UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Efficient Transmitters for Wireless Communications in Nanoscale CMOS Technology

Permalink

https://escholarship.org/uc/item/4fd9r83v

Author

Chowdhury, Debopriyo

Publication Date

2010

Peer reviewed|Thesis/dissertation

Efficient Transmitters for Wireless Communications in Nanoscale CMOS Technology

by

Debopriyo Chowdhury

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

in

Engineering - Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ali M. Niknejad, Chair
Professor Elad Alon
Professor Paul Wright

Fall 2010

Efficient Transmitters for Wireless Communications in Nanoscale CMOS Technology

Copyright 2010
by
Debopriyo Chowdhury

Abstract

Efficient Transmitters for Wireless Communications in Nanoscale CMOS Technology

by

Debopriyo Chowdhury

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Ali M. Niknejad, Chair

The last decade has witnessed a tremendous growth in wireless communications. Todays consumers demand wireless systems that are low-cost, power efficient, reliable and have a small form-factor. This quest for ubiquitous wireless connectivity and the trend toward highly integrated solutions have opened up a new wave of challenges and opportunities for RF (radio-frequency) integrated circuit design. Since it often dictates both battery life and form factor, the transmitter and in particular the power amplifier (PA) is often the most challenging block in this integrated radio design. The grand vision for wireless transmitters is to merge as many components as possible, if not all, into a single die in an inexpensive technology. There is therefore growing interest in utilizing CMOS technologies for power amplifiers (PAs). However, the low-supply voltage of nanoscale CMOS technology, the loss of on-chip passives and the conductive silicon substrate make a fully-integrated PA design challenging. This thesis focuses on the design of fully-integrated PAs for modern wireless communication systems at RF (2.4GHz) as well as 60GHz frequencies. Transformer based matching networks have been studied for PA design and new modeling methods proposed in this work. It has been shown that there is tremendous area benefit of using transformers at 60GHz, while still preserving high performance. A prototype of a transformer-coupled PA has been designed at 60GHz in 90nm CMOS technology. The transformer design and modeling proposed at 60GHz is equally valid at RF frequencies. However, the high output power and high linearity requirements at RF frequencies create further challenges. Conventional power amplifier architectures are showing limitations in terms of achievable efficiency and area reduction. In particular, such architectures are not benefiting much from technology scaling since the area is dominated by passive elements. In this thesis, we investigate a mixed-signal power amplifier architecture. By merging our work on transformer-coupled PAs with a digital signal processing framework, a truly scalable, efficient transmitter architecture can be created. Such a prototype has been designed and tested in 65nm CMOS technology.

To The All-Mighty Goddess Ma Kali

Contents

List of Figures

List of Tables

Acknowledgments

This work would not have been possible without the help and influence of many people. First, I would like to thank my advisor Prof. Ali M. Niknejad for his help and support. As a graduate student in a new country, there are many challenges that one needs to handle, in addition to academic work. I am thankful to my advisor for being understanding at all times and for helping me stay focused and motivated. I would like to thank Prof. Elad Alon for his extensive help and feedback. I would also like to thank Prof. Seth Sanders and Prof. Paul Wright for being on my qualifying exam committee.

I would like to take this opportunity to thank all of the people I have met and worked with at BWRC. I have learnt more here in the last few years than I could have ever imagined and a lot of it is due to my classmates and seniors at BWRC: Cristian Marcu, Amin Arbabian, Ehsan Adabi, Wei-Hung Chen, Mohan Dunga, Zhiming Deng, Ali Afshar, Jason Stauth and others. Whether its helping me with work or just offering a friendly advice, I could not have done it without all of your help. Working with juniors like Lu Ye, Jiashu Chen, Lingkai Kong, Chintan Thakkar, Maryam Tabesh and others has been an enriching experience as well. My collaboration with Lu Ye in designing the digital transmitter was particularly fruitful and enjoyable.

I would like to offer my special thanks to Peter Haldi, a visiting scholar at BWRC in 2006. Special thanks also goes to Dr. Patrick Reynaert, a post-doc in our group in 2007. We had a wonderful time talking about various details and I got to learn a lot from him.

Finally, I would like to acknowledge the love and support of my parents Suprita and Bijan. Without their love and encouragement, I would have never been where I am now. Special thanks goes to my wife Souti for all her patience, love and encouragement. She always helped stay motivated during the tiring tapeout and measurement phases. Thanks to my sister Debjani for her love. Thanks a lot and this thesis is to you all.

Chapter 1

Introduction

The last decade has witnessed phenomenal growth of the wireless industry, and wireless connectivity has now reached a state where it is considered to be an indispensable service. A vast plethora of applications such as mobile internet, gaming, home networking, and multimedia streaming have motivated the growth of high data rate wireless communication systems. A plot highlighting the increasing demand for high data rate communication both for long-range (cellular) and short-range (like WLAN, Bluetooth etc) applications is shown in Fig. 1.1 [1]. Similarly, as depicted in Fig. 1.2, the demand for converged wireless devices like smartphones and PDAs has been growing year after year [2].

Driven by the need for higher data-rates, standards have evolved towards packing more bits per second per Hertz by using higher order modulation. Newer standards are also being developed that seek to utilize higher bandwidths, like the 7GHz of unlicensed bandwidth around 60GHz. Standards such as 802.11ad and WiGig are targeting this mm-wave band to create very high data rate short-range wireless personal area networks (WPANs) [4].

These trends towards ubiquitous, high data-rate wireless connectivity and highly integrated solutions have opened up a new wave of challenges and opportunities for RF (radio-frequency) and mm-wave integrated circuit design. Mobile terminals inherently operate on batteries having limited charge, and hence power-efficiency is one of the key metrics [5]. In addition, small form-factor, low-cost, and reliability are key characteristics of radios used in portable devices.

