 40% of V_{DS} requires an increase in current of roughly 25%.

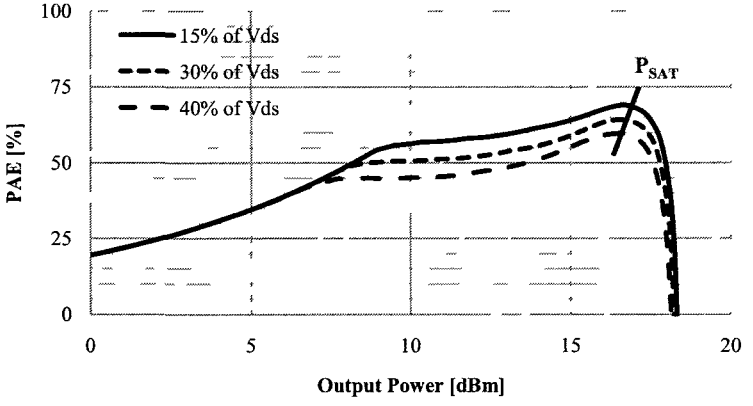


Figure 3.16: Effects of knee voltage on efficiency [1].

3.4.4 Output Reactive Parasitics

The original model from Figure 3.13 does not include any reactive components other than the output combiner. To capture the effects of the parasitic capacitance of the

device and the reactive matching network, a tank circuit is added at the drain of each Doherty stage. Ideal tank components are used since the losses that result from finite Q components can be considered as equivalent to shunt resistances and these losses have the same effect on efficiency as the analysis performed in Section 3.4.1. From Figure 3.17 it can be noticed that the tank circuit mainly impacts the efficiency characteristics beyond the P_{SAT} point where a significant amount of harmonics are generated. It is expected that the tank circuit will have a more significant effect if the analysis included a modulated signal instead of a single tone. In such a case the terminations at the harmonics would affect the odd order IM components landing near the fundamental. This analysis confirms that removing the reactive components from the model does not impact the trends.

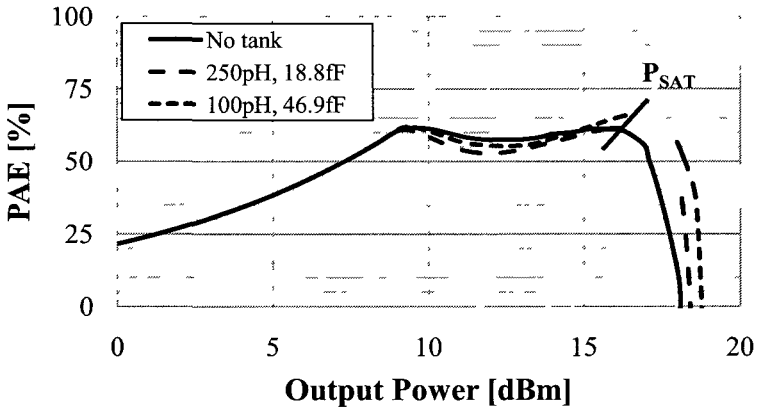


Figure 3.17: Effects of a drain tank circuit on efficiency.

3.5 Conclusions

In this chapter the high-frequency device limitations of deep sub-micron CMOS and GaN processes were investigated. It was found that factors such as output impedance, knee voltage, and supply voltage affect the performance of PAs. These factors were then incorporated into a Curtice-Ettenberg based model of the Doherty to quantify their effect individually. From the analysis it was found that at high-frequencies the finite output impedance has the biggest impact on the performance, and that it can reduce the ideal Doherty PAE by as much as 25%.

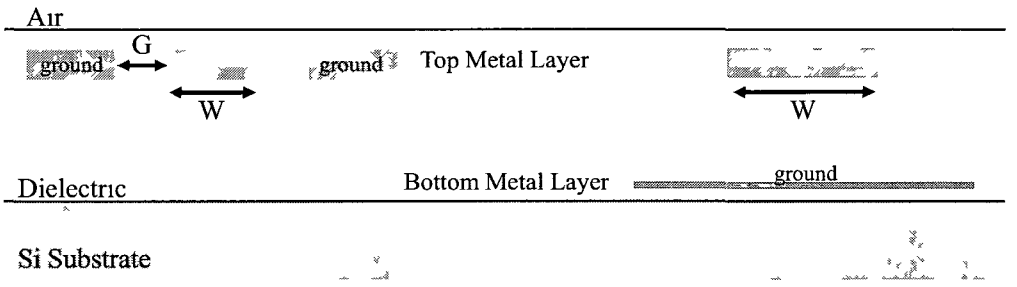
Chapter 4

Component Design and Characterization

This chapter discusses the design and characterization of passive and active components used in the Doherty implementations. In total, two IC technologies (CMOS and GaN) are used to implement the circuits in this work. Therefore both technologies are included in this chapter. The passive components are discussed in Section 4.1, the active components in Section 4.2, and finally, the de-embedding techniques used to compare the measured results to the simulations are discussed in Section 4.3.

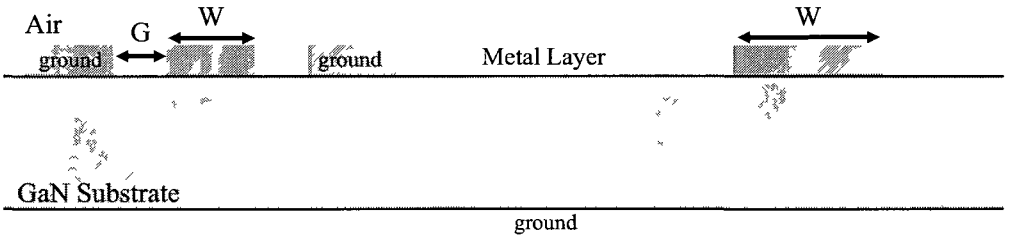
4.1 Passive Components

As previously mentioned in Section 2.1.3, low frequency IC's typically employ lumped matching techniques to save area. But as the frequency of operation increases and the guided wavelength becomes substantially shorter, it becomes practical to pursue matching techniques that involve distributed elements. Furthermore, the lumped assumption of capacitive and inductive elements begins to break down at high frequencies and their distributed effects become more pronounced. Therefore a hybrid methodology is employed in this work for the implementation of passive elements. This involves using transmission lines wherever possible and relying on lumped elements only in cases that result in a significant area reduction.



(a) CPW line. (b) Microstrip line.

Figure 4.1: CMOS implementation of (a) CPW lines and (b) microstrip lines.



(a) CPW line. (b) Microstrip line.

Figure 4.2: GaN implementation of (a) CPW lines and (b) microstrip lines.

4.1.1 Why Coplanar Waveguide (CPW) Structures?

Out of the various transmission line technologies, the most appropriate for MMIC's are microstrip lines and coplanar waveguide (CPW) lines [42]. The two transmission line technologies for the CMOS and GaN processes used in this work are shown in Figures 4.1 and 4.2. CPWs are used in this work since they have the following advantages compared to microstrip lines [42], [43]:

- CPW lines have more degrees of freedom since two dimensions affect the characteristic impedance - the width (W) of the center conductor and the gap (G) separating it from the ground traces. This results in a larger achievable characteristic impedance range.

- CPW lines have a larger inductive quality factor (Q_L) in cases such as silicon where the substrate has a low-resistivity. This is especially important since

inductive transmission lines are very often used to resonate the capacitive transistor parasitics.

- CPW lines are naturally shielded by the ground traces which allows more compact integration.
- The fundamental quasi-TEM mode of CPW lines exhibits less dispersion.
- CPW lines can be implemented with less process steps since they do not require VIA holes and are less dependent on the substrate thickness (so the substrate thinning step can be skipped).

Their main disadvantage is in terms of electromagnetic (EM) complexity since CPWs consist of three conductors as well as a ground plane at a finite distance. Due to the multiple conductors, CPW lines naturally support multiple zero cut-off frequency quasi-TEM modes. Also, in comparison to microstrip lines, higher order TE and TM modes can be excited more easily [42].

4.1.2 CPW-based Transmission Lines

Given that a transmission line satisfies quasi-TEM mode of propagation, it can be fully characterized by the appropriate frequency-dependent RLGC model [44],

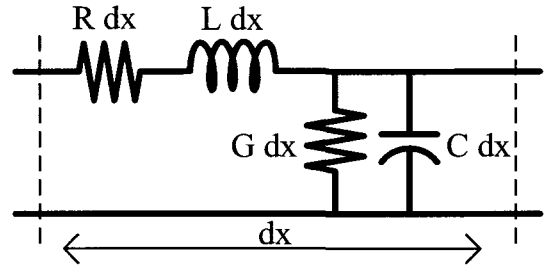


Figure 4.3: RLGC transmission line model.

where parameters R, L, G, and C are the per unit length series resistance, series inductance, shunt conductance, and shunt capacitance, respectively. From the RLGC model the transmission line can be characterized with the following equations [44]:

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \approx \sqrt{\frac{L}{C}} \tag{4.1}$$

$$\lambda_g = \frac{\lambda}{\sqrt{LC}} \tag{4.2}$$

$$\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{4.3}$$

When loss is small, the attenuation (α) can be approximated with (4.4).

$$\alpha \approx \frac{R}{2Z_0} + \frac{GZ_0}{2} \tag{4.4}$$

These RLGC based equations are valid for any quasi-TEM transmission line. But to size CPW lines, the analytical expressions for characteristic impedance (Z_0), as derived in [45], are shown in (4.5)-(4.8). These expressions are derived using conformal mapping and are valid for the ideal case where the CPWs are formed with sheet conductors. Nevertheless they demonstrate the trends that allow intuitive sizing of CPW lines.

$$Z_0 = \frac{30}{\sqrt{\varepsilon_{eff}}} \ln\left(2\frac{1+\sqrt{k}}{1-\sqrt{k}}\right), \frac{1}{2} \leq k^2 \leq 1 \tag{4.5}$$

$$Z_0 = \frac{30\pi^2}{\sqrt{\varepsilon_{eff}} \ln\left(2\frac{1+\sqrt[4]{1-k^2}}{1-\sqrt[4]{1-k^2}}\right)}, 0 \leq k^2 \leq \frac{1}{2} \tag{4.6}$$

where

$$k = \frac{G}{G + 2W} \tag{4.7}$$

$$\varepsilon_{eff} = \frac{1 + \varepsilon_r}{2} \tag{4.8}$$

From (4.5), (4.6), and (4.7) it can be noticed that the CPW gap-to-width (G/W) ratio determines Z_0, which means the same characteristic impedance can be achieved with various CPW sizes. This flexibility allows to optimize a given CPW for a certain performance measure (i.e. loss). Also, it can be noticed that a low G/W ratio results in low impedances and a high G/W ratio results in high impedances. These trends are now used to characterize three CPW transmission lines ($Z_0/\sqrt{2}$, Z_0, $\sqrt{2}Z_0$) for the 90-nm CMOS and 0.5-μm GaN processes used in the Doherty circuits. The 90-nm CMOS CPWs are implemented on the Top *Cu* Metal layer (M9), while the

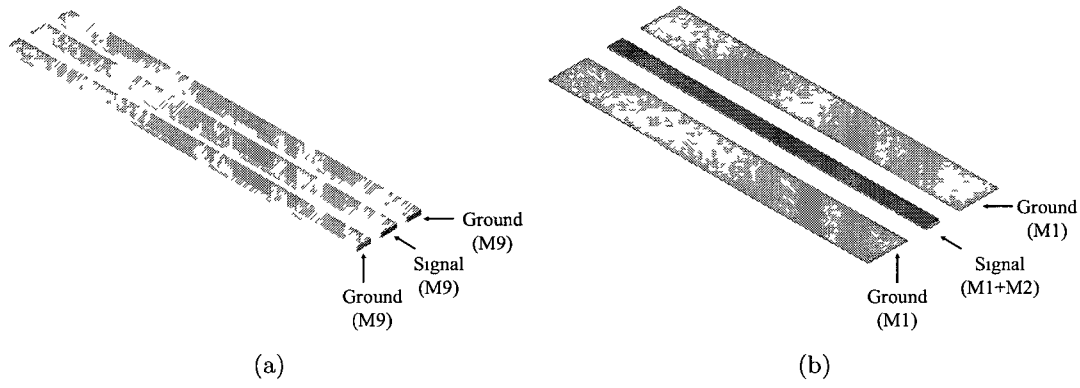


Figure 4.4: CPW lines in (a) 90-nm CMOS and in (b) 0.5-μm GaN.

0.5-μm GaN CPWs are implemented with both *Au* Metal Layers available in the process (M1 + M2); the layouts of the CPW lines are shown in Figure 4.4. Also, the dimensions and EM simulated performance from *Agilent Momentum* [46] are summarized in Tables 4.1 and 4.2.

Table 4.1: EM simulated results for 90-nm CMOS CPW lines.

W/G	12μm/3.5μm	12μm/7μm	8μm/12μm
Z_0	36.1Ω	51.3Ω	70.6Ω
Attenuation @ 75 GHz	1.03 dB/mm	0.73 dB/mm	0.67 dB/mm
Phase shift @ 75 GHz	191.7°/mm	178.3°/mm	181.9°/mm

Table 4.2: EM simulated results for 0.5-μm GaN CPW lines.

W/G	45μm/5μm	20μm/6μm	11μm/15μm
Z_0	39.8Ω	49.0Ω	70.7Ω
Attenuation @ 10 GHz	0.23 dB/mm	0.19 dB/mm	0.14 dB/mm
Phase shift @ 10 GHz	29.7°/mm	29.0°/mm	29.6°/mm

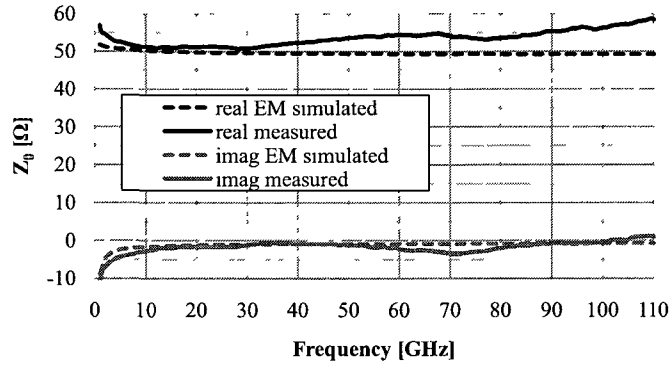


Figure 4.5: Measured and EM simulated characteristic impedance of 50-Ω CPW line in 90-nm CMOS (W = 12μm, G = 7μm).

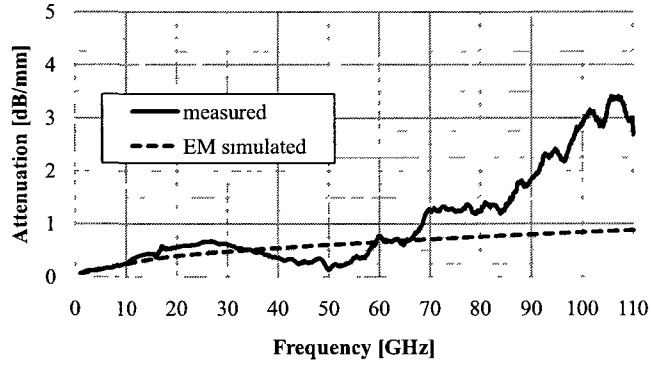


Figure 4.6: Measured and EM simulated attenuation of 50-Ω CPW line in 90-nm CMOS (W = 12μm, G = 7μm).

To validate the characterization process, a 50-Ω CPW line in 90-nm CMOS is measured and compared to the EM simulated results from *Momentum*. The measured CPW line is 100 μm long and the results in Figures 4.5, 4.6, and 4.7 show good agreement with simulations up to 80 GHz. The discrepancy beyond 80 GHz is caused by the relatively short CPW structure used for the measurements. For more accurate attenuation measurements it would be suggested to use 1 mm long CPW lines (similarly to the work in [43]).

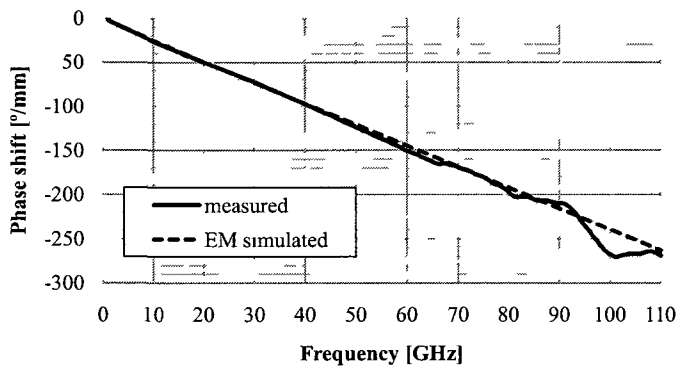


Figure 4.7: Measured and EM simulated phase shift of 50-Ω CPW line in 90-nm CMOS (W = 12μm, G = 7μm).

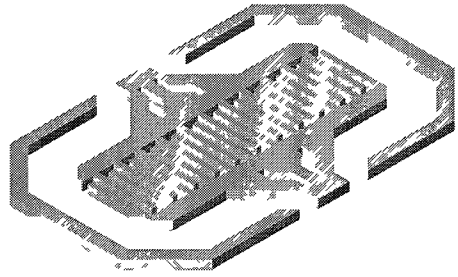


Figure 4.8: Custom CPW-based AC coupling capacitor in 90-nm CMOS

4.1.3 AC Coupling Capacitors

Standard CMOS processes typically include two types of capacitors Metal-Insulator-Metal (MIM) capacitors and Metal-Oxide-Metal (MOM) capacitors The MIM capacitors rely on additional masking steps that include high-K dielectric films, while MOM capacitors are implemented with the standard oxides layers [47] Since MOM capacitors employ less process steps they result in a lower cost. According to the foundry work published in [47], the MIM and MOM capacitors in the 90-nm process show similar densities, while the MOM shows better temperature dependency but slightly worse matching across samples From a high frequency power amplifier point of view there isn't much of an advantage to using MIM capacitors; therefore the cheaper MOM capacitors are used for the CMOS Doherty implementation.

Figure 4.8 shows the layout of the AC coupling MOM capacitor used for the 71 to 76 GHz CMOS Doherty amplifier The self-resonant frequency (SRF) in AC coupling capacitors is not a strict design requirement as long as 1) the capacitor blocks the DC

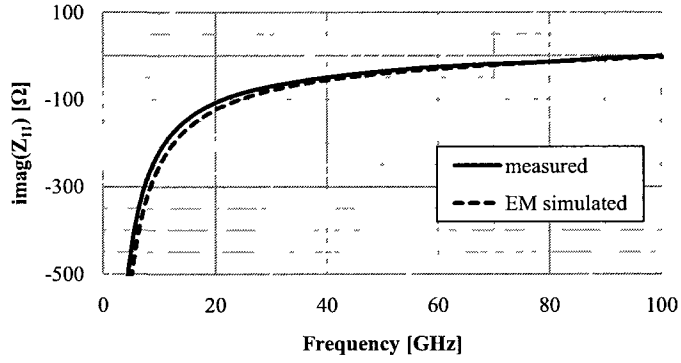
signal and 2) the precise impedance is known at all frequencies. However, only a single 90-nm CMOS fabrication run was available for this implementation and thus to ensure predictable performance the AC coupling capacitor was optimized for a high SRF. The required SRF is achieved by employing many short fingers in parallel. This results in the largest capacitance while maintaining the SRF high and the losses to a minimum. Comparison of the measured results with the EM simulations from *Momentum* is shown in Figure 4.9. The results show good matching between measurements and simulations for the reactive part of Z_{11} but a discrepancy for the resistive part of Z_{11}. It is probable that this discrepancy is due to either de-embedding error since the magnitude of the reactive part is much larger, or to a larger metal conductance that underestimates loss in the EM simulation. To verify the cause of the disagreement more test structures would need to be measured.