Most of the power dissipation in mobile wireless products is due to the transmitter, and specifically, due to the power amplifier (PA). The PA, which is the last building block in a conventional transmit chain (as shown in Fig. 1.3), is often times the most power-hungry. This is because the PA needs to deliver several tens to hundreds of milliwatts to a fixed load (like the antenna) and thus its power efficiency significantly impacts the battery life of a mobile device. In addition, PAs are often implemented as multi-stage designs, requiring a large number of passive elements (like inductors, capacitors, transmission lines etc.) for inter-stage matching. These passives consume significant area, increasing the die size and cost. This naturally leads to the conclusion that increasing the power efficiency while reducing
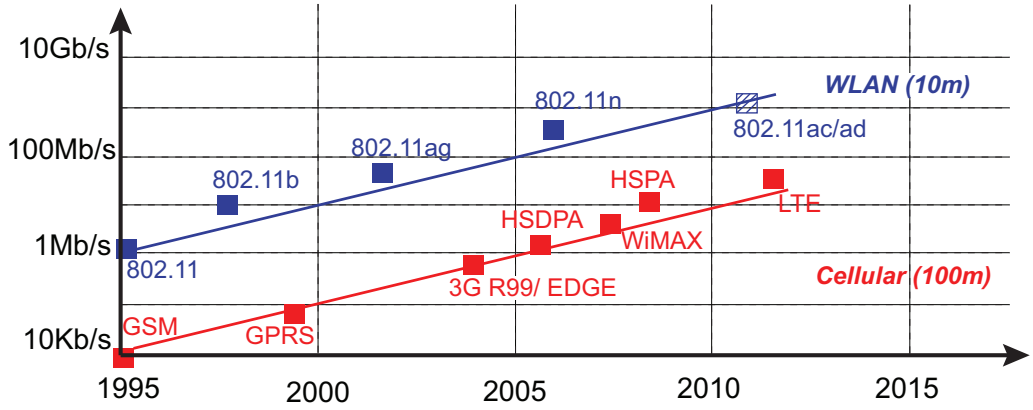
Figure 1.1: Growth of high data rate wireless communications [1].
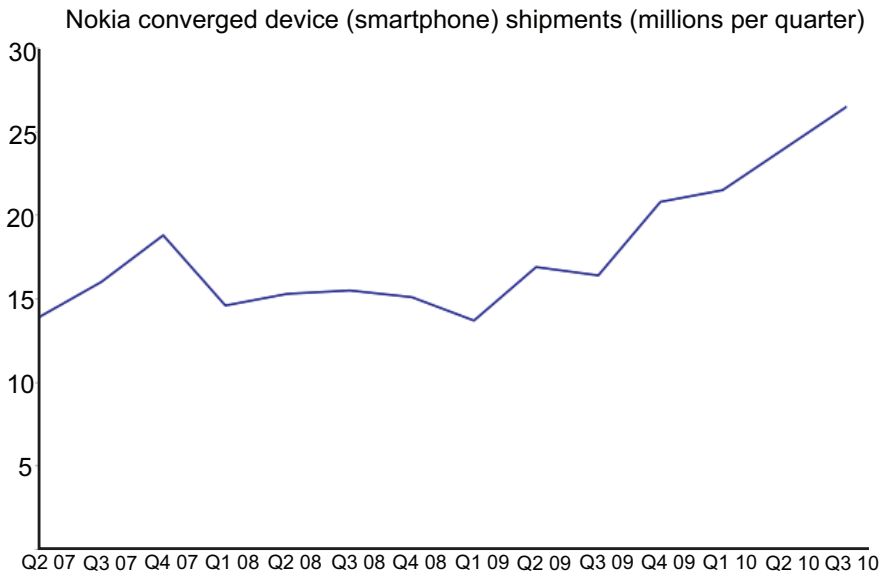
Figure 1.2: Growth in the sale of Nokia converged devices (smartphones) [2].
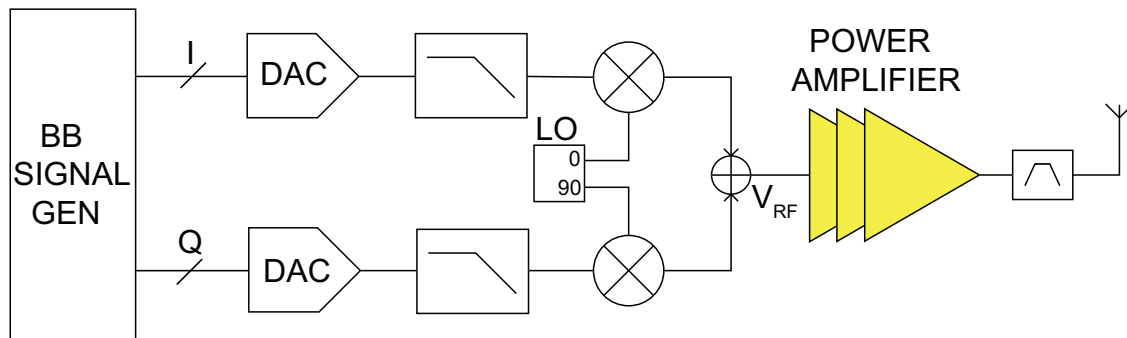
Figure 1.3: Block diagram of a wireless transmitter.

the size and cost of the power amplifier in particular - and the transmitter in general- is a key challenge in the evolution of mobile wireless transceivers.

One of the most effective ways to achieve compactness and reduce cost is a high level of integration. Integration of multiple circuit blocks into a single chip would reduce the size of the final product, the number of packages required, packaging and testing costs and printed circuit board footprint [5]. Thus, there has been tremendous effort by both industry and academia to realize a true single-chip radio in CMOS technology.

Over the last decade, most of the building blocks of an RF transceiver have been successfully integrated into CMOS technology. However, one piece that is missing from the puzzle is the power amplifier. The PA is often implemented as a stand-alone module in expensive processes like GaAs HBT (heterojunction bipolar transistors) in conjunction with external passives. The primary reasons for these are the high output power and power efficiency required by the PA, which are difficult to achieve in silicon based technologies. Although technology scaling typically provides faster transistors every generation, the breakdown voltage of the transistors also reduces every technology node. Thus, the power supply voltage needs to scale, making high power generation in silicon extremely challenging. In addition, on-chip passives suffer from low quality factor due to thin metals as well as the conductive nature of the silicon substrate. These losses degrade both the output power and efficiency of the PA. The design of a fully-integrated power amplifier thus remains a complex challenge. The primary goal of this thesis is to investigate efficient, integrated transmitter design, with special emphasis on integrated power amplifier design in nanoscale CMOS technology.

Since they play a crucial role in PAs, we will begin by exploring the design of passive matching networks. Matching networks are required for performing impedance transformation at the output of the PA. The PA needs to interface with a fixed external load like the filter or the antenna (which are often times 50Ω). If there is no matching network in between the PA and the fixed load, the output power that can be delivered to the load is limited. For example, a PA operating from a 1-V supply can deliver no more than 10dBm into a 50Ω load. To increase the output power beyond this, a matching network is required which will transform the 50Ω impedance to a lower value.

Such matching networks typically consist of passive components like inductors and capacitors which have finite quality factors. Hence matching networks are fundamentally lossy circuits. In PA applications, low insertion loss of the matching network is highly desirable since this loss directly degrades both the output power and the power efficiency of the PA. Furthermore, it is desirable to minimize the area of the passive components in a fully-integrated design.