For the 10 GHz Doherty implementation in GaN, AC coupling is performed off-chip since the wave-length is too long to implement a high power bias network on-chip.

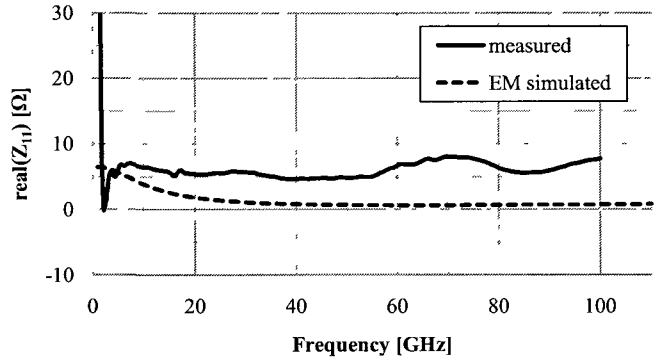
4.1.4 Inductors

The main function of an inductance is to resonate the capacitive parasitics of active devices [19]. This allows the actives to be used at frequencies much closer to their f_{MAX}. An inductance can be generated with either a lumped inductor or with a distributed transmission line. At lower frequencies lumped inductors are preferred due to their compact footprint. However, in the mm-wave range elements with higher inductive Q's can be generated by employing distributed transmission lines [43]. Therefore, the cancelation of parasitics at 71-76 GHz for the Doherty implementation in CMOS is performed with CPW lines and the cancellation at 10 GHz for the GaN version of the Doherty is performed with lumped inductors.

Figure 4.10 shows the layout of a CPW-based lumped octagonal inductor designed in the 0.5-μm GaN process. The inductor is designed to have a large hollow center region to increase the positive mutual inductance. Also, both Au Layers (M1 + M2) are used to minimize the conductor losses thus resulting in a larger Q. Unlike the AC coupling capacitor, the SRF is actually important since a constant inductor value is needed to resonate out the transistor's capacitive parasitics. Therefore the distance between each turn and the distance away from the ground planes is set such that the resulting SRF is much higher than 10 GHz.



(a) $Im\{Z_{11}\}$.



(b) $Re\{Z_{11}\}$.

Figure 4.9: Measured and EM simulated Z_{11} of capacitor in 90-nm CMOS.

A summary of the inductor dimensions is shown in Table 4.3. The inductance and Q as simulated with *HFSS* [48], *Momentum* [46], and *Sonnet* [49] are shown in Figure 4.11. All three EM simulators are setup with identical parameters, so the small discrepancy between the results must be from differences in the solution method.

Table 4.3: Dimensions of 600 pH inductor in 0.5-μm GaN.

sides	n	line width	line spacing	outer dimension	distance to ground
8	1.5	10 μm	5 μm	150 μm	30 μm

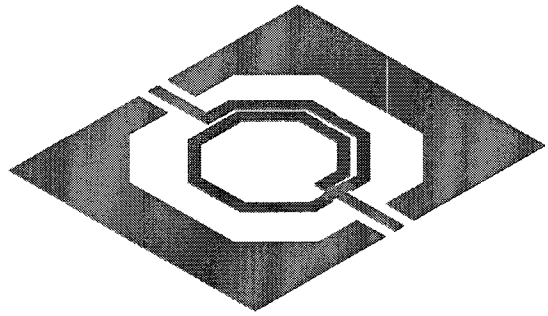
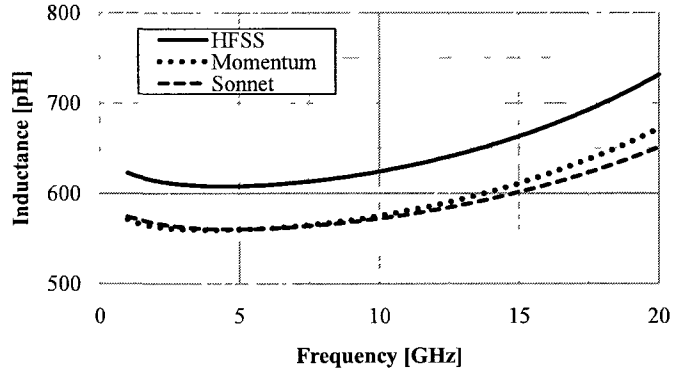
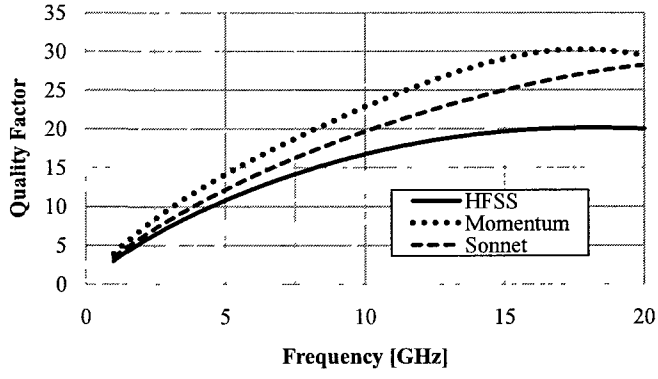


Figure 4.10: Custom CPW-based spiral inductor in 0.5-μm GaN.



(a) Inductance.



(b) Quality factor (Q).

Figure 4.11: EM simulated performance of inductor in 0.5-μm GaN.

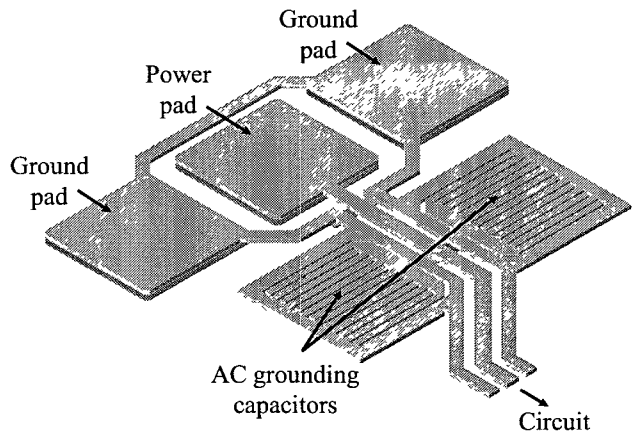


Figure 4.12: Custom CPW-based supply decoupling network in 90-nm CMOS.

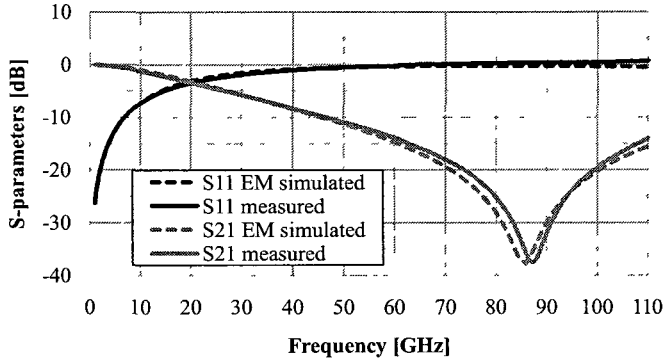
4.1.5 Supply Decoupling

Supply decoupling networks reduce the circuit dependance on signals and external impedances appearing on the supply. The decoupling is achieved by closing the AC loop locally (on-chip) through an element that creates an AC short between supply and ground. When the AC loop is closed locally, the supply noise as well as any impedance appearing at the supply are eliminated by grounding them. This ensures that the circuit is predictable and its behaviour reproducible.

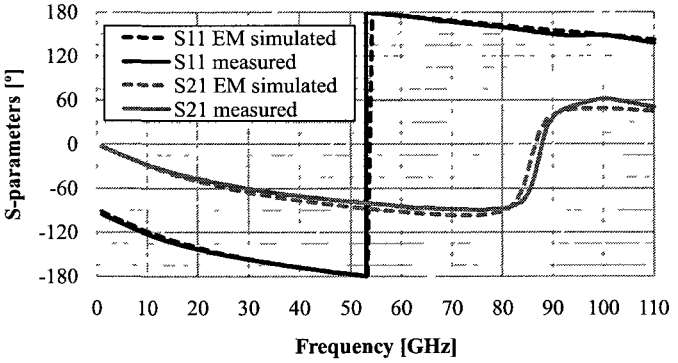
Either a large lumped capacitor or an open quarter-wave transmission line can be used to produce the AC ground. The main difference between the two is the bandwidth and the required layout area. For an integrated design, an open quarter-wave line appearing on each pads costs too much in terms of real-estate hence typically large lumped capacitors are used instead. As opposed to the capacitors discussed in Section 4.1.3, the design of these capacitors is much simpler because the capacitive parasitic to ground is actually desired. The main figure of merit for the AC grounding component is the resistance to ground in the band of interest.

Figure 4.12 shows the layout of the CPW-based supply decoupling network used for the 71-76 GHz CMOS Doherty amplifier. Implementing this network with CPW lines instead of microstrip lines results in a much more elegant design since the grounds appear close to the signal line. The measurements and simulations are performed on a CPW line that is AC grounded with the capacitors from Figure 4.12. Comparison of the measured S-parameters to the EM simulations from *Momentum* is shown in

Figure 4.13. It can be noticed that the results in Figure 4.13 agree very well with simulations. The series resistance to ground over the bandwidth is EM simulated to be from 0.8 to 1.0Ω and measured to be from -0.1 to -0.6Ω. The measured negative resistance is possibly due to calibration error.



(a) Magnitude components of S_{11}, S_{21}.



(b) Phase components of S_{11}, S_{21}.

Figure 4.13: Measured and EM simulated S-parameters of supply decoupling network in 90-nm CMOS.

Figure 4.14 shows the layout of the CPW-based supply decoupling network used for the 10 GHz GaN Doherty amplifier. Since the bias network for this version is implemented off-chip on the PCB, the size constraint is much less direct. Therefore the space allows for the use of quarter-wave long radial stubs instead of large capacitors. From the EM simulation results, this supply decoupling network produces a 0.3Ω series resistance to ground at 10 GHz.

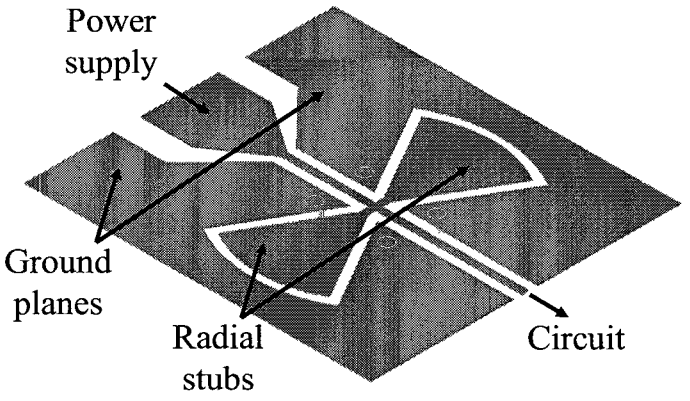
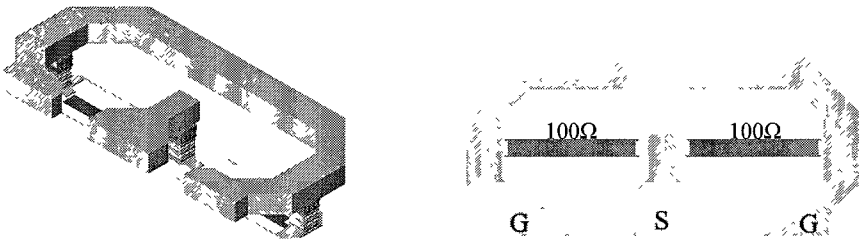


Figure 4.14: Supply decoupling network implemented with radial stubs on 6010 RT/duroid material.



(a) Isometric view. (b) Top view.

Figure 4.15: Custom CPW-based 50-Ω resistive termination in 90-nm CMOS.

4.1.6 Resistive Termination

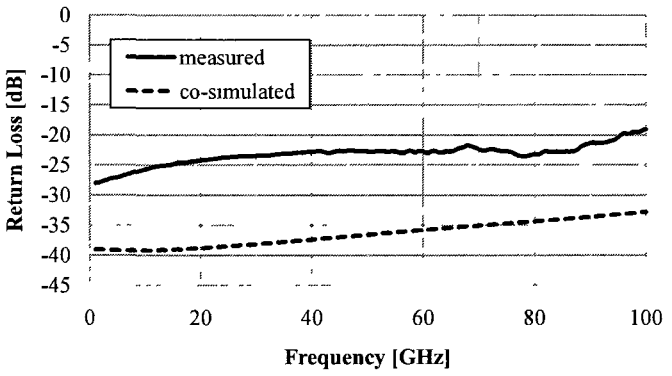
Resistors are often avoided in power amplifiers since they are lossy and reduce the efficiency. However in some cases they are absolutely necessary. One such case is the resistive termination present in directional four-port couplers and Wilkinson dividers. The reason for pursuing couplers/dividers that require resistive terminations is because it is impossible to have a three-port power divider/coupler that is simultaneously lossless, reciprocal, and matched at all ports [44]. The power division networks employed to drive the two parallel stages of the Doherty circuits in this work are implemented with four-port branch-line couplers. Since the coupler for the 71-76 GHz 90-nm CMOS implementation is designed to work as a 50-Ω element it also requires a 50-Ω termination. As the frequency is increased, the difficulty of achieving a purely

50-Ω termination also increases because of the parasitics. Therefore the strategy employed to design the termination at 71-76 GHz consists of using a distributed feed network that has a characteristic impedance close to 50-Ω. This can be achieved by spacing the vias feeds such that they act as a continuation to the CPW transmission line. The overall feed structure reduces the discontinuities that the signal will encounter thus introducing almost no reflections. For symmetry reasons and to improve the circuit robustness two parallel 100-Ω resistors are used instead of a single 50-Ω resistor. Figure 4.15 shows the resulting layout of the 71-76 GHz termination. Comparison of the measured return loss to the co-simulated return loss is done in Figure 4.16. The co-simulation is performed in the *Agilent ADS* environment and it consists of an EM simulation of the feed-network in *Momentum* connected to the design kit model of the 100-Ω resistor. Even though the measured return loss is about 12 dB worse than simulated, it still maintains a return loss better than 20 dB from 0 to 100 GHz. Also, the measured return loss is flat proving that there are no resonating discontinuities in the feed structure.

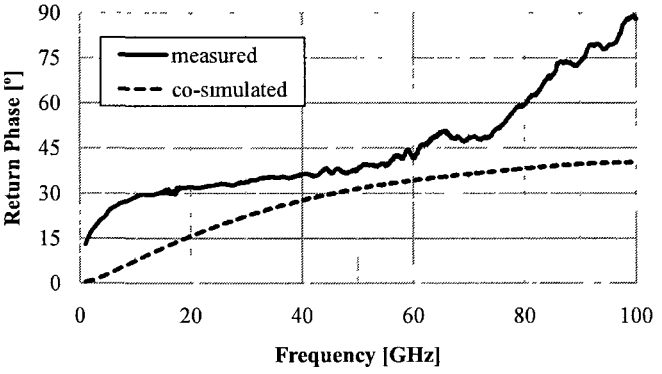
4.1.7 Non-Loading RF Pad Structures

To ease testing, distributed pad structures with a 50-Ω characteristic impedance are employed. Since all the circuits in this work are designed to be part of a 50-Ω system, the pads do not load the circuits reactively. This eliminates any frequency shifts caused by standard pad capacitances. The width and gap dimensions of sections A and C in Figure 4.17 are designed similarly to standard CPW lines, whereas the section B transition is adjusted to have an effective inductance and capacitance resulting to 50-Ω. Verification of the pad structures design can be done by electromagnetically simulating the *ABCD* parameters of the whole structure in Figure 4.17 and computing the characteristic impedance (Z_0). If the desired Z_0 is met then sections A, B, and C inherently satisfy (4.9) and the pad structures will only exhibit an insertion loss with a delay.

$$Z_0 = \left(\sqrt{\frac{L}{C}}\right)_A = \left(\sqrt{\frac{L_{eff}}{C_{eff}}}\right)_B = \left(\sqrt{\frac{L}{C}}\right)_C \tag{4.9}$$



(a) Magnitude component of S_{11}.



(b) Phase component of S_{11}.

Figure 4.16: Measured and EM simulated S_{11} for 50-Ω resistive termination in 90-nm CMOS .

Figure 4.18 shows the non-loading RF pads designed for 130-nm CMOS circuits as well as the ones designed for 0.5-μm GaN circuits. The measured results for two sets of 130-nm CMOS pads connected in a *Thru* configuration are compared to the EM simulated results produced in *Momentum* in Figure 4.19. It can be observed that the measured return loss is better than 18.3 dB up until 110 GHz proving that the non-loading pads are indeed 50-Ω. The measured insertion loss per pad from 30-110 GHz ranges between 0.08 and 0.55 dB.

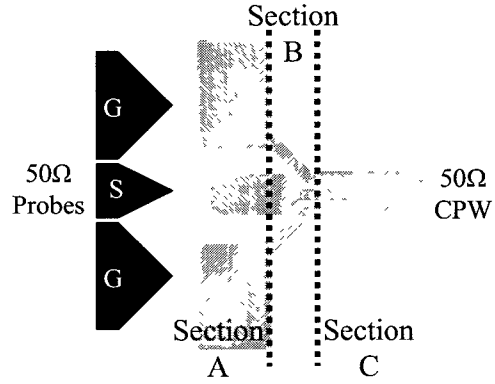
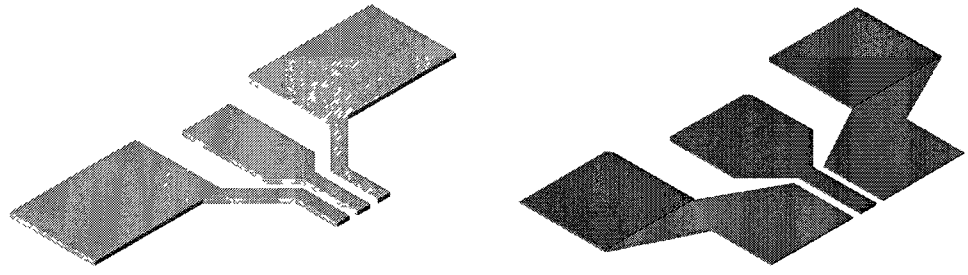
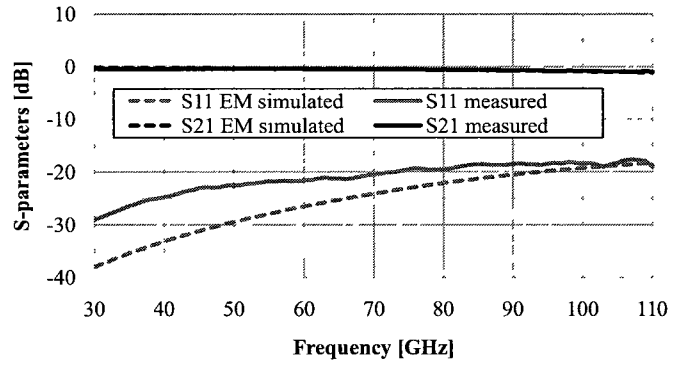


Figure 4.17: Non-loading pad structures.