The conventional way of performing such an impedance transformation is by using 'LC' matching networks [6]. However, it has been shown in [7] that the insertion loss of an 'LC' matching network increases rapidly as the impedance transformation ratio (ITR) is increased. This is clearly a problem in low-voltage high output power PA design because of the required high impedance transformation.

Transformers using coupled inductors are another class of impedance transformation circuit. It has been shown in [7] that transformers can ideally break the efficiency-ITR trade-off inherent in 'LC' matching networks. Hence, transformers have recently become popular at RF frequencies [8] and their application in PA design has also been demonstrated [9] [10] [11].

In this work, we want to extend the operating regime of transformers to mm-wave frequencies (specifically to 60GHz). Most of the reported circuits at mm-wave frequencies employ distributed matching networks using transmission lines [12] [13] [14] [15]. Such matching networks are easier to model, but are bulky and occupy large area, even at 60GHz. In contrast, we will demonstrate in Chapter 2 that by choosing an appropriate size and geometry, it is feasible to design a low-loss, compact transformer even at 60GHz. The small size of such transformers is a clear win over the distributed matching networks.

In order to utilize these transformers to design efficient power amplifier circuits, a simple, scalable model needs to be developed. The conventional way of modeling a transformer is by using lumped elements [16] [17]. We will show in Chapter 2 that such a lumped element model of the transformer can predict its performance faithfully even up to mm-wave frequencies. Despite its accuracy, the lumped model suffers from the fact that it employs many parameters - some of which can be predicted by analytical equations but some need to be fitted based on simulations. Thus it is challenging to make such a model truly scalable with transformer size and hence difficult to use during circuit design.

In this work, we propose a distributed modeling methodology for on-chip transformers. Utilizing transmission-line equations, we will show that we can predict the insertion loss and impedance transformation ratio of the transformer over a wide range of frequency. The proposed model has far fewer parameters and is also length-scalable, making it useful during the design phase of the PA.

In order to verify the utility of the designed transformers and the proposed modeling method, a 60GHz fully-integrated, transformer-coupled power amplifier will be presented in Chapter 3. It will be shown that in addition to performing impedance transformation, transformers offer additional benefits in PA design like differential-to-single ended conversion and convenient biasing through the center tap of the winding. A systematic design

methodology will be developed to facilitate the co-design of PAs and transformers. Utilizing this algorithm, a two-stage as well as a high-gain three stage power amplifier will be demonstrated in 90nm CMOS technology. Operating from a 1V supply, this PA delivers more than 12dBm of output power, which is one of the highest reported for a 60GHz CMOS PA without power combining. The compact size of the reported PA clearly shows the area benefit of transformers at these frequencies.

The design and modeling of transformers developed in this work is also useful at RF frequencies. However, the output power requirements in the low-GHz wireless systems is often higher as they are meant for long-range communication. In low voltage CMOS process, higher output power generation translates to a higher impedance transformation ratio in the output matching network. An efficient way to realize such a high transformation ratio with low insertion loss is to use power-combining.

Different types of power combining architectures have been proposed in literature [18] [19]. Aoki *et al* [7] proposed a novel on-chip impedance matching and power-combining method called the distributed active transformer (DAT). It combines several low-voltage differential amplifiers efficiently with their outputs in series to produce a larger output power while maintaining a 50Ω match. However, in the original design, the amplifiers driving the DAT were not independent of each other and hence individual amplifiers could not be turned off when less output power was required. [20] proposed a modified version of the DAT where individual 1:1 transformers were cascaded in series to realize an on-chip power combiner. While similar in principle to the DAT operation, the power combiner in [20] allowed independent amplifiers to be turned off. In this thesis, we have adopted a similar power combiner architecture to generate high output power at RF frequencies. A modified layout strategy that can ensure better PA linearity will be discussed in Chapter 4.

Indeed, the linearity of the power amplifier is becoming increasingly important in newer wireless systems. The use of higher order modulation like 64-QAM, often coupled with OFDM (orthogonal frequency division multiplexing), leads to very high dynamic range signals. Processing such signals with high integrity while delivering high output power poses a significant challenge for the PA. Some of the older wireless standards like GSM employed constant envelope modulation and hence linearity was not the chief concern. Consequently many of the original CMOS power amplifiers published in literature were switching PAs [21] [22]; in fact, the original PA using the DAT [23] was also such a switch-mode. Because of the evolution of wireless standards employing both amplitude and phase modulation, it is only more recently that PA linearity requirements have become so stringent.

There has been a lot of work to improve the linearity of the power amplifier [24] [25] [26]. In this thesis, we have tried to come up with simple circuit techniques that can increase the linearity of the core PA block. We have analyzed the desired amplitude and phase response that allows the PA to meet the linearity requirements with the maximum efficiency. We will show in Chapter 4 that the choice of an optimal gate bias for the output stage, together with capacitive phase compensation, can enable a highly-linear PA design without sacrificing efficiency. Additional linearization techniques like those proposed in [24] and other work can

be applied to further improve the linearity.

In order to verify the effectiveness of our proposed scheme, a two-stage 2.4GHz Class-AB power amplifier has been fabricated in 90nm CMOS technology. Transformer-based power combining has been used for the matching network design. With 37% efficiency, the PA delivers more than 1W of output power, which is one of the highest reported in literature for a linear CMOS PA. The PA can meet the linearity requirements of modern wireless standards like WiMAX with 18% efficiency.

It is well-known that there exists a fundamental linearity-efficiency trade-off in conventional power amplifier design. We have shown in Chapter 4 that PAs like Class A/AB/B can achieve high linearity with proper design. However, the efficiency of such PAs is fundamentally lower than that of switching power amplifiers like Classes D/E/F [27]. Unfortunately, traditional transmitter implementations (Fig. 1.3) do not allow the use of such switching PAs in systems that employ amplitude modulation. Therefore, numerous methods have been proposed to either improve the efficiency of linear amplifiers or linearize the more efficient nonlinear amplifiers without significantly degrading their efficiency.

One such technique is envelope-elimination and restoration (EER) and is based on performing efficient constant-envelope amplification of the phase-modulated portion of the input signal and then restoring the envelope by amplitude modulation at the output [28]. In effect, an EER system is a polar transmitter in which the signal is decomposed into a constant-envelope RF phase-modulated signal and an envelope component. Several papers have demonstrated the suitability of such architectures for CMOS implementation [29] [30]. However, the efficiency-bandwidth trade-off in analog supply modulators has restricted the popularity of this architecture to relatively narrowband modulation standards.