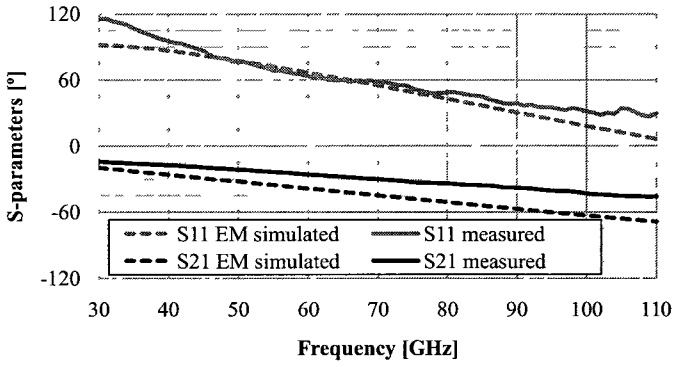


(a) (b)

Figure 4.18: Non-loading RF pads in (a) 130-nm CMOS and (b) 0.5-μm GaN.



(a) Magnitude components of S_{11}, S_{21}.



(b) Phase components of S_{11}, S_{21}.

Figure 4.19: Measured and EM simulated S-parameters of 130-nm CMOS RF pads in a *Thru* configuration.

4.2 Active Components

Due to time constrains, the 90-nm CMOS and the 0.5-μm GaN Doherty designs are both implemented with single fabrication runs. Therefore, the design of actives is heavily based on available RF models through the design kits instead of dedicating separate fabrication runs for parameter extraction as in [43]. The design methodology involves finding the optimum devices through a comparative study of various transistor peripheries, followed by the electromagnetic design of the suitable transistor feed-networks. The performance measures used to find the optimum devices are power gain, efficiency, and ability to match the input. This procedure is used for the active devices in the 90-nm CMOS process as well as the 0.5-μm GaN process.

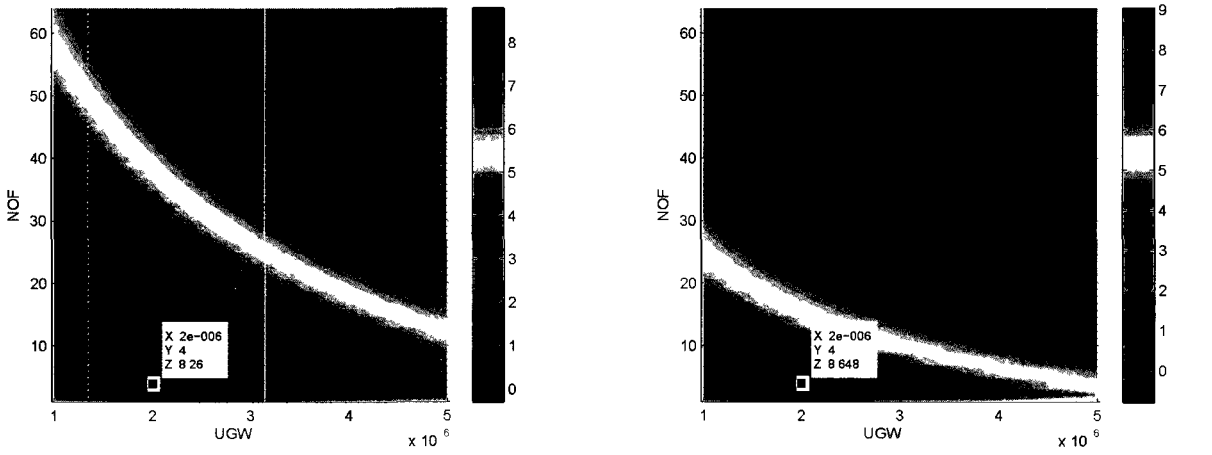
4.2.1 Power Devices in 90-nm CMOS

To compare the gain of various 90-nm devices, the maximum stable gain (MSG) figure of merit is used [44]. MSG is meaningful since it presents the amount of gain that can be actually used in an amplifier. The surface plots in Figure 4.20 show the MSG at mid-band (73.5 GHz) for two very similar cases (both biased in class-A); (a) shows the MSG for a common-source (CS) device and (b) shows the MSG for a CS device with a degenerative inductor of 25 pH. For this surface and for all following surfaces, the x-axis is the number of fingers (NOF), the y-axis is the unit gate width (UGW), and the red region represents the desired performance. Inductive degeneration is used to represent the non-ideal ground that is likely to be present on the source terminal of the actual circuit implementation. It can be noticed from Figure 4.20 (b) that the high gain region shrinks when degeneration is present. Therefore for adequate gain, especially with the presence of parasitics, the device size has to be selected somewhere in the red region of Figure 4.20 (b).

The next significant parameter is efficiency. In Section 3.4.1, it was found that the ratio between the output resistance and the optimum load (r_{ds}/R_{OPT}) plays an important role in the efficiency characteristics of the Doherty. Thus, in Figure 4.21 (a) the r_{ds}/R_{OPT} figure of merit is used to characterize the efficiency of various 90-nm devices at 73.5 GHz. The red region of this surface plot represents the highest r_{ds}/R_{OPT} ratio - meaning that the devices in that region have the smallest portion of their total RF power lost through r_{ds}.

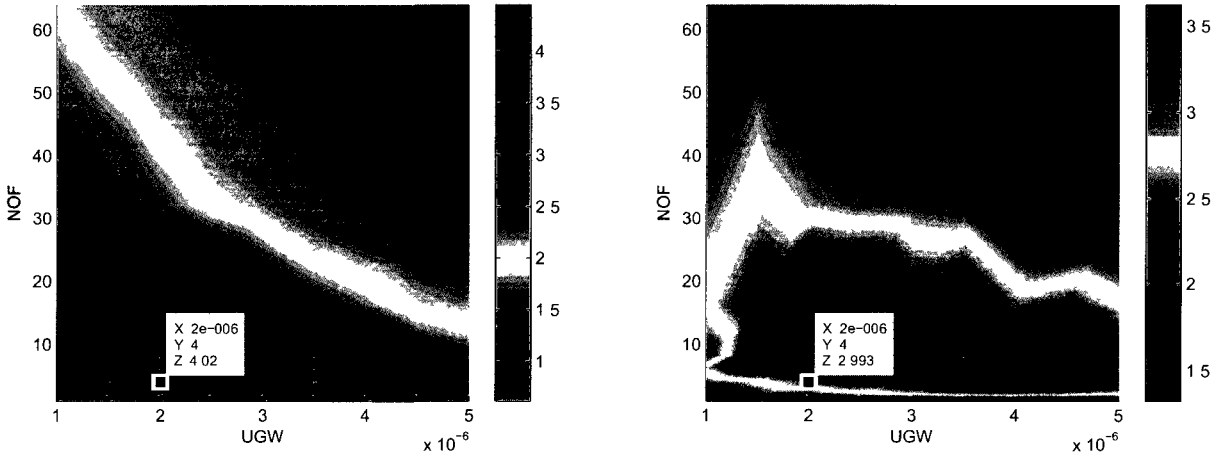
Finally, to determine the difficulty of designing the input matching network, the input series Q is plotted on the surface in Figure 4.21 (b). Unlike the MSG and the efficiency plots which show best performance for smaller device sizes, the input series Q surface shows that larger devices are the easiest to match. As a result, this leads to a trade-off between gain/efficiency and input matching that can be attributed to gate resistance (r_g).

The 71-76 GHz frequency range at which the CMOS Doherty is implemented represents a significant portion of the process f_T and f_{MAX}. Thus, the optimum device is mainly selected for gain but the efficiency and matching are also considered. By considering all surface plots, the best suited unit device for the Doherty is selected to have four fingers and a UGW of 2 μm. This device produces 8.26 dB of power gain while having a r_{ds}/R_{OPT} ratio of about 3 and an input series Q of about 4. Since the power produced by this unit device is less than the 10 dBm specification, four of these devices are combined in parallel through the feed-network shown in Figure 4.22. The feed-network is designed to interface the Top Metal layer (M9) CPW lines with the active devices laying on the substrate. The design methodology of the feed-network is based on maintaining equal delay to all transistor fingers and keeping the discontinuities to a minimum. From Figure 4.22 it can be noticed that the ground traces of the CPW feeds are kept closer to the substrate while the signal traces are kept away from the substrate as much as possible. This minimizes capacitive loading at the gate and drain of the transistor which results in a lower f_T drop.



(a) Common-source. (b) Common-source with 25 pH of degeneration.

Figure 4.20: Surface plot of maximum stable gain (MSG) (a) in common-source configuration and (b) in common-source with 25 pH of degeneration for various devices sizes in 90-nm CMOS at 73.5 GHz (bias: $V_{GS} = 0.75$ V, $V_{DS} = 1.5$ V).



(a) input series Q (b) r_{ds}/R_{OPT} ratio.

Figure 4.21: Surface plot of (a) input series Q and (b) r_{ds}/R_{OPT} ratio for various device sizes in 90-nm CMOS at 73.5 GHz (bias: $V_{GS} = 0.75$ V, $V_{DS} = 1.5$ V).

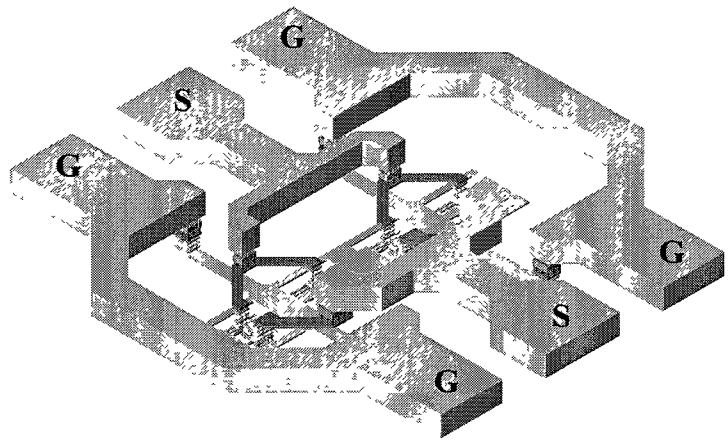


Figure 4.22: CPW-based feed-network for the power device in 90-nm CMOS.

4.2.2 Power Devices in 0.5-μm GaN

Similarly to the CMOS device selection, the GaN devices are compared with the same figure of merits: MSG, r_{ds}/R_{OPT}, and input series Q. The surface plots in Figures 4.23, 4.24 (a), and 4.24 (b) show the three figure of merits for all available HEMT device sizes. Again, for all surface plots the x-axis is the NOF, the y-axis is the UGW, and the red region represents the desired performance. However, in comparison to using the 90-nm CMOS process at 71-76 GHz, the 0.5-μm GaN process allows much more power gain when it operates at 10 GHz. Thus, the device sizes are mainly optimized for efficiency and matching. From the surface plots, the large Q's of small UGW devices ($\leq$ 100 μm) make them difficult to match, while the power gains and r_{ds}/R_{OPT} ratios of large UGW devices (300 μm) are poor. This results in an optimum UGW of 200 μm. The NOF on the other hand has marginal effects on the figures of merit so the NOF parameter is used to set the output power. Unlike in CMOS, feeding the GaN HEMTs is much simpler since the metallization layers are practically on the same level as the actives. The feed-networks shown in Figure 4.25 are designed as an extension to the CPW lines.

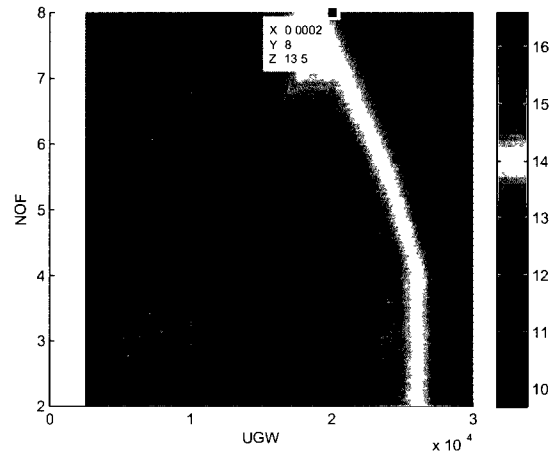
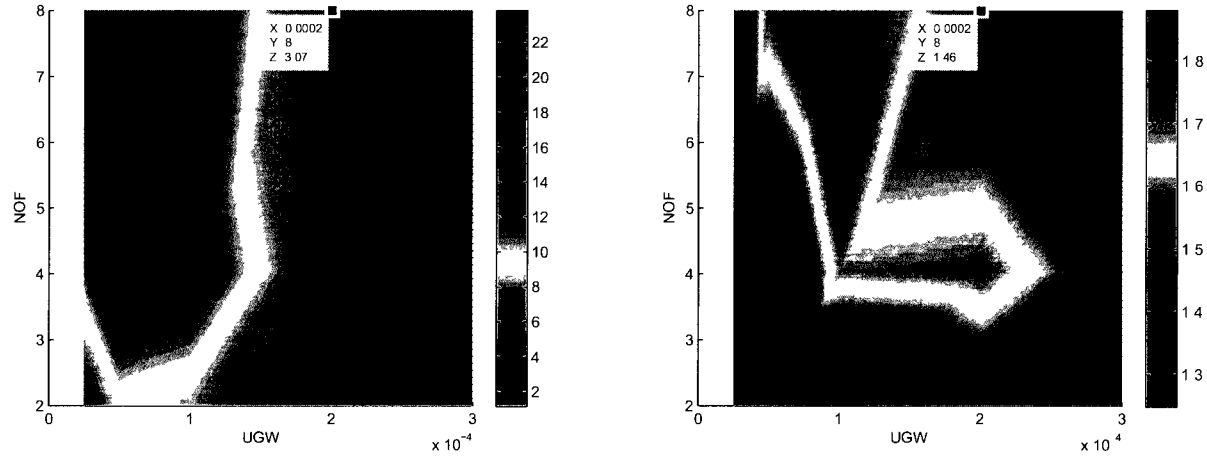


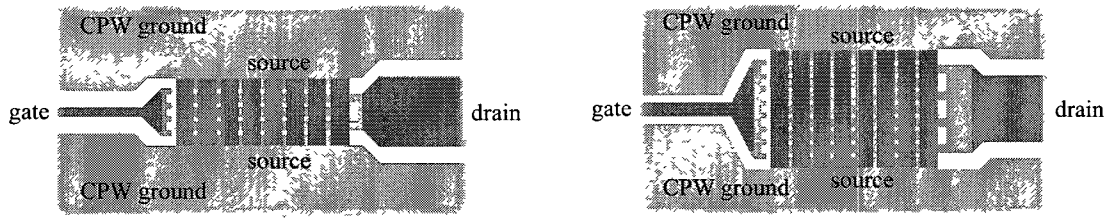
Figure 4.23: Surface plot of maximum stable gain (MSG) for various 0.5-μm GaN device sizes at 10 GHz (bias: V_{GS} = -3.5 V, V_{DS} = 20 V).



(a) input series Q

(b) r_{ds}/R_{OPT} ratio

Figure 4.24: Surface plot of (a) input series Q, and (b) r_{ds}/R_{OPT} ratio for various 0.5-μm GaN device sizes at 10 GHz.



(a) UGW = 200μm, NOF = 4. (b) UGW = 200μm, NOF = 8.

Figure 4.25: CPW feeds for power devices in 0.5-μm GaN.

4.3 De-embedding

The difficulty with microwave measurements is that there are no non-invasive methods for characterizing circuits. In order to physically measure the performance of circuits additional testing structures must to be added, such as pads or connectors. For the purpose of this work, only on-wafer measurements are performed so the test structures represent the pads where the probes would land. Since these pads would normally be unnecessary in a finished product, their effect should be removed from the measurements to obtain the true performance of the circuit. The process of negating their effects is termed *de-embedding* and can be performed with the following procedure:

1. Calibrate the performance network analyzer (PNA) up to the probe tips to remove the effects of the cables and probes (done with a *Short-Open-Load-Thru* (SOLT) calibration substrate);

2. Measure the device under test (DUT) with the help of pads;

3. Measure the pad structures without the DUT;

4. Mathematically deduce the characteristics of the pads as they would be connected to the DUT;

5. Mathematically de-embed the pads from the measurements of the DUT.

The hardest parts of this procedure are steps 3 and 4 since single-ended pad structures such as the ones shown in Figure 4.17 cannot be directly measured. Therefore, a series of other measurements are performed in order to collect enough information

to mathematically deduce the matrix defining the single-ended pads. These measurements are performed on structures that contain the actual pads arranged in specific configurations. The type of structures used depends on the de-embedding method employed. Each de-embedding method relies on different assumptions and is applicable to different situations. The block diagram in Figure 4.26 demonstrates the general scenario under which the DUT is measured. The following subsections discuss the methods employed to deduce the *ABCD* matrices of the pads for cases when the DUT represents an active or a passive component or when it represents a transmission line. Once the *ABCD* matrices of the pads are known, the true matrix of the DUT can be calculated through some basic matrix manipulations [44].