In order to expand such a polar architecture to wideband systems such as 802.11g, 802.11n, and others, in this thesis we study a mixed-signal polar transmitter in 65nm CMOS technology. The proposed architecture does not employ a supply modulator but instead operates the power amplifier as an RF-DAC. The phase-modulated signal being constant-envelope, the core PA can still be efficient but non-linear (from the standpoint of amplitude modulation); the amplitude information is transmitted by turning on and off an appropriate number of unit cells.

In addition to allowing the use of a non-linear, efficient PA, such a mixed-signal architecture also requires less passive matching networks. This is important because the passives do not scale well with technology and hence in a multi-stage PA the area is often dominated by the passive elements. A truly scalable architecture, which will benefit from technology scaling like digital circuits, is thus highly desirable for low-cost wireless solutions. A mixed-signal polar architecture offers such a potential.

Some previous publications have discussed the concept of such a digitally modulated transmitter [31] [32] [33] [34]. However, not much attention has been paid on the design of a high output power, high efficiency, fully integrated switching PA. For example, [31] uses a Class-A amplifier as the unit cell. But, it would be much more efficient to use switching PAs in such an architecture. [32] proposes the use of Class-E power amplifiers. However, the

targeted output power level is much lower and the amplifier basically acts as a PA driver. There is thus little work done on a detailed study of this digital polar architecture with an emphasis on an efficient, high-power, fully-integrated PA design. The real benefit of such a mixed-signal transmitter can be leveraged only if the PA is more efficient that what can be achieved by a linear amplifier like the one designed in Chapter 4. The core PA design is thus an integral part of such a transmitter.

This thesis therefore proposes the use of an efficient, switching inverse Class-D PA as the core of such a mixed-signal system. We will analyze its advantages, design methodology, linearity, and back-off characteristics in details in Chapter 5. The power of digital signal processing will be leveraged to compensate for the PA non-idealities.

The other piece missing from previous digitally modulated transmitter designs is a consideration of the out-of-band noise for co-existence issues. Meeting the spectral mask specified by the standard is a necessity; however, care also needs to be taken to ensure that the far out-of-band noise is low enough so that it does not desensitize the receiver of another radio in a multi-standard solution. In this work, we have analyzed the required amplitude and phase path resolution to meet not only the linearity requirements, but also to ensure low out-of-band noise. Furthermore, we will be demonstrate the design of a low-power interpolation filter that can be used to prefilter any replicas of the baseband signal that remain after oversampling.

A prototype of this mixed-signal transmitter with integrated filtering has been fabricated in 65nm CMOS technology. The PA achieves about 22dBm of output power with over 44% efficiency. It requires only one transformer in the whole design, even though it has an effective power gain of more than 30dB. As will be presented in Chapter 5, the performance of the designed transmitter is even comparable to some commercial power amplifiers implemented in expensive compound semicondcutor processes.

Chapter 2

Passive Matching Networks for Power Amplifier Design

In the previous chapter, passive matching networks were identified as one of the key blocks enabling an efficient, low-voltage power amplifier in CMOS technology. Because lossless reactive components do not exist, all impedance matching networks will have some insertion loss. These losses can be quite pronounced in PAs since they directly degrade the efficiency. This issue is even more serious when designing PA with state-of-the-art low-voltage CMOS technologies because of the high required impedance transformation ratio and the lossy on-chip passive components. Two different types of matching networks have been used mostly in PA design: LC resonant matching networks and transformer-based matching networks. In this chapter, the benefit of transformers vis-a-vis conventional LC matching networks will be elaborated. Specific design examples for low-loss RF and 60GHz transformers will be presented, and a simplified modeling method for on-chip transformers will be introduced. Finally, the chapter will conclude with the potential of transformers as power combiners to build high output-power amplifiers in low-voltage CMOS process.

2.1 LC Matching Networks

LC resonant matching networks are straightforward to implement. A single stage L-match is shown in Fig. 2.1, consisting of a capacitor and an inductor. A lossy inductor can be analyzed as a series equivalent circuit or a parallel equivalent circuit by adding a resistor to model the loss (Fig. 2.2) [35], where

$$Q_L = \frac{\omega L_s}{R_s} \tag{2.1}$$

$$R_p = R_s(Q_L^2 + 1) \tag{2.2}$$

$$L_p = L_s \frac{Q_L^2 + 1}{Q_L^2} \approx L_s \tag{2.3}$$
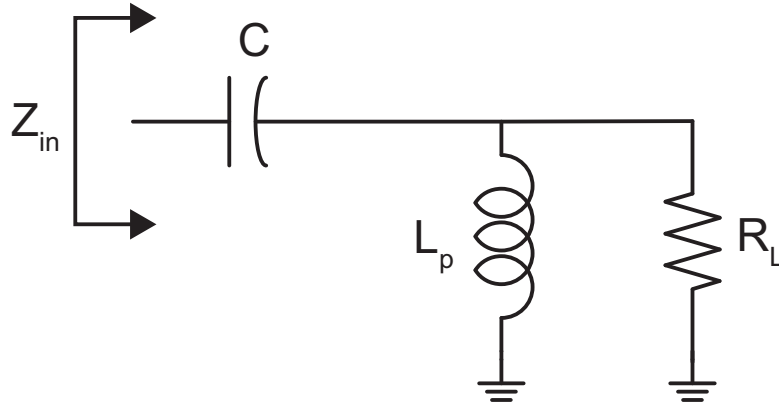
Figure 2.1: A single stage "L-match" with a series capacitor and a parallel inductor.

Assuming that at the frequency of interest L_p and C resonate with each other, it can be shown that

$$Q_{network} = \frac{R_p || R_L}{\omega L_p} = \frac{R_L}{\omega L_p + \frac{R_L}{Q_L}} \tag{2.4}$$

$$R_{in} = Re(Z_{in}) = \frac{R_p || R_L}{1 + Q_{network}^2} = \frac{1}{1 + Q_{network}^2} \cdot \frac{R_L}{1 + \frac{R_L}{\omega L_p Q_L}} \tag{2.5}$$

The impedance transformation ratio (ITR) is given by:

$$ITR = \frac{R_L}{R_{in}} = \frac{R_L}{R_p || R_L}(1 + Q_{network}^2) = (1 + \frac{R_L}{\omega L_p Q_L})(1 + Q_{network}^2) \tag{2.6}$$

The above equation shows that as the impedance transformation ratio increases, the $Q_{network}$ also increases. This has an impact on the network efficiency as shown next.