4.3.1 Active and Passive Components

The electrical properties of the pads used to probe all active and passive elements (except transmission lines) are determined through a simplified *Thru-Reflect-Line* (TRL) method. In cases where the two pad matrices from Figure 4.26 are equivalent, the original TRL method in [50] can be simplified to only a *Thru* and a *Line*. The pad matrices are assumed equivalent since 1) the reference planes of the PNA are set to the

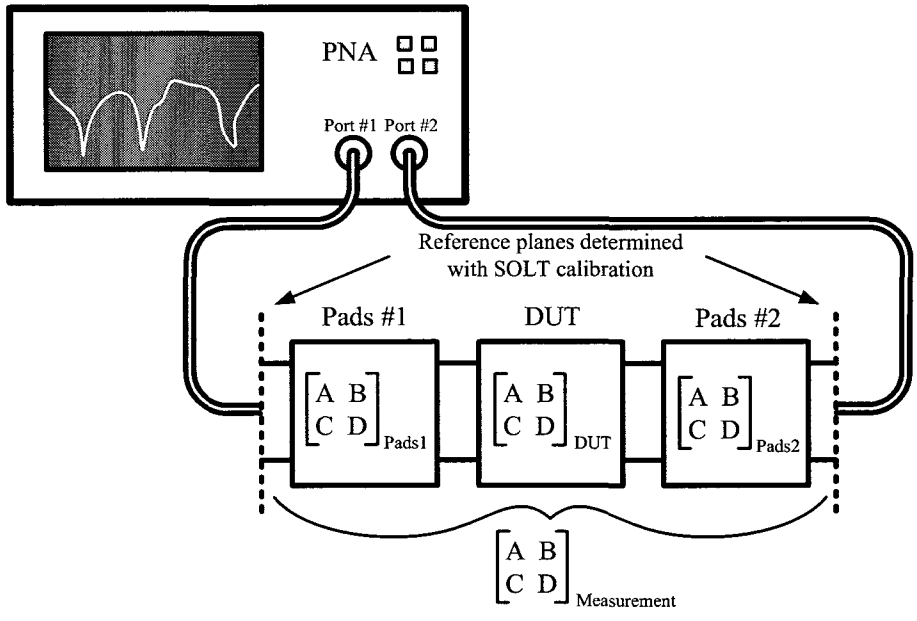


Figure 4.26: Block diagram of the measurement set-up.

tips of the probes by SOLT calibration and 2) physically identical pad structures are used on both sides of the DUT. Using identical matrices allows to reduce the original six-variable TRL problem to only three variables which makes the *Reflect* structure optional. Figure 4.27 shows the resulting *Thru* and *Line* structures. However, the assumptions under which the original TRL method was derived are still present. One of these assumptions is using a perfect 50-Ω line for the *Line* measurement. Although achieving close to 50-Ω with an EM design tool may be possible, relying on an exact impedance to de-embed transmission lines can introduce errors; therefore this method should not be used to de-embed transmission lines. After measuring the structures in Figure 4.27, the S matrix of the pads ($[P]$) can be determined with the following equations from [44]:

$$P_{22} = \frac{T_{11} - L_{11}}{T_{12} - L_{12}e^{-\gamma l}} \tag{4.10}$$

$$P_{11} = T_{11} - P_{22}T_{12} \tag{4.11}$$

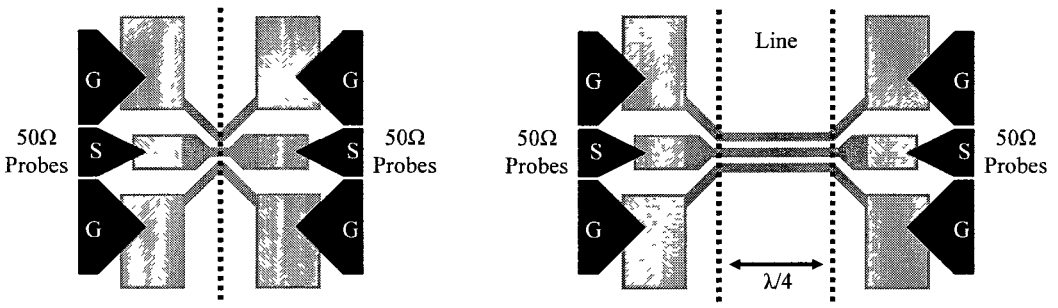
$$P_{12} = P_{21} = \sqrt{T_{12}(1 - P_{22}^2)} \tag{4.12}$$

$$e^{-\gamma l} = \frac{L_{12}^2 + T_{12}^2 - (T_{11} - L_{11})^2 \pm \sqrt{[L_{12}^2 + T_{12}^2 - (T_{11} - L_{11})^2]^2 - 4L_{12}^2T_{12}^2}}{2L_{12}T_{12}} \tag{4.13}$$

where $[T]$, $[L]$ are the measured *Thru*, *Line* S matrices, respectively.

4.3.2 Transmission Lines

From the discussion in Section 4.3.1, transmission lines should not be de-embedded with a method that relies on the line itself. For better accuracy, the method should be based on assumptions about the pads. One such method is shown in [51], however it models the pads with a lumped admittance which goes against the distributed design methodology used in Section 4.1.7. It is more appropriate to assume that the pads are reciprocal and symmetric. This is valid since 1) the pads are passive structures ($S_{21} = S_{12}$), and 2) the pads are designed to have a characteristic impedance close to 50-Ω with Section 4.1.7 ($S_{11} \approx S_{22}$). With these assumptions only two independent



(a) *Thru* structure. (b) *Line* structure.

Figure 4.27: Simplified TRL de-embedding method with (a) *Thru* and (b) *Line* structures.

variables (i.e. S_{11}, S_{21}) are needed to characterize the pads and they can be acquired with only the *Thru* measurement from Figure 4.27 (a). The $ABCD$ matrix of the pads ($[ABCD]_P$) is then found with (4.14) where $[ABCD]_T$ is the $ABCD$ matrix of the *Thru* measurements. Mathematically a square-root of a two-dimensional matrix can have many solutions, but since the $ABCD$ matrix is based on two independent variables only a single solution exists.

$$[ABCD]_P = \sqrt{[ABCD]_T} \tag{4.14}$$

4.4 Conclusions

In this chapter the design of a complete CPW-based passive and active component library was discussed. All components necessary for the implementation of CMOS and GaN Doherty amplifiers were designed and the passive components were measured. To size power devices suited for Doherty amplifiers, three performance measures which include maximum stable gain, input series Q, and the r_{ds}/R_{OPT} ratio were proposed. Finally, the de-embedding techniques used to compare the measured results to the simulations were discussed.

Chapter 5

Circuit Design and Measurements

In this chapter, the previously developed passive and active components are used to design a branch-line coupler and two Doherty implementations. The coupler is an important building block to the Doherty system and is described first in Section 5.1. Then, Sections 5.2 and 5.3 are used to describe the design of two Doherty systems which employ two separate integration approaches. The first system uses an SoC approach to integrate a 71-76 GHz Doherty PA in a 90-nm CMOS process, while the second system uses an SoP approach to integrate a 10 GHz GaN Doherty PA with an RT/duroid board.

5.1 High-Performance and Compact 70-80 GHz 90° Branch-Line Coupler in CMOS

Branch-line couplers are not only important to Doherty amplifiers but also to many RF front-end circuits that require power division or combining. In spite of that, in CMOS they are typically avoided due to their bulky size and poor performance. The bands available at millimeter-wave (mm-wave) frequencies such as the 60 GHz ISM band, the 77 GHz band, and the more recent E-bands (71-76, 81-86 GHz) have finally reached a range where the wavelengths of integrated transmission lines are reasonably small. Therefore, it now becomes practical to pursue the implementation of high-performance CMOS 90° branch-lines at these frequencies.

This section presents a new layout folding technique that reduces the size of branch-line couplers while maintaining the same performance as conventional designs. Reactively loaded and lumped element hybrids exhibit a degradation in phase

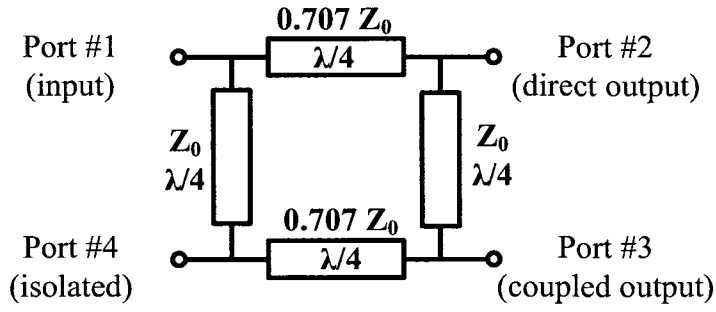


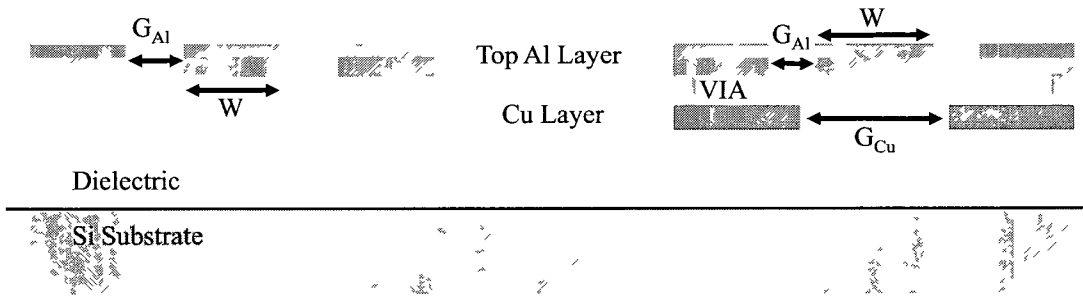
Figure 5.1: Schematic of branch-line hybrid.

imbalance [52], so unlike [53, 54] this method employs distributed-only elements to achieve maximum bandwidth.

5.1.1 Circuit Description

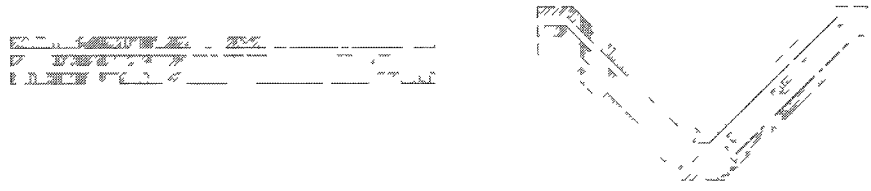
Branch-line couplers consist of four $\lambda/4$ transformers as shown in Figure 5.1 [44]. When the branch-line is designed to operate in a 50-Ω system, the characteristic impedance (Z_0) of the branch-line $\lambda/4$'s and the through-line $\lambda/4$'s are 50Ω and 35Ω, respectively.

One of the difficulties of implementing branch-line couplers in CMOS is that the design rules can be very restrictive. For example, in the standard IBM 130-nm CMOS process used for this design, they limit the lower bound of the characteristic impedance by restricting the spacing between metal traces on the same layer. Therefore, producing low impedance CPW lines, such as the 35Ω through-lines, requires capacitive loading which can adversely limit the line's usable bandwidth. In order to avoid the bandwidth reduction caused by reactive loading, Z_0 is lowered through distributed loading of the top Al metal layer with the Cu metal layer located right underneath as shown in Figure 5.2(b). This results in CPW lines with an open ground under the center conductor, as termed here "hybrid CPWG lines". To obtain the required Z_0 with this type of transmission lines, the gap (G_{Al}) between the top Al traces is set to the minimum allowable by the design rules and the gap of the lower Cu traces (G_{Cu}) is reduced until 35Ω is reached.



(a) CPW line. (b) Hybrid CPWG line.

Figure 5.2: Cross section of (a) 50Ω CPW line and (b) 35Ω hybrid CPWG line.



(a) Standard CPW. (b) Folded CPW.

Figure 5.3: Layout of (a) standard CPW and (b) folded CPW.

5.1.2 Size Reduction

As mentioned, branch-line couplers have a large footprint, so to improve the feasibility of the design, the size of the $\lambda/4$ lines is reduced through folding. This same technique is applied to both CPW 50Ω lines and hybrid CPWG 35Ω lines and can be seen in Figure 5.3. The impact of the folds is verified through EM simulations with *Ansoft HFSS* [48] and Table 5.1 summarizes the results. The folded versions of the CPW and hybrid CPWG lines are designed to have identical width (W) and gap (G_{Al}, G_{Cu}) parameters, as well as the same effective lengths as the standard lines. From the summary of results, it can be noticed the folded lines result in equivalent performance as standard lines. This equivalent performance is attributed to using 45° bends as well as the use of small gaps between the traces. It is found that when the lines are sized such that $G_{Al} < W$, the discontinuity formed by 45° bends have practically no impact on the electrical performance.

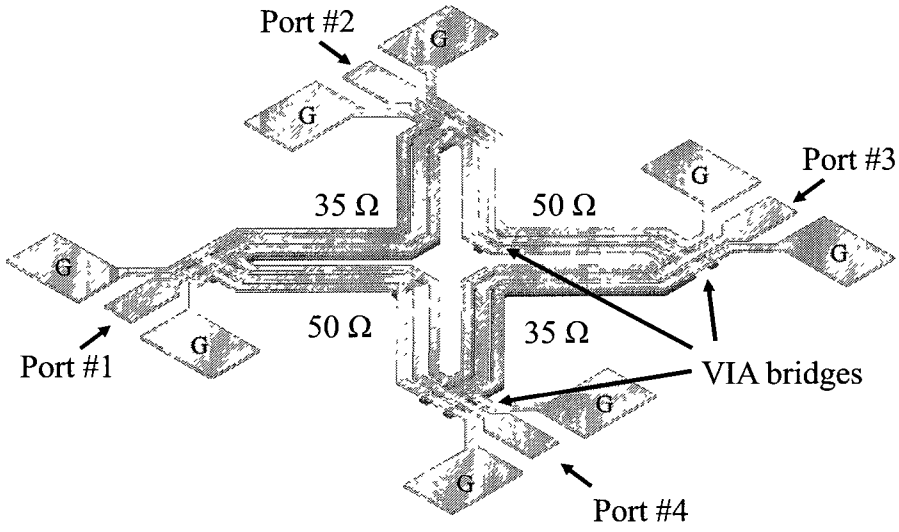


Figure 5.4: Complete branch-line coupler.

By folding each branch of the coupler and arranging them to meet in the middle as in Figure 5.4, the coupler makes use of the typically wasted center area. The total area is reduced by 35% compared to an non-folded conventional design. The resulting branches appear close to one another which can result in unwanted coupling. But, by employing CPW lines with ground traces on the same layer as the central conductor, the isolation at 80 GHz between two adjacent $\lambda/4$ CPW's separated by 25 μm is simulated with *HFSS* to be better than 50 dB. To suppress the slot-line mode of propagation present when the grounds of a CPW line have opposite potentials, ground equalizing VIA bridges are used on all discontinuities (e.g. corners, central folds).

Table 5.1: EM simulated results evaluating the effects of folding.

Parameter	Standard CPW	Folded CPW	Standard hybrid CPWG	Folded hybrid CPWG
W/G_{Al}, G_{Cu} (μm)	12/7.5, N/A	12/7.5, N/A	15/6, 19	15/6, 19
Z_0 (Ω)	46.6	46.2	32.8	33.1
θ @ 75 GHz (°/mm)	190.1	189.1	193.4	192.2
α @ 75 GHz (dB/mm)	1.1	1.1	0.69	0.69

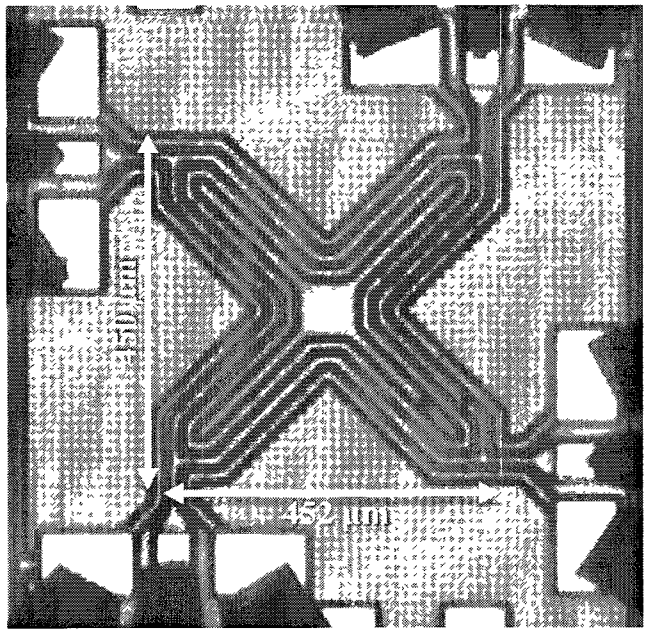


Figure 5.5: Photomicrograph of fabricated 70-80 GHz 130-nm CMOS coupler [2].

5.1.3 Simulation and Measurement Results

Before submitting the design for fabrication, the *HFSS* simulation results of the complete coupler are validated with a second EM simulator (*Agilent Momentum* [46]). This reduces the probability of design tool errors. The photomicrograph of the fabricated branch-line coupler is shown in Figure 5.5.

Figure 5.6 compares the measured S-parameters to the *HFSS* simulated ones. The pads are de-embedded with custom on-chip TRL structures along with the procedure described in Section 4.3.1. The measured insertion loss from 70-80 GHz on the direct port and coupled port are in the range of 1.1-1.4 dB and 1.0-1.1 dB, respectively. The maximum measured isolation is 42.8 dB at 78 GHz and is better than 14.8 dB across the whole bandwidth. Figure 5.7 compares the measured return loss for the case when pad de-embedding is used and when it is not used. There is a minimal frequency shift between the two cases proving that the non-loading pads designed in Section 4.1.7 are indeed 50-Ω. The measured return loss across the whole bandwidth is better than 19.5 dB on the input port and better than 17.5 dB on all ports. Figures 5.8 (a) and (b) compare the measured phase and amplitude imbalances with EM simulations. From 70-80 GHz the phase imbalance ranges between -2° and 0.2°, while the magnitude

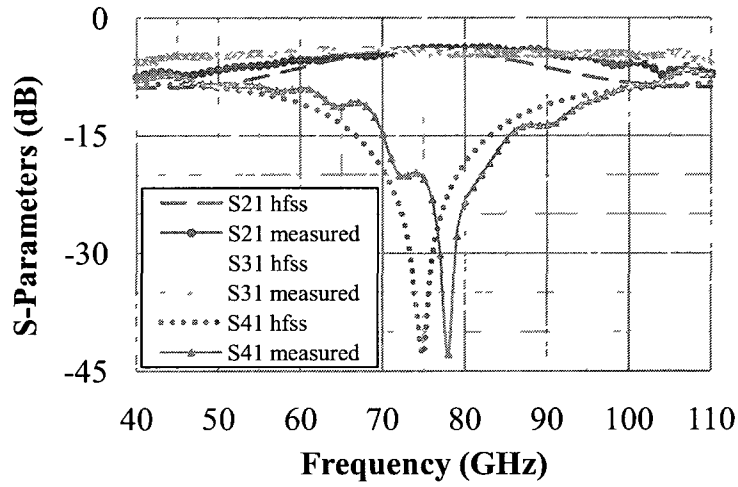


Figure 5.6: Measured and EM simulated S-parameters of 130-nm CMOS coupler [2].

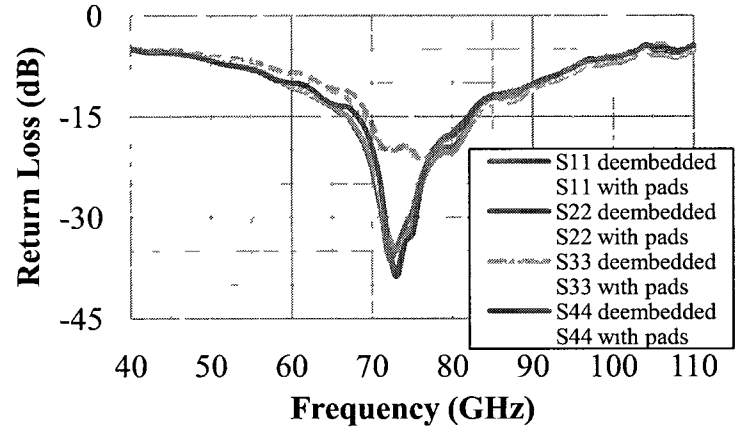
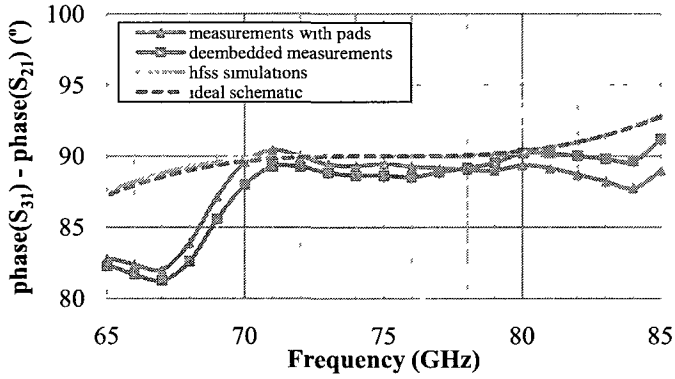
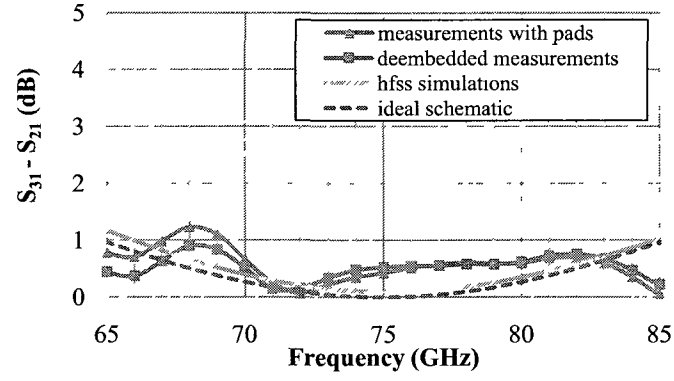


Figure 5.7: Measured and de-embedded return loss of 130-nm CMOS coupler [2].

imbalance between 0.1 and 0.6 dB. All the measurements are in very good agreement with the EM simulations. Table 5.2 shows a performance comparison with previously reported CMOS work at mm-wave frequencies. The fabricated coupler exhibits better insertion and return loss, as well as lower imbalances between outputs.