The efficiency of a matching network can be calculated as the ratio of the power delivered to the load to the power delivered into the network. The efficiency (η) can be computed as
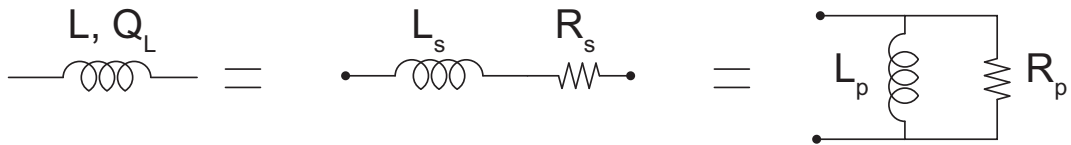
Figure 2.2: Equivalent models of an inductor with quality factor Q_L.
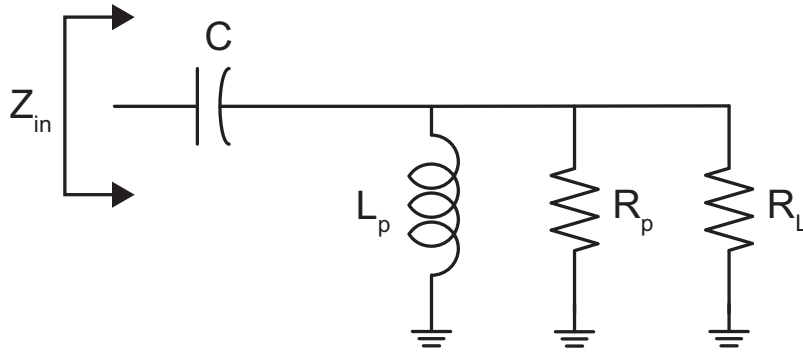
Figure 2.3: Single stage L-match network with inductor loss.

[7]

$$\eta = \frac{P_{out}}{P_{in}}$$
$$= \frac{|V_l|^2/(2R_L)}{|V_l|^2/(2(R_p||R_L))}$$
$$= \frac{1/R_L}{1/R_L + 1/R_p}$$
$$= \frac{1}{1 + \frac{R_L}{\omega L_p Q_L}} \tag{2.7}$$

Equations 2.6 and 2.7 can be solved for ωL_p in terms of the desired impedance transformation ratio (ITR), load resistor (R_L), and inductor quality factor (Q_L).

$$\omega L_p = \frac{R_p}{Q_L} = \frac{2(Q_L + 1/Q_L)}{ITR - 2 + \sqrt{ITR^2 + 4Q_L^2(ITR - 1)}} \cdot R_L \tag{2.8}$$

The efficiency of the network can be simplified in terms of ITR and Q_L by substituting Eq. 2.8 in Eq. 2.7.

$$\eta = \frac{1 + Q_L^2}{Q_L^2 + \frac{ITR + \sqrt{ITR^2 + 4Q_L^2(ITR-1)}}{2}} \approx \frac{1}{1 + \frac{ITR}{Q_L^2}} \tag{2.9}$$

The efficiency of the LC matching network is plotted in Fig. 2.4 for different inductor quality factors. As seen from the plot, for a given inductor quality factor, the power efficiency drops quickly as ITR increases. The reason for this is that in the resonant network, $Q_{network}$ times higher energy circulates in the LC tank than the energy delivered to the load [35]. As ITR increases, $Q_{network}$ increases, so more energy is dissipated in the lossy inductor. This is a fundamental problem in PA design, since higher impedance transformation is required for low-voltage operation. This problem can be partially mitigated using multi-stage matching networks, but at the expense of higher passive component count and higher silicon area.
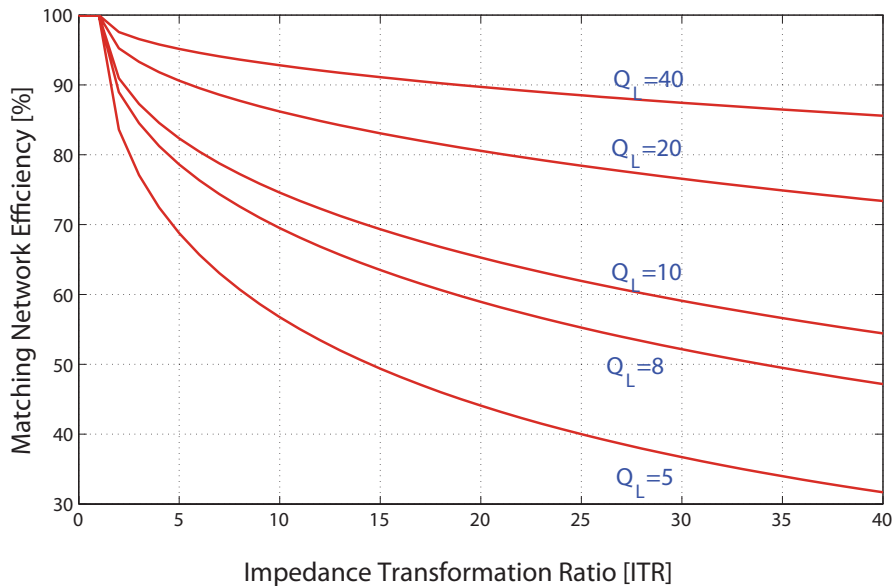
Figure 2.4: Efficiency of an LC matching network for different inductor quality factors.

2.2 Transformer-Based Matching Networks

Transformers- as their name implies- are another class of impedance transformation network. In this section, we will demonstrate that transformers, to first order, can overcome the ITR - efficiency tradeoff fundamental to LC matching networks. This is the reason why transformers have recently attracted interest from the power amplifier design community.

Transformers are based on the principle of electro-magnetic induction. In silicon, a transformer is implemented by two coupled inductors [36]. In a coupled-inductor transformer (Fig. 2.5), the magnetic field created by primary inductor L_p generates a voltage in the secondary inductor L_s. At the same time, the current through the secondary winding I_s will magnetically induce a voltage in the primary circuit. The port voltages of the transformer V_p and V_s in Fig. 2.5 are related to its port currents through

$$\begin{bmatrix} V_p \\ V_s \end{bmatrix} = \begin{bmatrix} j\omega L_p & -j\omega M \\ j\omega M & -j\omega L_s \end{bmatrix} \cdot \begin{bmatrix} I_p \\ I_s \end{bmatrix} \tag{2.10}$$

$$M = k \cdot \sqrt{L_p L_s} \tag{2.11}$$

$$n = \sqrt{\frac{L_p}{L_s}} = \frac{I_p}{I_s} = \frac{V_s}{V_p} \tag{2.12}$$

where M is the mutual inductance, k is the coupling factor, and n is the turn ratio between the primary and secondary windings. Due to the finite quality factor of on-chip inductors
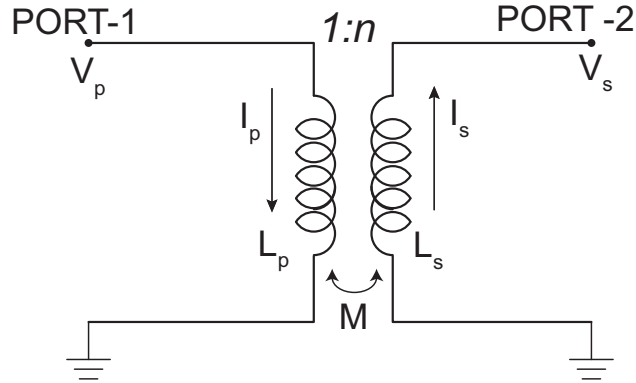
Figure 2.5: Ideal transformer model.

and the non-unity coupling between primary and secondary windings, a real transformer deviates from the ideal equations described above. Fig. 2.6 shows a simplified model of a real transformer, where Q_p and Q_s represent the quality factor of the primary and secondary windings, R_L is the load resistance and C_p is the shunt tuning capacitance. We can convert
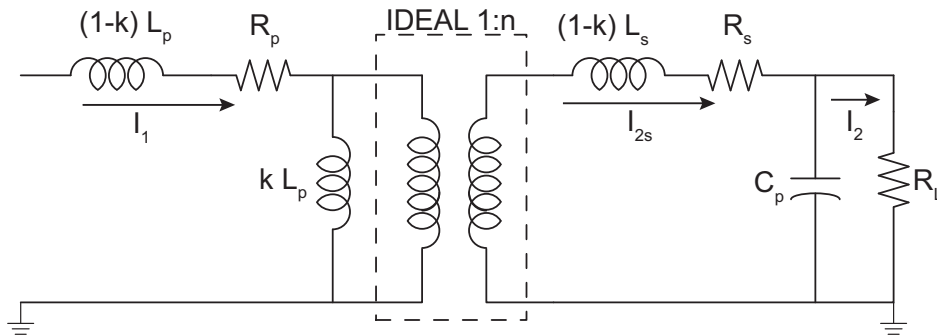
Figure 2.6: Simplified model of a 1:n transformer.

the shunt capacitor-resistor network into an equivalent series network given by [37]

$$R_{eq} = \frac{R_L}{1 + (\omega R_L C_p)^2} \tag{2.13}$$

$$C_{eq} = \frac{1 + (\omega R_L C_p)^2}{\omega^2 R_L{}^2 C_p} \tag{2.14}$$

The efficiency η of the transformer can be written as

$$\eta = \frac{P_{load}}{P_{load} + P_{diss}} \tag{2.15}$$

where P_{load} and P_{diss} are the power delivered to the load and power dissipated in the parasitic resistance of the transformer. Calculating the values of P_{load} and P_{diss} from the circuit and putting them in the above equation, the value of efficiency can be simplified to

$$\eta = \frac{\frac{1}{1+(\omega R_L C_s)^2} R_p}{\frac{1}{1+(\omega R_p C_p)^2} R_p + R_s + \left| \frac{Z_s + j\omega k L_p}{j\omega k L_p} \right|^2 R_p} \tag{2.16}$$

where Z_s is the secondary side impedance transformed to the primary side, as shown below.

$$Z_s = \frac{R_s + j\omega(1-k)L_s + \frac{1}{j\omega C_{eq}} + R_{eq}}{n^2} \tag{2.17}$$

The efficiency will be maximized if

$$\omega L_s = \frac{1}{\omega C_{eq}} \tag{2.18}$$

assuming $L_p \simeq L_s/n^2$. This gives a design equation which says that the optimum value of C_{eq} is one that tunes out L_s and not $(1-k)L_s$, as might appear from a first look at the equivalent circuit of Fig. 2.6. Basically, this condition minimizes the current I_1 through R_p and its dissipated power by resonating inductors $(1-k)L_s/n^2$ and kL_p with the capacitor C_{eq}. Note that because we use shunt tuning and not series tuning, the value of the tuning capacitor depends on the load resistor, even for a fixed size transformer at a given frequency.

In order to find the optimum winding inductance, we can take the derivative of the efficiency with respect to L_p and set $\frac{\partial \eta}{\partial L_p}$ to zero [7]. This gives an optimum inductance value of

$$\omega L_p = \frac{\alpha}{1+\alpha^2} \frac{R_L}{n^2} \tag{2.19}$$

where

$$\alpha = \frac{1}{\sqrt{\frac{1}{Q_s^2} + \frac{Q_p}{Q_s} k^2}} \tag{2.20}$$

and Q_p and Q_s represent the quality factor of the primary and secondary windings.

Using this optimum inductance and writing the parasitic winding resistances using winding quality factor (Q), the optimum efficiency is [7]:

$$\eta_{max} = \frac{1}{1 + \frac{2}{Q_p Q_s k^2} + 2\sqrt{\frac{1}{Q_p Q_s k^2}\left(1 + \frac{1}{Q_p Q_s k^2}\right)}} \tag{2.21}$$