(a) Phase imbalance.



(b) Amplitude imbalance.

Figure 5.8: Measured and EM simulated (a) phase and (b) amplitude imbalance of 130-nm CMOS coupler.

Table 5.2: Measured performance compared to previous work.

Ref.	CMOS Tech.	Freq. (GHz)	S_{21} / S_{31} (dB)	R.L. (dB)	Mag. Imb. (dB)	Phase Imb. (°)	Area (μm^2)
[53]	90-nm	60	-4.8 / -5.5	> 17	0.7	-2	355x287
[54]	90-nm	57-64	-5.1 / ?	> 18	-0.8 to 0.5	0 to 5	194x414
[55]	180-nm	**31.1-35.6**	≈-4.4 to -4.8 / ≈-5.1 to -5.5	> ≈14	≈0.3 to 1.0	±10	**300x360**
This work	130-nm	70-80	**-4.1 to -4.4 / -4.0 to -4.1**	**> 19.5**	**0.1 to 0.6**	**-2 to 0.2**	450x452

5.1.4 Measurement Set-up

Measurements are performed using a 110 GHz performance network analyzer (Agilent N5250A) and an on-wafer programmable probe station (Karl Suss PA200) with 150-μm pitch Ground-Signal-Ground (G-S-G) probes. The network analyzer has four ports but they are only available up to 67 GHz. For frequencies above 67 GHz only two ports measurements can be performed. Therefore to characterize the branch-line coupler fully, a series of two port measurements are performed while terminating the remaining two ports of the coupler with 50-Ω resistive loads. The full S-parameter matrix of the coupler in (5.1) is then combined during post-processing from the available measurements. The list of two port measurements is summarized in Table 5.3. It should be noted that the 50-Ω terminations were not de-embedded from the measurements. This step was omitted because the return loss in the band of interest of the two terminations is measured to be better than -15 dB and -17 dB, respectively.

$$\begin{bmatrix} S_{11} & S_{21} & S_{31} & S_{41} \\ S_{12} & S_{22} & S_{32} & S_{42} \\ S_{13} & S_{23} & S_{33} & S_{43} \\ S_{14} & S_{24} & S_{34} & S_{44} \end{bmatrix} \tag{5.1}$$

Table 5.3: Four port S-parameter measurement set-up with two port PNA.

PNA p1	PNA p2	Probe orientation	S-parameters
Coupler p1	Coupler p2	90°	S_{11}, S_{12}, S_{21}, S_{22}
Coupler p2	Coupler p3	90°	S_{23}, S_{32}, S_{33}
Coupler p3	Coupler p4	90°	S_{34}, S_{43}, S_{44}
Coupler p4	Coupler p1	90°	S_{41}, S_{14}
Coupler p1	Coupler p3	180°	S_{13}, S_{31}
Coupler p2	Coupler p4	180°	S_{24}, S_{42}

5.1.5 Summary

A new performance-conserving layout reduction technique for 90° branch-line couplers has been proposed and confirmed through the measurements of a 70-80 GHz coupler in 130-nm CMOS. The layout area reduction is 35% while exhibiting best reported performance. The design is based on CPW and hybrid CPWG transmission lines. Experimental measurements of the 0.203 mm^2 coupler from 70-80 GHz show a maximum insertion loss of 1.4 dB, an amplitude imbalance less than 0.6 dB, a phase imbalance less than 2°, and a return loss better than 19.5 dB. With these measured results, the proposed coupler is well suited for the design of high-performance front-end SoC circuits.

5.2 E-band Reconfigurable Doherty Amplifier SoC in CMOS

The previous section confirmed that high-performance CMOS branch-line couplers are feasible at 70-80 GHz. Now this section makes use of a similar coupler to present the implementation of a complete E-band (71-76 GHz) reconfigurable Doherty SoC with a commercial 90-nm process.

5.2.1 Circuit Description

The design equations (3.1) and (3.2) from Section 3.1 are used to combine two symmetric single-ended PAs into the Doherty circuit shown in Figure 5.9. The electrical lengths of the transmission lines in the schematic are all at mid-band (i.e. 73.5 GHz).

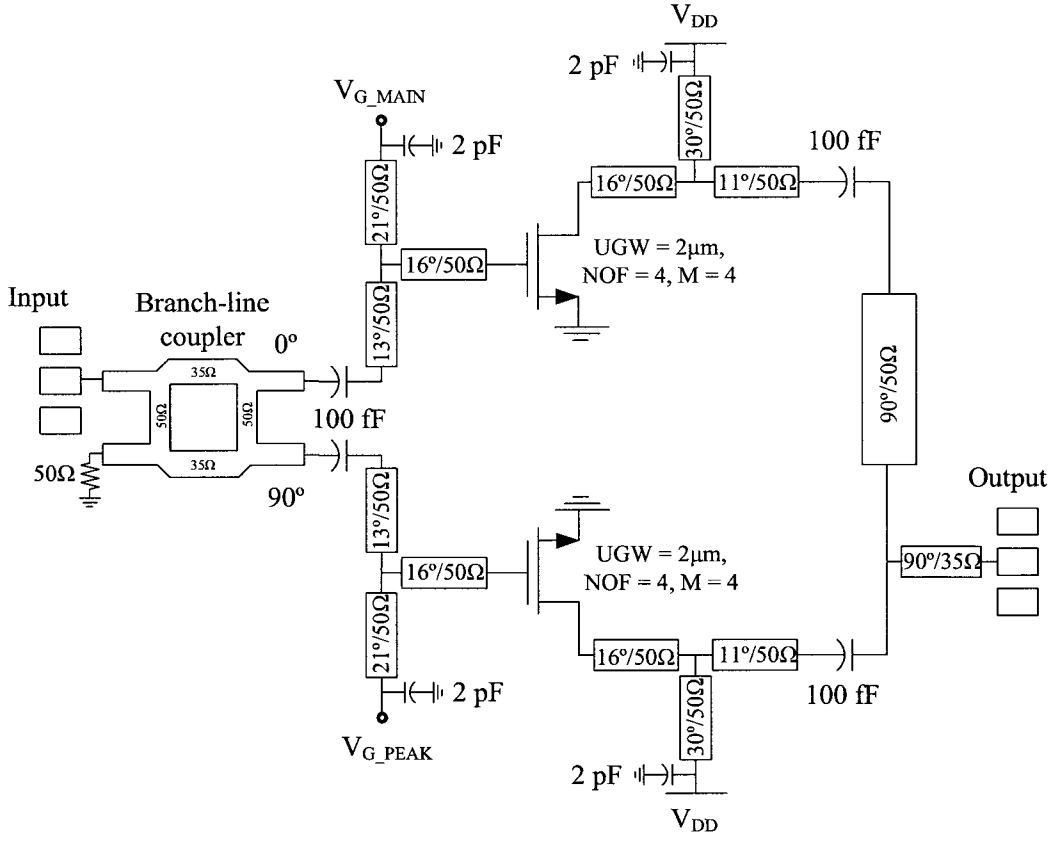


Figure 5.9: Schematic of E-band Doherty power amplifier.

The power device used in the single-ended PAs is selected based on maximum stable gain (MSG), on the r_{ds}/R_{OPT} ratio, and on the input series Q, as it was discussed in Section 4 2 1 A load-pull analysis of the power device is performed using the *Agilent ADS* environment to find the optimum input and output matching impedances All matching components are designed using 50-Ω lines since Table 4 1 showed that they have low loss and convenient physical dimensions The input matching consists of a series CPW line with a shunt inductive CPW line that conjugately match the 50-Ω source impedance to the input impedance of the power device, the output network uses a series CPW line with a shunt inductive CPW line to transform the 50-Ω load impedance into the optimum output load The shunt lines are also used to provide the drain and gate biases Combining the bias network with the matching network simplifies the overall complexity of the circuit and results in lower losses Since the complete single-ended PA is designed for a 50-Ω system, the resulting branch-line coupler and output combiner also require 50-Ω impedances The branch-line coupler is designed similarly to the one in Section 5 1 However, due to the relaxed design rules of the TSMC process, the 35Ω through branches of the coupler are implemented with standard CPWs instead of hybrid CPWGs

The main PA stage is biased in class-AB ($V_{G\,MAIN} = 0\ 55$V) similarly to the work in [15] This is done to improve the power gain at mm-wave frequencies On the other hand, the peak amplifier is biased in shallow class-C ($V_{G\,PEAK} = 0\ 3$V) such that it begins to modulate the load at the point where the main amplifier starts to saturate By biasing the peak stage slightly below class-B, its peak current is almost the same as the main stage which eliminates the need to have a larger periphery peak device

5.2.2 Reconfigurable Option

From the schematic in Figure 5 9 it can be noticed that the peak and the main stages of the Doherty are physically symmetric and their output signals combine in phase This means that the Doherty can be re-biased such that both stages are operating in the same class which would result in a balanced PA The advantage of having this option is that balanced PAs with same device peripheries have an improved gain and linearity On the other hand, the drawback is that the load-modulation effect no longer occurs so the back-off efficiency is worse The applications for such an option are cases where low PAPR modulation schemes are used One example can be during initial receiver/transmitter pairing where the extra gain and simpler

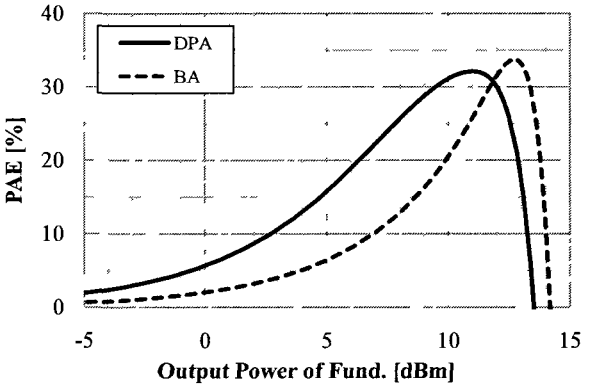


Figure 5.10: Power added efficiency (PAE) of E-band DPA when biased as a Doherty and as a balanced amplifier (BA), simulated at 73.5 GHz with ideal schematic.

modulation scheme can be used to acquire range for easier device discovery.

Since this work is only based on the PA component of the transceiver, the reconfigurable option is only used as a method to verify the effectiveness of the Doherty. In Figures 5.10 and 5.11 the ideal schematic from Figure 5.9 is used to plot the PAE and the gain as a function of output power for both configurations. In the Doherty configuration the circuit produces 10.5 dBm of output power at the 1-dB compression point (P_{1dB}) with 31.8% PAE, while in the balanced configuration the circuit produces 11.6 dBm at P_{1dB} with 29.7% PAE. Both peak PAEs are close, but the advantage of using the Doherty occurs at the 6-dB back-off from P_{1dB} where the PAE is improved by 7.2%. The improvement in gain of the balanced configuration is 2.5 dB. Figure 5.12 shows the current consumption as a function of output power for the two configurations when they are supplied from 1.5V. Finally, all the results from the ideal schematic are summarized in Table 5.4.

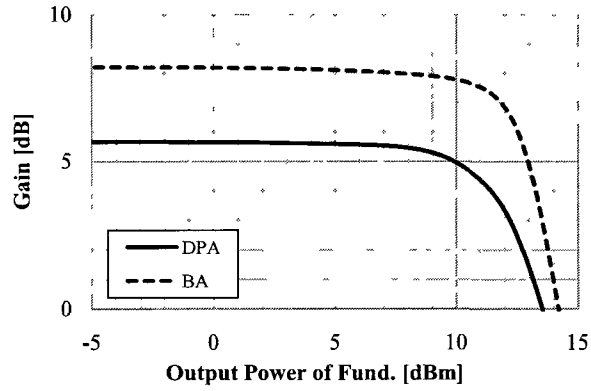


Figure 5.11: Power gain of E-band DPA when biased as a Doherty and as a balanced amplifier (BA); simulated at 73.5 GHz with ideal schematic.

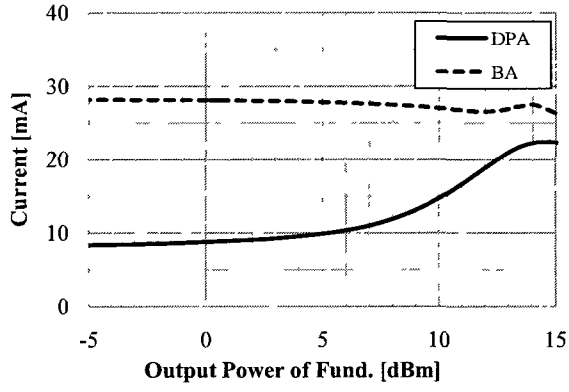


Figure 5.12: Current consumption of E-band DPA when biased as a Doherty, and as a balanced amplifier (BA); simulated at 73.5 GHz with ideal schematic.

Table 5.4: Summary of simulated results at 73.5 GHz with ideal schematic.

	Doherty PA	Balanced PA
Main bias	class-AB	class-A
Peak bias	class-C	class-A
Power dissipation	8.2 mA @ 1.5V	28.2 mA @ 1.5V
Gain	5.7 dB	8.2 dB
P_{1dB}	10.5 dBm	11.6 dBm
PAE @ P_{1dB}	31.8%	29.4%
PAE @ 6-dB back-off	14.6%	7.4%

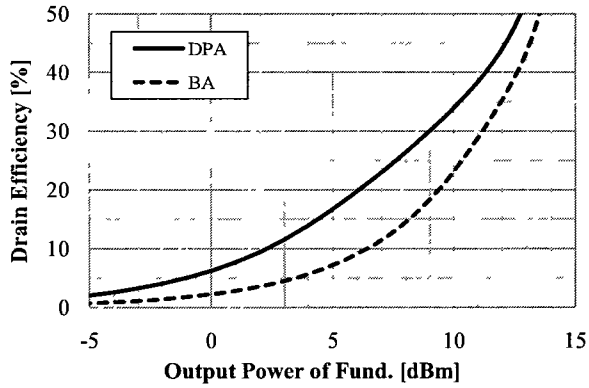


Figure 5.13: Drain efficiency of E-band DPA system when biased as a Doherty and as a balanced amplifier (BA); simulated at 73.5 GHz with schematic that includes CPW losses.

5.2.3 Schematic Simulations Including CPW Losses

Since the 71-76 GHz Doherty design heavily relies on transmission lines, an accurate estimate of the final results can be obtained by including the CPW attenuation losses. The CPW losses are included in the Doherty schematic from Figure 5.9 by using the characterization data in Table 4.1. An attenuation of 0.0054 dB/° and 0.0041 dB/° is added to the 50Ω and 35Ω CPW lines, respectively.

After including the losses, the gain reduces by 2.6 dB for the Doherty configuration and by 2.5 dB for the balanced configuration. It was realized prior to completing the layout that the gain is lower than a practical PA. However, the design work continued, since switching to a faster CMOS process was not possible. Therefore the Doherty was designed with the intent to verify the operation through the drain efficiency which is independent of gain. The simulated drain efficiency characteristics that include CPW losses are shown in Figure 5.13. It can be noticed that the Doherty exhibits a back-off efficiency improvement of 5.2% over the balanced configuration. Table 5.5 shows the complete summary of simulated results that include CPW losses. The S-parameter simulations for the balanced configuration are shown in Figure 5.14. Both configurations are unconditionally stable since their K stability factors are greater than 1 (as shown in Figure 5.15).

Table 5.5: Summary of simulated results at 73.5 GHz with schematic that includes CPW losses.

	Doherty PA	Balanced PA
Main bias	class-AB	class-A
Peak bias	class-C	class-A
Gain	3.1 dB	5.7 dB
P_{1dB}	8.9 dBm	10.3 dBm
η @ P_{1dB}	29.7%	24.7%
η @ 6-dB back-off	11.4%	6.2%

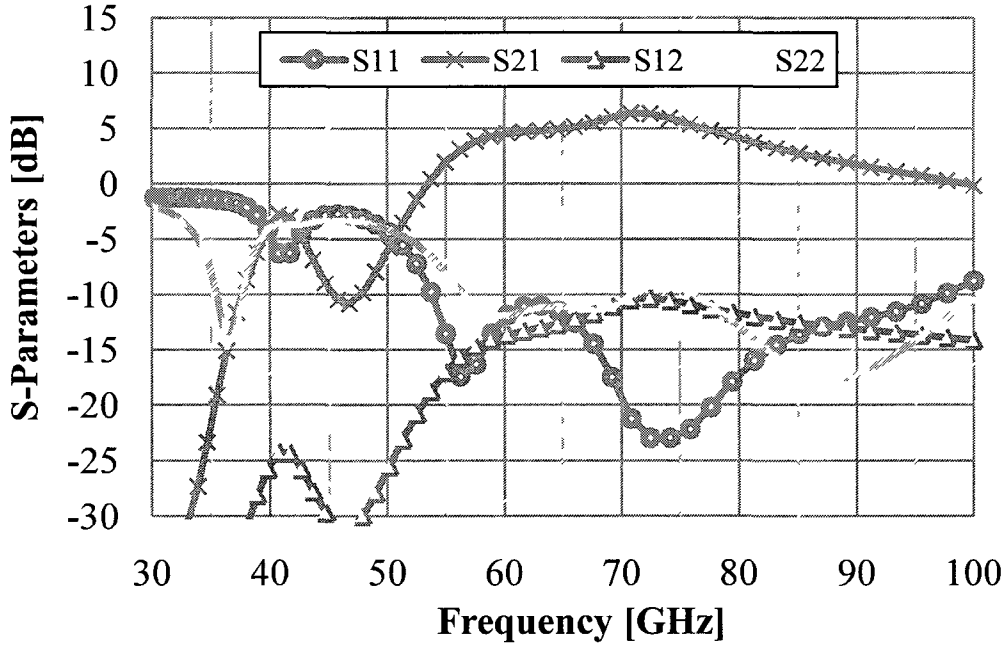
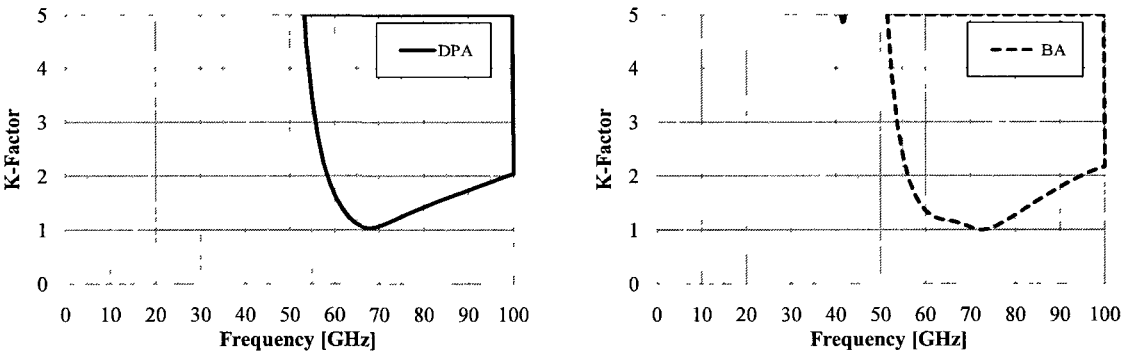


Figure 5.14: S-parameters of E-band DPA system; simulated with schematic that includes CPW losses.