The above equation shows that the passive efficiency (η) can be maximized using a k as close as possible to unity. This is because the smaller the k, the larger is the fraction of the primary inductor current I_1 that will go through the magnetizing inductor kL_p. This reduces the power that is delivered to the load resistor. More importantly, unlike LC matching networks,
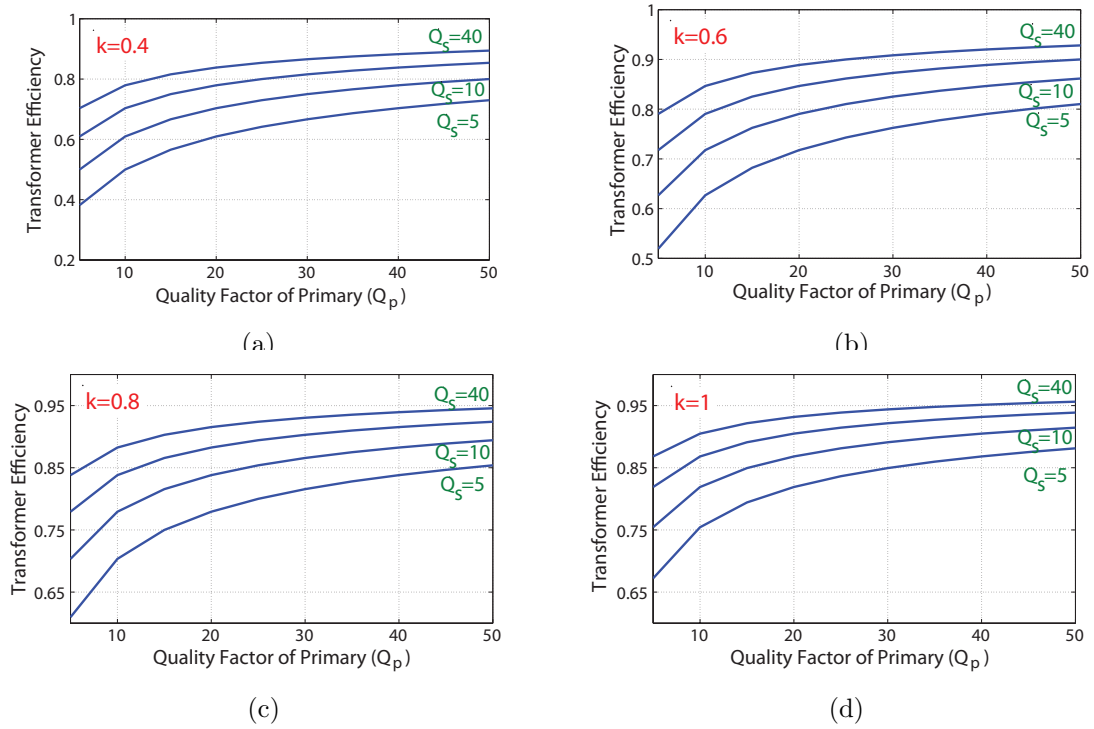
Figure 2.7: Transformer efficiency for different primary and secondary winding quality factors for a) $k = 0.4$ b) $k = 0.6$ c) $k = 0.8$ and d) $k = 1.0$.

the transformer efficiency, to first order, is not affected by the transformation ratio, as seen from Eq. 2.21.

In addition to impedance transformation with low insertion loss, transformers also allow easy DC biasing using center taps of windings and inherently perform differential-to-single ended conversion for interface to external filters or the antenna. These features have made transformers attractive for power amplifier design in CMOS technology.

2.3   Transformers at 60GHz

Transformers have gained some attention at lower GHz RF frequencies, but their use in mm-wave circuit design had not been demonstrated until recently [38]. Traditional microwave circuits employ distributed elements like microstrip lines, coplanar waveguides etc. to perform impedance transformation [3] [12] [14]. However, even at 60GHz, such transmission lines are bulky and increase silicon area, thereby incurring higher cost. Furthermore, they tend to be lossy at these frequencies due to the semi-conducting nature of the silicon substrate.

We have already seen that transformers fundamentally can perform impedance transformation as well as differential-to-single ended conversion. They are also very compact at mm-wave frequencies. In order to understand why transformers can be made much smaller in size at 60GHz, let us take a look at the equivalent circuit of a lossy transformer, redrawn in Fig. 2.8. Here the leakage inductance ($(1 - k^2)L_p$) and magnetizing inductance ($k^2 L_p$) have been transferred to the primary side. In order to ensure that most of the signal current on the primary (I_1) is transferred on to the load resistor (R_L), we need to ensure that

$$k^2 \omega L_p \gg R_L \tag{2.22}$$

The load resistance R_L (usually the antenna or the filter input impedance) is typically the same at RF and mm-wave frequencies (i.e., 50Ω). Hence as the operating frequency (ω) increases, the required primary inductance (L_p) can be reduced proportionally, while still keeping the impedance ratios the same. This is why transformers have the potential of being very compact at mm-wave frequencies.

Despite the potential for transformers to be very compact at mm-wave frequencies, for transformer-based mm-wave design to be practical, it needs to be demonstrated that the insertion loss of transformers at these frequencies is better than or comparable to distributed transmission lines. Furthermore, a predictable and scalable modeling methodology- good up to mm-wave frequencies- needs to be developed so that transformers can be easily co-designed with the active elements in power amplifier circuits.