(a) Doherty PA. (b) Balanced PA.

Figure 5.15: Stability of E-band DPA system when biased as a (a) Doherty power amplifier (DPA) and as a (b) balanced amplifier (BA); simulated with schematic that includes CPW losses.

5.2.4 Complete Circuit and Post-Layout Verification Issue

The custom 90-nm CMOS components designed in Chapter 4 are put together to form the complete Doherty PA layout shown in Figure 5.16 and the resulting layout has a total footprint of 1590 μm by 965 μm. The large pads on the *North* and *South* sides of the circuit are for DC bias, whereas the pads on the *West* and *East* sides of the circuit are for the RF input and output, respectively.

Before the design was submitted for fabrication, *Agilent Momentum* [46] was used for the post-layout EM verification. However, after the circuit was already sent for fabrication it was realized that the *Momentum* layer definition was wrong. The mistake occurred because *Momentum* expands the thickness of the dielectric material in cases where the metallization layer is defined as a *thick* layer. This was not expected since it was assumed that once a dielectric layer is defined its dimension would stay fixed.

As a consequence, there was an under estimation of the capacitive parasitics of the VIA bridges that are used through out the circuit to equalize the CPW grounds. Without disclosing information about the process layers, the total under estimation per VIA bridge is about 5 fF. When 5 fF is added on each side of a quarter-wave 50Ω line at 73.5 GHz, the frequency shift is calculated with the size reduction equation from [52] to be about -5 GHz. Thus, a frequency shift of the same order of magnitude can be expected from the measurements.

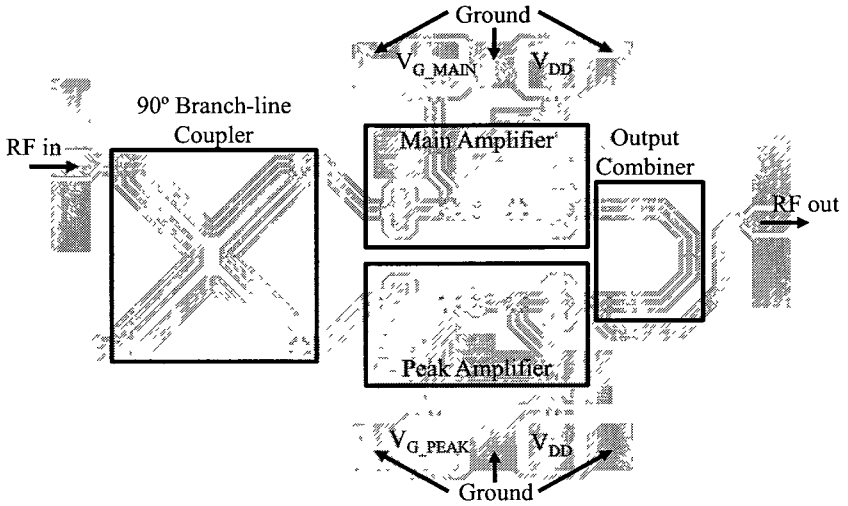


Figure 5.16: Layout of E-band Doherty power amplifier.

5.2.5 EM Simulation and Measurement Results

The photomicrograph of the fabricated Doherty SoC is shown in Figure 5.17. As it was discussed in Section 5.2.4, the measurements expect a frequency shift due to the unaccounted VIA bridge capacitances. Figure 5.18 compares the measured S-parameters to the EM simulated S-parameters with a corrected *HFSS* model. A frequency reduction of about 8 GHz can be noticed in both the measurements and the EM simulations. Overall, the measurements agree very well with the EM simulations which validates the EM modeling technique. The only existing disagreement is for S_{22} in the 45-67 GHz frequency range. This inconsistency is caused by a drift in the calibration of the network analyzer and is discussed in the following section.

When biased as a balanced amplifier with 1.5V supplies and 0.75V at the gates, the measured DC power consumption of the circuit is 24.4 mA while the simulated is 28.2 mA. This verifies that the circuit is properly implemented from a DC point of view.

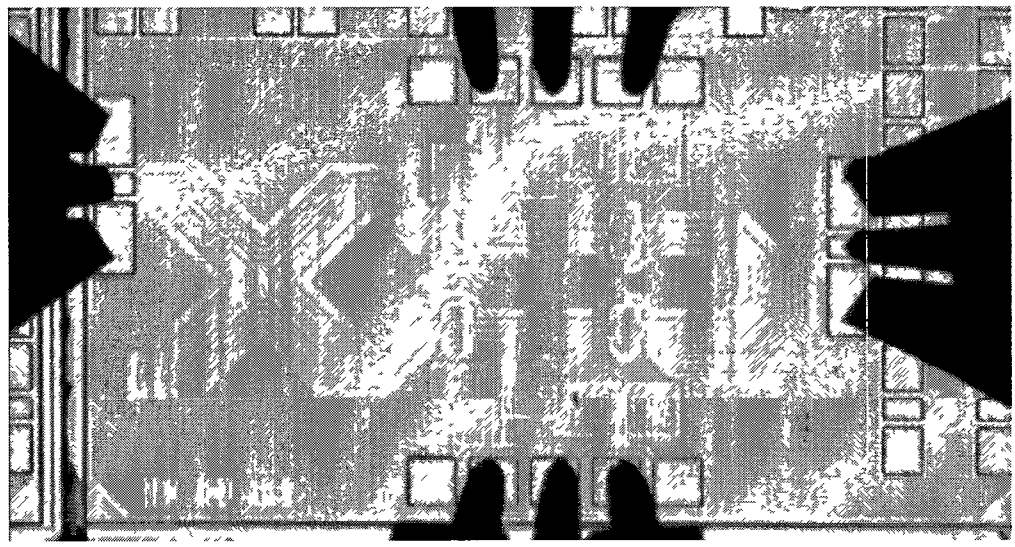


Figure 5.17: Photomicrograph of fabricated E-band Doherty power amplifier SoC in 90-nm CMOS.

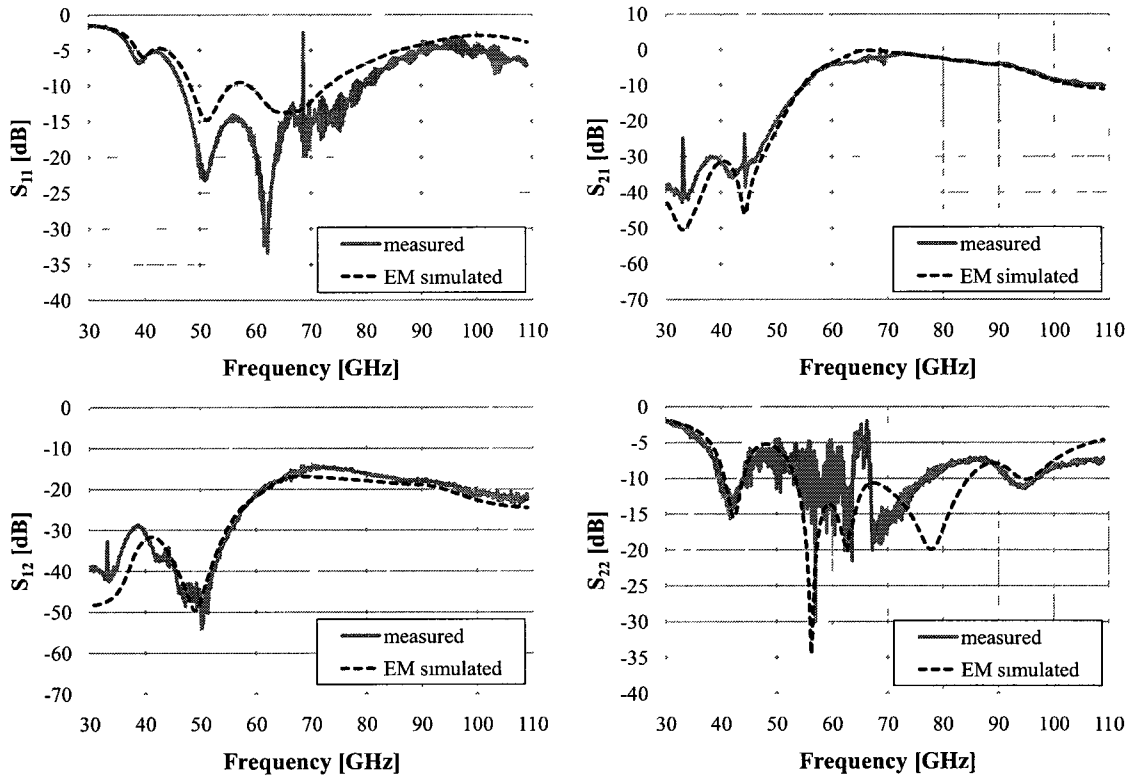


Figure 5.18: Measured and EM simulated S-parameters of E-band Doherty PA under balanced biasing conditions.

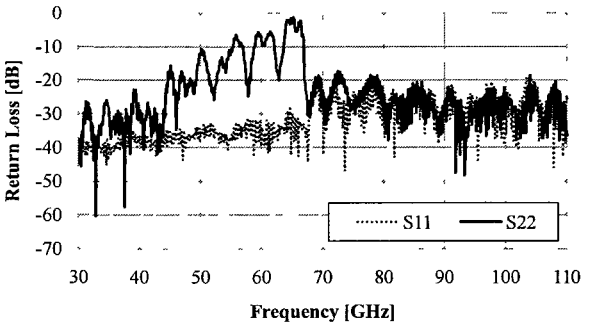


Figure 5.19: Measured return loss of 50-Ω calibration standard showing an issue with port#2 of the network analyzer.

5.2.6 Measurement Set-up

Measurements are performed using the same 110 GHz performance network analyzer and probe station as described for the branch-line coupler in Section 5.1.4. The RF performance is measured using the *West* and *East* ports of the probe station by connecting them to the input and output of the Doherty with G-S-G 150-μm pitch probes. The *North* and *South* ports of the probe station supply the main and peak amplifier stages through Power-Ground-Power (P-G-P) 150-μm pitch probes. A four port Agilent DC power analyzer (N6705A) is used to supply and measure the DC power.

As noticed from the Doherty measurements of S_{22}, there is a calibration issue between 45-67 GHz on port#2 of the network analyzer. This problem can be seen more clearly when measuring the 50-Ω calibration standard in Figure 5.19. Even though a SOLT calibration was performed to de-embed the probes all the way to the tips, the issue could not be fully eliminated since it drifted with time. Therefore the discrepancy in S_{22} at 45-67 GHz and the small dip in S_{21} at 60-67 GHz can be attributed to this problem.

5.2.7 Summary

A 71-76 GHz reconfigurable Doherty PA is proposed and implemented in a standard 90-nm CMOS process. The footprint of the Doherty PA is 1.53 mm^2 and the simulation results showed a P_{1dB} of 8.9 dBm, a peak drain efficiency of 29.7%, and a 6-dB back-off drain efficiency of 11.4%. Due to underestimation of VIA bridge parasitics that caused a frequency shift in the fabricated prototype, the measurements did not replicate the schematic results. The schematic simulations did however demonstrate a 5.2% back-off efficiency improvement compared to the case where the circuit is biased as a balanced amplifier. This shows that there is an advantage to using the Doherty technique, even at mm-wave frequencies. Finally, after correcting the VIA bridge issue, the shifted measurements agreed very well with the EM simulations which puts confidence in the EM modeling technique.

5.3 10 GHz Doherty Amplifier SoP

This section presents a 10 GHz GaN Doherty PA integrated with a system-on-package (SoP) approach on a RT/duroid board. The SoP integration technique is applied in an attempt to tackle the limitations of implementing single-chip high-power RF PAs; the limitations include low on-chip Q's and current-handling capabilities [56], as well as high cost of GaN real estate.

5.3.1 Circuit Description

Similarly to the CMOS SoC implementation, the input power splitter of this Doherty is implemented with a branch-line coupler. But instead of implementing it on-chip, it is designed as part of the package due to its bulky size at 10 GHz. The gate bias networks are also placed on the package to save area, whereas the drain biasing and output combiner are placed on the package to circumvent the current-handling limitations of the GaN metallization layers. The resulting Doherty PA system is shown in Figure 5.20.

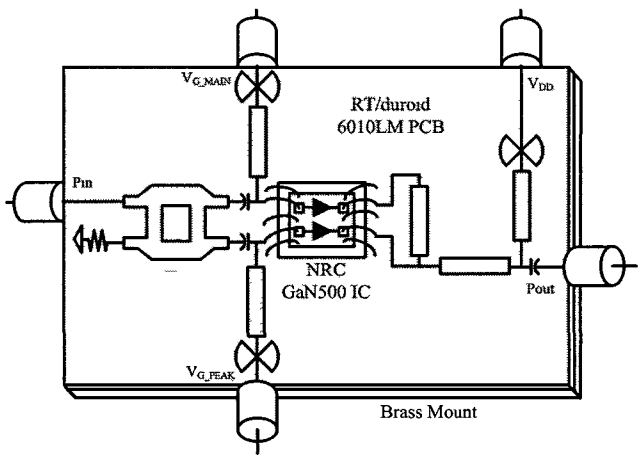


Figure 5.20: 10 GHz Doherty PA system-on-package (SoP).

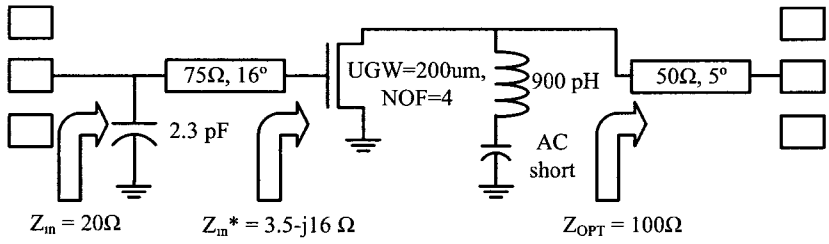
A GaN MMIC is used to implement the two single-ended PA stages. Some passive components are included on the MMIC to match the GaN devices and to absorb the bond wire discontinuities. As it was shown in Section 4.2.2, the active devices in the 0.5-μm GaN process have very low output impedances. Therefore, the recommendations from Section 3.4.1 are used to improve the Doherty load modulation effect by

increasing the periphery of the peak stage. As a result, a peak device with twice the main device periphery is used.

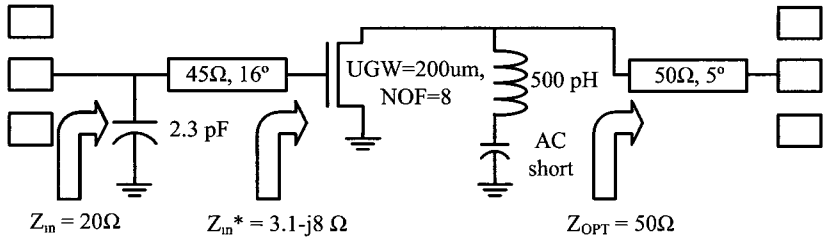
5.3.2 Single-ended GaN PA Design

The GaN devices used to design the single-ended PAs are selected based on the criteria discussed in Section 4.2.2. It was found that the devices with unit gate widths (UGWs) of 200 μm provide optimum trade-off between efficiency, gain, and input matching network. Therefore, the main device is sized with a UGW of 200 μm and 4 fingers, while the peak device is sized with a UGW of 200 μm and 8 fingers.

The schematics for the two stages are shown in Figure 5.21. The Doherty topology relies on having equal delays for both PA stages so the matching networks are made symmetric. A shunt inductor is used at the output of each stage to resonate the capacitive parasitics. At low power this inductor could also be used for biasing, but due to the low current-handling of the metal layers it is only used for matching and the DC flow is blocked by placing AC ground. With the inductors, the optimum load for the peak and main stage are 50Ω and 100Ω, respectively. The input matching is implemented with a series CPW line and a shunt capacitor. To maintain the required electrical and physical symmetry between stages, the same length CPW lines are used. The peak stage is matched with a 75Ω line and a 2.3 pF capacitor, while the main stage is matched with a 45Ω line and a 2.3 pF capacitor. The matching networks bring up the impedance of the two stages to 20Ω. Increasing the intrinsic input impedance of the GaN device makes the transition off-chip much less abrupt. The resulting layout for the two single-ended PA stages is shown in Figure 5.22 and the total area consumption of the GaN MMIC is 1050 μm by 1120 μm.



(a) Schematic of main stage.



(b) Schematic of peak stage.

Figure 5.21: Single-ended GaN PA schematic for (a) the main stage and for (b) the peak stage.

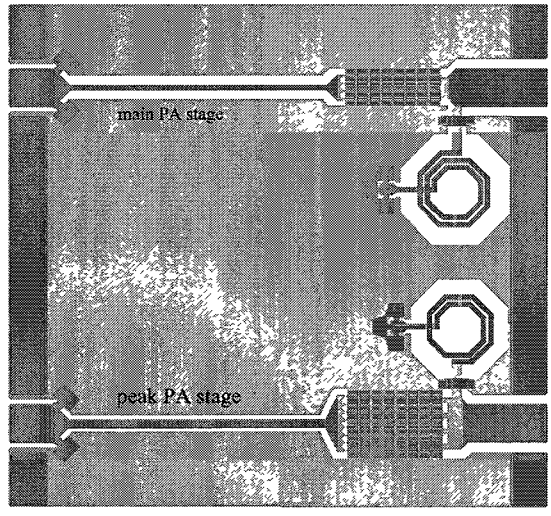


Figure 5.22: Single-ended PAs implemented on 0.5-μm GaN MMIC.

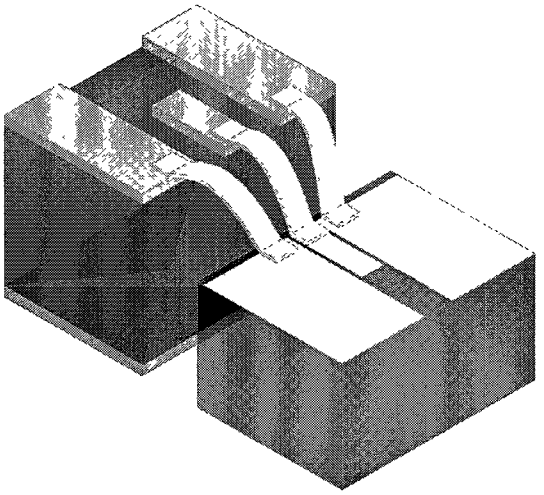


Figure 5.23: Coplanar ribbon bond model in *HFSS*.

5.3.3 System-on-Package Integration

The GaN MMIC is integrated on an RT/duroid package with coplanar ribbon bonds instead of traditional wire bonds to minimize the discontinuity of the transition. The available ribbon bonds are 3 mil wide and 0.5 mil thick. Naturally, the relative permittivity (ε_r) of air is lower than the ε_r of GaN and RT/duroid materials. Thus, the coplanar ribbon bonds will exhibit characteristic impedances that are higher and cause discontinuities. To minimize this effect, the characteristic impedance is reduced by placing the outer ground ribbons as close as possible to central signal ribbon. By following the tolerances, the pitch of the ribbon bonds on the MMIC side are set to 4.5 mil and the pitch on the package side are set to 8.5 mil.

Next, the ribbon bonds are electromagnetically simulated with *HFSS* such that they can be included into the design. The *HFSS* model is shown in Figure 5.23. From the simulation results in Figure 5.24 it can be noticed that the ribbon bonds behave approximately as 75Ω transmission lines. Therefore, as in Figure 5.25, the 100-μm long ribbon bonds are treated in the design as transmission lines with electrical lengths of 12° at 10 GHz and characteristic impedances of 75Ω.

Figure 5.26 shows the schematic for the input network that includes the ribbon bonds. The 20Ω input impedance of the single-ended GaN stages is brought to 50Ω using the ribbon bond together with a series CPW line and a parallel open-ended

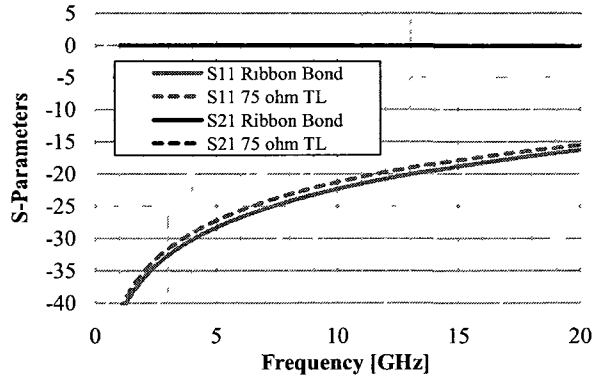


Figure 5.24: EM simulated S-parameters for 100-μm coplanar ribbon bond compared to ideal 75Ω transmission line.

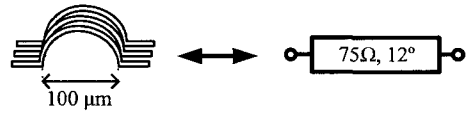


Figure 5.25: Simplified circuit component for a 100-μm coplanar ribbon bond.

stub. The ribbon bond is absorbed into the matching network by using it as a series 12° 75Ω line. Finally, the previously mentioned branch-line coupler is placed as shown in Figure 5.26 to drive the two 50Ω stages

The schematic for the output power combining network is shown in Figure 5.27. In order for the Doherty circuit to result in proper load modulation, the signals at the drains of the two stages need to be combined through an exact 90° output combiner. The difficulty is that 34° of the output combiner is already accounted for by the ribbon bonds and the small on-chip segments following the drains. Furthermore, the main and the peak stages require different optimum loads (i.e. $R_{OPT,MAIN} = 100\Omega$, $R_{OPT,PEAK} = 50\Omega$). To simultaneously solve both of these issues, an additional 163° impedance transforming segment is added to the main and peak stage such that the output is combined through a 360° + 90° output combiner. This technique makes use of the output ribbon bonds as part of the output combiner.

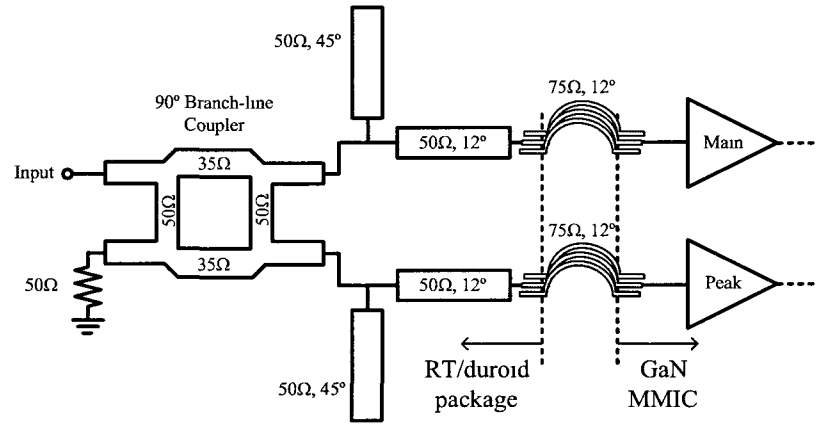


Figure 5.26: Input schematic for 10 GHz SoP Doherty.

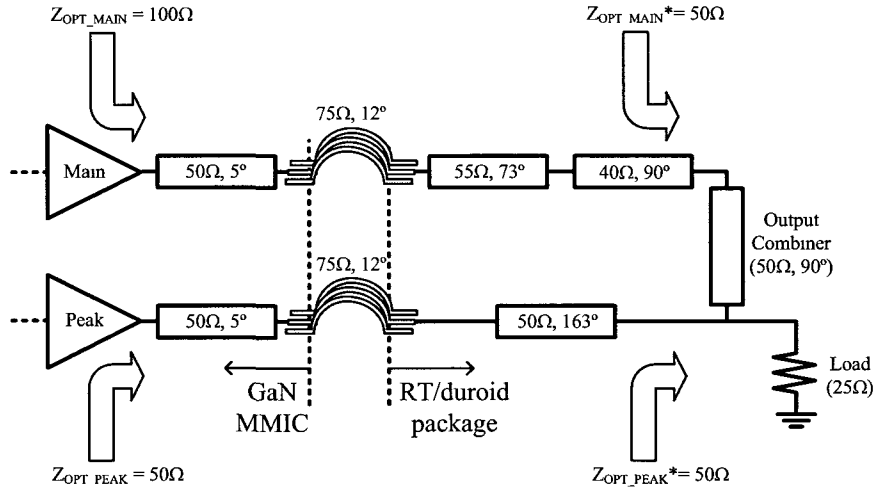


Figure 5.27: Output schematic for 10 GHz SoP Doherty.

5.3.4 Packaging Material and Surface Mounted Components

The RT/duroid material used to implement the input and output networks is selected to have an ε_r similar to the one on the GaN MMIC to ensure that the dimensions of components are on the same order of magnitude. For example, if the ε_r of the package is much lower, the resulting transmission lines would be physically large and interfacing to the tiny MMIC would be impossible. Therefore, the 6010LM RT/duroid material is selected since it has an ε_r of 10.2. A material thickness of 10 mil is used since it is comparable to the MMIC.

The biasing networks were not included in Figures 5.26 and 5.27 because they are almost "invisible" from an RF point of view. The bias voltages are supplied to

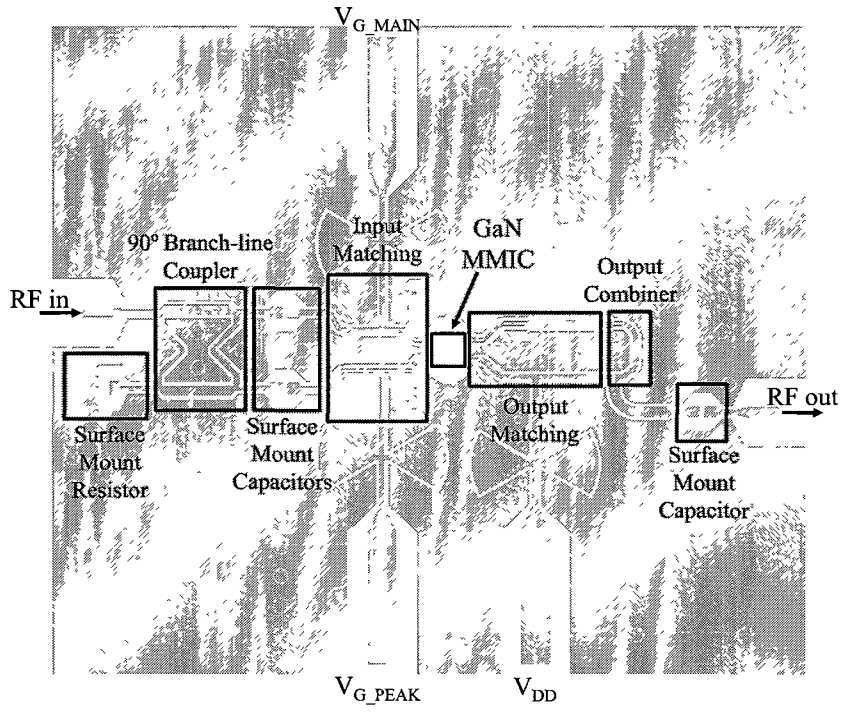


Figure 5.28: Layout of 6010LM RT/duroid package for 10 GHz Doherty.

the circuit through the radial stub network discussed in Section 4.1.5. For the input network, surface mounted AC coupling capacitors are placed right after the branch-line coupler to eliminate DC power dissipation in the coupler's resistive termination. For the output network, the surface mounted AC coupling capacitor is placed after the output combiner. The surface mounted capacitors are selected based on break down voltage, physical size, and self-resonant frequency. The 50-Ω resistive termination of the coupler is done through a surface mounted resistor, which is selected based on the same criteria. A summary of surface mounted components used is shown in Table 5.6. Finally, the complete layout of the RT/duroid package is shown in Figure 5.28.

Table 5.6: Summary of surface mounted components.

Component	Case size	Manufacturer	Part number	Quantity
1 pF capacitor	0.2 by 0.1 mm	muRata	GRM0335C1H1R0WD01D	3
50Ω resistor	0.4 by 0.2 mm	Vishay	FC0402E50R0F	1

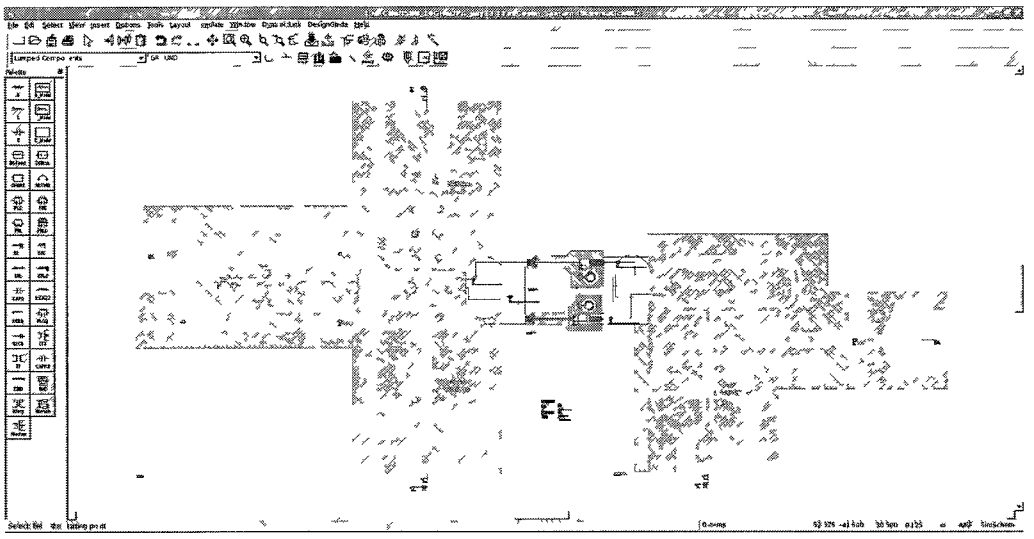


Figure 5.29: Co-simulation test-bench in *Agilent ADS*

5.3.5 Co-Simulation Results

Due to unexpected manufacturing delays, measurement results are not available at the time of writing this thesis However, very thorough co-simulations are performed to verify the operation of the 10 GHz Doherty SoP The circuit is co-simulated in the *Agilent ADS* environment [40] by combining EM simulation results for the ribbon bonds, RT/duroid package, and GaN MMIC with the foundry models of the GaN devices and manufacturer models of the surface mounted components. The test-bench used for the co-simulations is shown in Figure 5.29

Using the co-simulation test-bench the Doherty system performance is evaluated and compared to when it is biased as a balanced amplifier The S-parameters and stability K factor simulation results are shown in Figures 5.30 and 5.31, respectively Under Doherty operation the circuit is unconditionally stable. From Figures 5 32, 5.33, 5.34, the Doherty produces 35.6 dBm of output power at P_{1dB} with a PAE of 19 3% and a 6-dB back-off PAE of 9 9% When biased as a balanced PA, it has 2 5 dB more gain but the P_{1dB} is 2.6 dB lower which results in a PAE of 13 5% and a back-off PAE of 4 2% Under both biasing conditions the maximum output power is 10 W (40 dBm). The current consumption from 20V supplies is 0 135A for the Doherty and 0 51A for the balanced amplifier

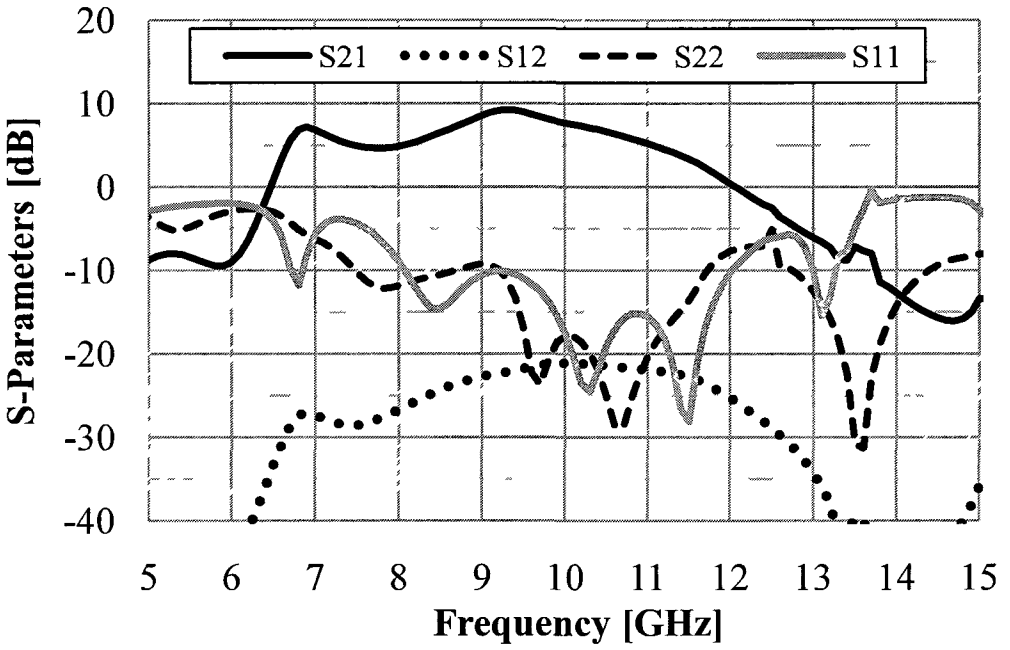


Figure 5.30: Co-simulated S-parameters of the 10 GHz GaN Doherty system.

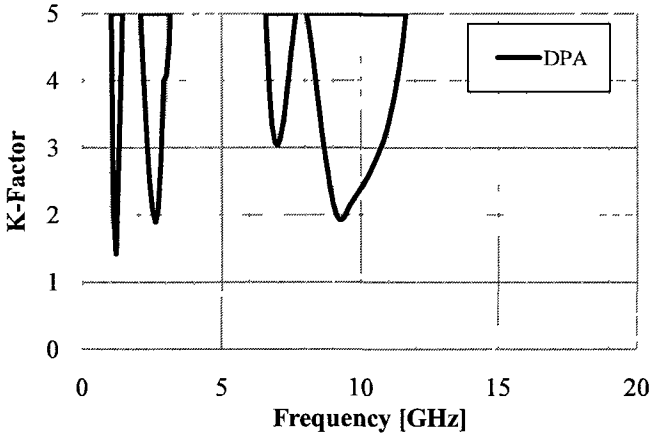


Figure 5.31: Co-simulated stability of the 10 GHz GaN Doherty system.

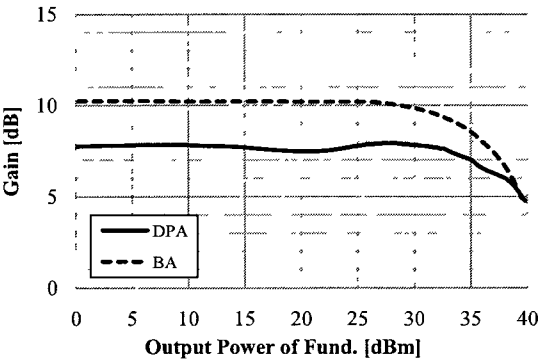


Figure 5.32: Co-simulated gain of the 10 GHz GaN Doherty system when biased as a Doherty and as a balanced amplifier (BA).

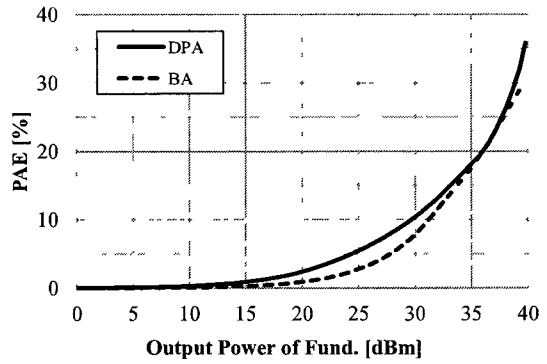


Figure 5.33: Co-simulated PAE of the 10 GHz GaN Doherty system when biased as a Doherty and as a balanced amplifier (BA).

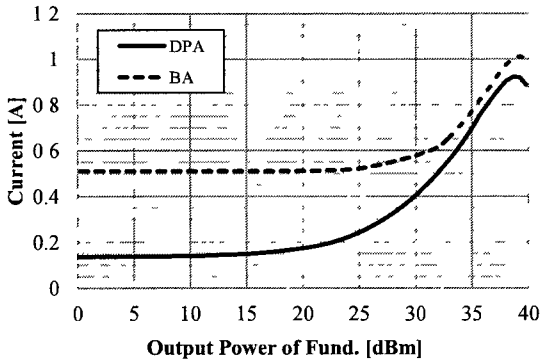


Figure 5.34: Co-simulated current consumption from 20V for the 10 GHz GaN Doherty system when biased as a Doherty and as a balanced amplifier (BA).

Finally, all the results are summarized and compared in Table 5.7.

Table 5.7: Summary of co-simulated results at 10 GHz.

	Doherty PA	Balanced PA
Main bias	class-AB (V_G = -3.5V)	class-A (V_G = -3V)
Peak bias	shallow class-C (V_G = -4.1V)	class-A (V_G = -3V)
Power dissipation	**0.135 A @ 20V**	0.51 A @ 20V
Gain	7.75 dB	**10.23 dB**
P_{1dB}	**35.6 dBm**	33.0 dBm
P_{SAT}	10 W	10 W
PAE @ P_{1dB}	**19.3%**	13.5%
PAE @ 6-dB back-off	**9.9%**	4.2%

5.3.6 Summary

A 10 GHz GaN Doherty amplifier was proposed and integrated with an SoP approach. The SoP technique was employed to tackle the limited current-handling, the low on-chip Q's, and the high cost of GaN MMICs. A 0.5-μm GaN MMIC was used for the two single-ended PA stages and a 6010LM RT/duroid package was used for the bulky input power splitting network as well as the high-power output combining network. The transitions between the GaN MMIC and the package were implemented with coplanar ribbon bonds which were absorbed into the design. The measurement results were not available at the time of writing but in-depth co-simulations were performed that included the effects of all components. The co-simulated results demonstrated a P_{1dB} of 35.6 dBm with a PAE of 19.3% and a 6-dB back-off PAE of 9.9%. The co-simulated maximum output power of the Doherty is 10 W. Compared to a balanced amplifier, the Doherty has a PAE that is 5.8% better at P_{1dB} and a PAE that is 5.7% better at 6-dB back-off from P_{1dB}.

5.4 Conclusions

This chapter started by discussing the design of a high-performance and compact 70-80 GHz branch-line coupler that can be used for system-on-chip CMOS Doherty implementations. The measured results of the coupler showed best reported performance in terms of insertion loss, phase imbalance, and amplitude imbalance. Then a similar coupler was used to design and implement a 71-76 GHz reconfigurable Doherty amplifier in CMOS. The simulation results showed a P_{1dB} of 8.9 dBm, a peak drain efficiency of 29.7%, and a 6-dB back-off drain efficiency of 11.4%. These results were not replicated in measurements due to a frequency shift in the fabricated prototype. After correcting the issue with the frequency shift, the *HFSS* simulated results showed very good agreement with measurements. A second Doherty circuit was then designed at 10 GHz using a GaN MMIC and an SoP integration technique in an attempt to get around the limitations of single-chip high power solutions. The technique absorbed the ribbon bonds as part of the input matching and output combining networks. Finally, the co-simulated results showed a maximum output power of 10 W as well as a PAE improvement of 5.8% at P_{1dB}, and an improvement of 5.7% at 6-dB back-off when comparing to a balanced amplifier.

Chapter 6

Conclusion

6.1 Summary

The research work in this thesis has been focused on improving the feasibility of MMIC based Doherty amplifiers at frequencies of 10 GHz and above. As shown in Chapter 2, the Doherty technique is especially efficient in power amplifying signals with high peak-to-average power ratios that are present in many modern high-data rate systems.

The high-frequency device limitations of deep sub-micron CMOS and GaN processes were studied in Chapter 3 to evaluate their effects on the Doherty system. Out of the high-frequency limitations, it was shown with a specifically developed Doherty model that the output conductance has the biggest impact on the efficiency characteristics. It was found that the ideal power added efficiency can degrade by as much as 25% when using realistic output conductances.

Chapter 4 served as a preamble to the Doherty circuit implementations by discussing the design of a complete CPW-based passive and active component library. All components necessary for the implementation of CMOS and GaN Doherty amplifiers were designed and measured.

Doherty circuits in two separate areas are improved in Chapter 5 by employing two integration techniques. A system-on-chip approach was used at 71-76 GHz to improve the efficiency of mm-wave CMOS Doherty PAs, while a system-on-package approach was used at 10 GHz to circumvent the current-handling limitations and high costs of completely integrated high power Doherty PAs. The CMOS SoC implementation was made possible by the successful design of a high-performance 70-80 GHz branch-line coupler. Measurement results of the coupler from 70-80 GHz showed an insertion

loss less than 1.4 dB, an amplitude imbalance less than 0.6 dB, a phase imbalance less than 2°, and a return loss better than 19.5 dB. Then a similar coupler was used in the design and implementation of a 71-76 GHz Doherty SoC; the Doherty SoC showed a 5.2% back-off efficiency improvement over a balanced amplifier. Final design section of this thesis presented the SoP implementation of a high power Doherty PA, which integrated a GaN MMIC with an RT/duroid package through the use of ribbon bonds. The distributed effects of the ribbon bonds were absorbed as part of the input matching and output power combining networks. Compared to a balanced amplifier, the high power SoP implementation resulted in twice the back-off efficiency.

6.2 Thesis Contributions

Major contributions developed with this thesis include:

1. An explanation is proposed in Section 3.2.1 for the high-frequency limitations of deep sub-micron CMOS Doherty PAs. The limitations include large knee voltages, reduced supply voltages, and an increase in output conductances. A Doherty system model capturing these limitations is developed and presented at IEEE NEWCAS [1].

2. An existing single trap model shown by [38] is extended in Section 3.2.2 to two traps to represent the output impedance collapse of the 0.5-μm GaN HEMT devices used for this work.

3. Three performance measures are proposed in Section 4.2 to size power devices for Doherty power amplifiers. These measures include maximum stable gain (MSG), input series Q, and the r_{ds}/R_{OPT} ratio which was found to be important for the back-off efficiency characteristics of the Doherty.

4. An article is submitted to IEEE Microwave and Wireless Components Letters [2] presenting a new high-performance area reduction technique for CMOS branch-line couplers. The presented coupler operates from 70-80 GHz and achieves best reported measured performance in terms of: insertion loss, phase imbalance, and amplitude imbalance.

5. An efficiency improvement of high-frequency Doherty amplifiers was demonstrated with two implementations. First, simulations of a 71-76 GHz Doherty

amplifier in Section 5.2 showed a 5.2% back-off efficiency improvement over a balanced amplifier. Secondly, a co-simulated high-power 10 GHz Doherty system in Section 5.3 showed more than twice the back-off efficiency of a balanced amplifier.

6. A system-on-package approach for Doherty power amplifiers is proposed in Section 5.3 to improve the output power as well as to reduce the real estate costs of GaN MMICs. The package absorbs the input power splitter to save area and the output power combiner to get around the poor current-handling of on-chip metallization layers.

7. A novel method to absorb ribbon bonds into the design of 10 GHz system-on-package Doherty power amplifiers is proposed in Section 5.3.3. Input bonds are used to increases the input impedance of the GaN power devices and output bonds are included as part of the output power combining network.

6.3 Suggestions for Future Work

Possible directions for future work could be as follows:

1. The 71-76 GHz CMOS implementation can be scaled to a 65-nm process node. This would certainly improve the gain but would also increase the drain efficiency as it was shown by Figure 3.6. Using a 65-nm process may also reduce the output impedance collapse because smaller node devices have less junction parasitic capacitance due to a physically smaller drain area.

2. The output impedance collapse which was found to affect CMOS devices beyond 10 GHz could be almost eliminated through the use of a silicon-on-insulator (SOI) process. Since SOI processes do not possess an R_{sub}, they are naturally protected from the output impedance collapse mechanism present in CMOS. This could result in the classic two peak efficiency response at mm-wave frequencies. However, it should be mentioned that the insulating layer in SOI processes is not very thermally conductive, so the thermal issues must be considered.

3. The physical size of the quarter-wave transformer forming the output power combiner of the Doherty could be reduced by absorbing the drain capacitive parasitics of the two stages. This would be an extension to the reduction method in [52] and could result in a significant area reduction of SoC implementations.

List of References

[1] S. Shopov, R. E. Amaya, J. W. M. Rogers, and C. Plett, "Adapting the Doherty Amplifier for Millimetre-Wave CMOS Applications," in *IEEE Int. NEWCAS Conference*, pp. 229–232, June 2011.

[2] S. Shopov, R. E. Amaya, J. W. M. Rogers, and C. Plett, "High-Performance and Compact 70-80 GHz CMOS Branch-Line Coupler using CPW's and Hybrid CPWG's," *submitted to IEEE Microwave and Wireless Components Letters*, July 2011.

[3] B. Hirosaki, "An Orthogonally Multiplexed QAM System Using the Discrete Fourier Transform," *IEEE Trans. On Communications*, vol. 29, pp. 982–989, July 1981.

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

[5] W. H. Doherty, "A New High Efficiency Power Amplifier for Modulated Waves," *Proceedings of the IRE*, vol. 24, pp. 1163–1182, September 1936.

[6] F. H. Raab, "Efficiency of Doherty RF Power-Amplifier Systems," *IEEE Trans. On Broadcasting*, vol. BC-33, pp. 77–83, September 1987.

[7] N. Wongkomet, L. Tee, and P. R. Gray, "A +31.5 dBm CMOS RF Doherty Power Amplifier for Wireless Communications," *IEEE Journal of Solid-State Circuits*, vol. 41, pp. 2852 – 2859, December 2006.

[8] L.-Y. Yang, H.-S. Chen, and Y.-J. E. Chen, "A 2.4 GHz Fully Integrated Cascode-Cascade CMOS Doherty Power Amplifier," *IEEE Microwave and Wireless Components Letters*, vol. 18, pp. 197 – 199, March 2008.

[9] E. Kaymakst and P. Reynaert, "A 2.4 GHz Fully Integrated Doherty Power Amplifier Using Series Combining Transformer," in *Proc. of the ESSCIRC*, pp. 302 – 305, 2010.

[10] J. Kang, D. Yu, K. Min, and B. Kim, "A Ultra-High PAE Doherty Amplifier Based on 0.13-μm CMOS Process," *IEEE Microwave and Wireless Components Letters*, vol. 16, pp. 505 – 507, September 2006.

[11] M. Elmala, J. Paramesh, and K. Soumyanath, "A 90-nm CMOS Doherty Power Amplifier With Minimum AM-PM Distortion," *IEEE Journal of Solid-State Circuits*, vol. 41, pp. 1323 – 1332, June 2006.

[12] M. Elmala and R. Bishop, "A 90nm CMOS Doherty Power Amplifier with Integrated Hybrid Coupler and Impedance Transformer," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, pp. 423 – 426, 2007.

[13] C. F. Campbell, "A Fully Integrated Ku-Band Doherty Amplifier MMIC," *IEEE Microwave and Guided Wave Letters*, vol. 9, pp. 114 – 116, March 1999.

[14] C. P. McCarroll, G. D. Alley, S. Yates, and R. Matreci, "A 20GHz Doherty Power Amplifier MMIC with High Efficiency and Low Distortion Designed for Broad Band Digital Communication Systems," in *IEEE MTT-S Int. Microwave Symposium Digest*, pp. 537 – 540, June 2000.

[15] J.-H. Tsai and T.-W. Huang, "A 38-46 GHz MMIC Doherty Power Amplifier Using Post-Distortion Linearization," *IEEE Microwave and Wireless Components Letters*, vol. 17, pp. 388 – 390, May 2007.

[16] B. Wicks, E. Skafidas, and R. Evans, "A 60-GHz Fully-Integrated Doherty Power Amplifier Based on 0 13-μm CMOS Process," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, pp. 69 – 72, 2008.

[17] A. V -G J.-W. Lai, "A 1V 17.9dBm 60GHz Power Amplifier in Standard 65nm CMOS," in *IEEE Int. Solid State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 424–425, 2010.

[18] A.-V. P. C.Y. Law, "A High-Gain 60GHz Power Amplifier with 20dBm Output Power in 90nm CMOS," in *IEEE Int. Solid State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 426–427, 2010.

[19] J. W. M. Rogers and C. Plett, *Radio Frequency Integrated Circuit Design, 2nd ed.* Norwood, MA: Artech House Publishers, 2010.

[20] Q. Gu, *RF System Design of Tranceivers for Wireless Communications.* New York, NY: Springer Science+Business Media, 2005.

[21] A. Tomkins, R. Aroca, T. Yamamoto, S. Nicolson, Y. Doi, and S. P. Voinigescu, "A Zero-IF 60 GHz 65 nm CMOS Transciever With Direct BPSK Modulation Demonstrating up to 6 Gb/s Data Rates Over a 2 m Wireless Link," *IEEE Journal of Solid-State Circuits*, vol. 44, pp. 2085–2099, August 2009.

[22] N. O. Sokal and A. D. Sokal, "Class E-A New Class of High-Efficiency Tuned Single-Ended Switching Power Amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 10, pp. 168–176, June 1975.

[23] A. V. Grebennikov, "Switched-Mode Tunes High-Efficiency Power Amplifiers: Historical Aspect and Future Prospect," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, pp. 49–52, 2002.

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

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

[26] L. R. Kahn, "Single-Sideband Transmission by Envelope Elimination and Restoration," *Proceedings of the IRE*, vol. 40, pp. 803–806, July 1952.

[27] F. H. Raab, P. Asbeck, S. Cripps, P. B. Kenington, Z. B. Popovic, N. Pothecary, J. F. Sevic, and N. O. Sokal, "Power Amplifiers and Transmitters for RF and Microwave," *IEEE Trans. On Microwave Theory and Techniques*, vol. 50, pp. 814–826, March 2002.

[28] Y. Yang, J. Cha, B. Shin, and B. Kim, "A Fully Matched N-Way Doherty Amplifier With Optimized Linearity," *IEEE Trans. On Microwave Theory and Techniques*, vol. 51, pp. 986 – 992, March 2003.

[29] H. Chireix, "High Power Outphasig Modulation," *Proceedings of the IRE*, vol. 23, pp. 1370–1392, November 1935.

[30] F. H. Raab, "Efficiency of Outphasing RF Power-Amplifier Systems," *IEEE Trans. On Communications*, vol. COM-33, pp. 1094–1099, October 1985.

[31] M. Iwamoto, A. Williams, P.-F. Chen, A. G. Metzger, L. E. Larson, and P. M. Asbeck, "An Extended Doherty Amplifier With High Efficiency Over a Wide Power Range," *IEEE Trans. On Microwave Theory and Techniques*, vol. 29, pp. 2472 – 2479, December 2001.

[32] T. C Kuo and B. B. Lusignan, "A 1 5 W Class-F RF Power Amplifier in 0.2μm CMOS Technology," in *IEEE Int. Solid State Circuits Conf. (ISSCC) Dig. Tech. Papers*, pp. 157–158, Feb. 2001.

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

[34] S. P. Voinigescu, T. O. Dickson, M. Gordon, C. Lee, T. Yao, A. Mangan, K. Tang, and K. H. K. Yau, "RF and Millimeter-Wave IC Design in the Nano-(Bi)CMOS Era," in *Si-based Semiconductor Components for Radio-Frequency Integrated Circuits (RFIC)* (W. Z. Cai, ed.), pp. 33–62, Transworld Research Network, 2006.

[35] T. O. Dickson, K. H. K. Yau, T. Chalvatzis, A. M. Mangan, E. Laskin, R. Beerkens, P. Westergaard, M. Tazlauanu, M.-T. Yang, and S. P. Voinigescu, "The Invariance of Characteristic Current Densities in Nanoscale MOSFETs and Its Impact on Algorithmic Design Methodologies and Design Porting of Si(Ge) (Bi)CMOS High-Speed Building Blocks," *IEEE Journal of Solid-State Circuits*, vol. 41, pp. 1830–1845, Aug. 2006.

[36] U. K. Mishra, L. Shen, T. E. Kazior, and Y.-F. Wu, "GaN-Based RF Power Devices and Amplifiers," *Proceedings of the IEEE*, vol. 96, pp. 288–305, February 2008.

[37] S. C. Binari, P. B. Klein, and T. E. Kazior, "Trapping Effects in GaN and SiC Microwave FETs," *Proceedings of the IEEE*, vol. 90, pp. 1048–1058, 2002.

[38] G. A. Umana-Membreno, J. M. Dell, B. D. Nener, and L. Faraone, "Low-Temperature Shallow-Trap Related Output-Admittance Frequency Dispersion in AlGaN/GaN MODFETs," in *Proc. Optoelectronic and Microelectronic Materials Devices*, pp. 252–254, 1999.

[39] W. R. Curtice and M. Ettenberg, "A Nonlinear GaAsFET Model for use in the Design of Output Circuits for Power Amplifiers," *IEEE Trans. On Microwave Theory and Techniques*, vol. 33, pp. 1383–1394, December 1985.

[40] "Agilent ADS." `http://eesof.tm.agilent.com/products/ads`.

[41] Y. Tsividis and C. Mcandrew, *Operation and Modeling of the MOS Transistor*. New York, NY: Oxford University Press, 2011.

[42] I. Wolf, *Coplanar Microwave Integrated Circuits*. Hoboken, NJ: John Wiley & Sons, 2006.

[43] C. Doan, S. Emami, A. Niknejad, and R. Broderson, "Millimeter-Wave CMOS Design," *IEEE Journal of Solid-State Circuits*, vol. 41, pp. 144–154, Jan. 2005.

[44] D. Pozar, *Microwave Engineering, 3rd ed.* Hoboken, NJ: John Wiley & Sons, 2005.

[45] R. Simons, *Coplanar Waveguide Circuits, Components, and Systems*. New York, NY: John Wiley & Sons, 2001.

[46] "Agilent Momentum." `http://eesof.tm.agilent.com/products/momentum`.

[47] C. Chen and et al., "A 90 nm CMOS MS/RF based foundry SOC technology comprising superb 185 GHz fT RFMOS and versatile, high-Q passive components for cost/performance optimization," in *IEEE IEDM Tech. Dig.*, pp. 2.5.1–2.5.4, 2003.

[48] "Ansoft HFSS." `http://www.ansoft.com/products/hf/hfss/`.

[49] "Sonnet Suites." `http://www.sonnetsoftware.com/products/sonnet-suites/`.

[50] G. Engen and C. Hoer, ""Thru-Reflect-Line": An Improved Technique for Calibrating the Dual Six-Port Automatic Network Analyzer," *IEEE Trans. on Microwave Theory and Techniques*, vol. 27, pp. 987–993, December 1979.

[51] A. Mangan, S. Voinigescu, M.-T. Yang, and M. Tazlauanu, "De-Embedding Tranmission Line Measurements for Accurate Modeling of IC Designs," *IEEE Trans. on Electron Devices*, vol. 53, pp. 235–241, February 2006.

[52] T. Hirota, A. Minakawa, and M. Muragachi, "Reduced-Size Branch-Line and Rat-Race Hybrids for Uniplanar MMIC's," *IEEE Trans. Microw. Theory Tech.*, vol. 38, pp. 270–274, March 1990.

[53] I. Haroun, J. Wight, C. Plett, A. Fathy, and D.-C. Chang, "Experimental Analysis of a 60 GHz Compact EC-CPW Branch-Line Coupler for mm-Wave CMOS Radios," *IEEE Microwave and Wireless Components Letters*, vol. 20, pp. 211–213, April 2010.

[54] K. Hettak, R. Amaya, and G. Morin, "A Novel Compact Three-Dimensional CMOS Branch-Line Coupler using the Meandering ECPW, TFMS, and Buried Micro Coaxial Technologies at 60 GHz," in *IEEE MTT-S Int. Microw. Symp. Dig.*, pp. 1576–1579, May 2010.

[55] M.-J. Chiang, H.-S. Wu, M.-L. Lee, and C.-K. Tzuang, "Design of Compact Ka-band Monolithic Branch-Line Coupler on Silicon Substrate," in *Asia Pacific Microw. Conf.*, pp. 2124–2127, Dec. 2009.

[56] A. Sutono, D. Heo, Y.-J. Chen, and J. Laskar, "High-Q LTCC-Based Passive Library for Wireless System-on-Package (SOP) Module Development," *IEEE Trans. Microw. Theory Tech.*, vol. 49, pp. 1715–1724, October 2001.