In this work, we use 'overlay' (vertical) configuration for the transformer and utilize the top two metal layers. First, a family of planar octagonal loop inductors was simulated using a full 3-D electromagnetic-field simulator (HFSS) [39]. The effective inductance (L_{eff}) and quality factor (Q_{eff}) of the windings at 60GHz were extracted from the simulated s-parameters. Fig. 2.9 shows how the inductance (L_{eff}) and self-resonance frequency (SRF)
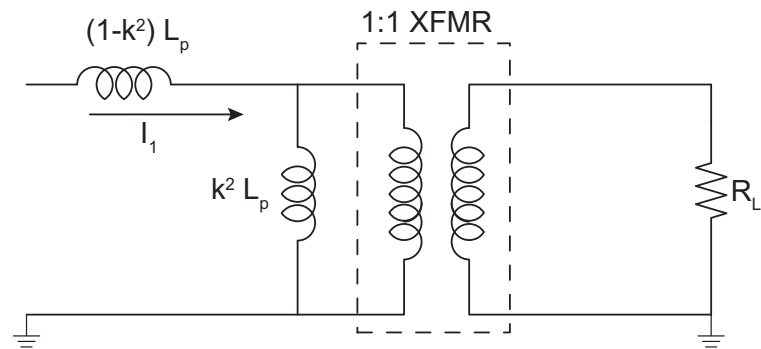
Figure 2.8: Equivalent circuit of a 1:1 transformer.
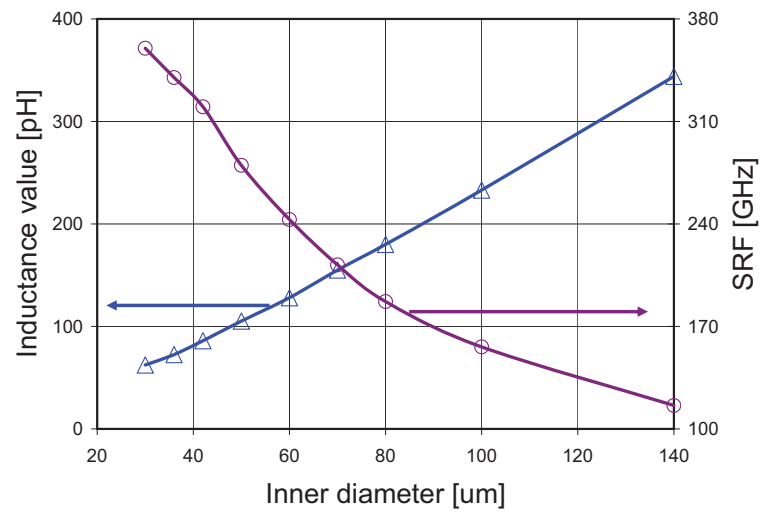
Figure 2.9: Variation of L_{eff} and SRF with diameter of inductor.
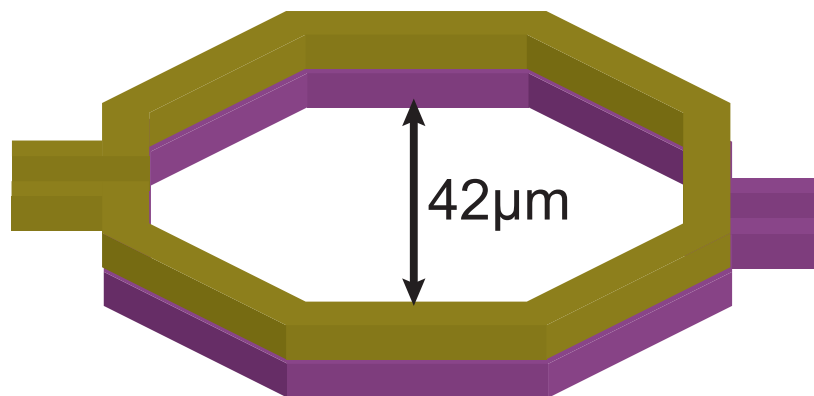
Figure 2.10: Layout of a 1:1 transformer using coupled inductors in a vertical configuration.
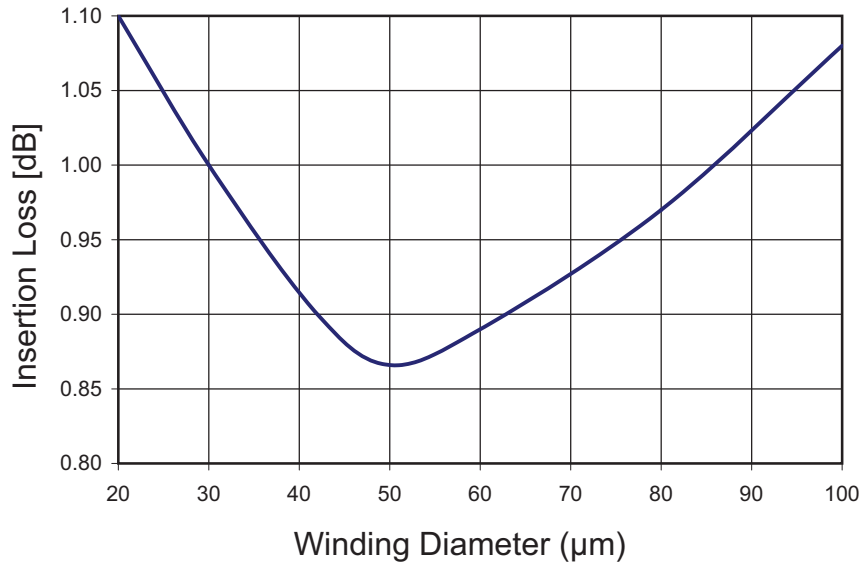
Figure 2.11: Variation of insertion loss of transformer at 60GHz as a function of winding diameter.

of an inductor varies as its inner diameter (d_{in}) is increased. Similar plots can be generated for different trace widths as well. From the plot, we see that for reasonable inductor sizes (up to about 350pH), the SRF remains above 100GHz.

The loop inductors discussed above were used in an overlay configuration to build vertical 1:1 transformers (Fig. 2.10). Transformers with different diameters and trace widths have been implemented with the top two metal layers. Fig. 2.11 shows the simulated minimum insertion loss of 60GHz transformers, clearly indicating how the size can be optimized. For very small sizes, the impedance of the shunt magnetizing inductance ($k^2 \cdot \omega \cdot L_p$) is much smaller than the reflected load impedance and most of the signal current is lost through it. On the other hand, a larger transformer entails more substrate losses and increased series leakage inductance ($(1 - k^2) \cdot \omega \cdot L_p$) which acts a low pass filter, reducing signal transfer to the secondary size [40]. In the mm-wave domain, where a 90nm CMOS transistor's MSG (maximum stable gain) is only around 8dB, choosing an optimum transformer size is critical to enabling high performance and efficient circuits.

Prototype transformers were fabricated in a 90nm CMOS process [38]. A die photo of the fabricated transformer is shown in Fig. 2.12. The measured and simulated s-parameters show good agreement up to high frequencies. Fig. 2.13 shows the comparison for a 1:1 vertical transformer having a inner diameter of $42\mu m$ and width of $8\mu m$. Fig. 2.14 shows the measured maximum available gain (MAG) and power gain (G_p for a 50Ω load) of the same transformer. From the measurement results, we see that it is possible to design area-efficient transformers at 60 GHz with less than 1 dB of insertion loss. It has been reported in literature that matching networks employing distributed elements like coplanar waveguides
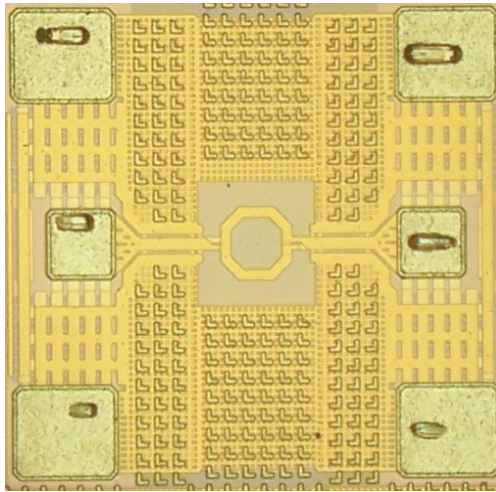
Figure 2.12: Die photo of a transformer with a diameter of $42\mu m$ and width of $8\mu m$.

have insertion loss of the order of 1.2-1.6 dB [3]. Thus we see that transformers can achieve similar performance, with dramatic reduction in silicon area.

Another advantage of transformers which becomes significant at 60GHz is related to the ease of biasing. The center-tap of the transformer winding in a differential circuit is a virtual ground at fundamental frequencies and can be conveniently utilized to provide DC bias. This allows us to eliminate the AC coupling capacitors, which have low quality factor at high frequencies and therefore incur significant loss.

The final advantage of lumped transformers stems from their lower sensitivity to the lossy silicon substrate. Lenz's law states that in an ideal transformer the magnetic flux produced by the current flowing in the primary winding must completely cancel the flux produced by the secondary current. Thus in an ideal transformer with unity coupling coefficient, there is no leakage flux in the area around the transformer. In reality, the coupling coefficient is less than one, and hence complete flux cancelation does not take place. However, the residual leakage flux is still much lower than in an inductor of similar area. This means that less magnetic fields impinge upon the lossy silicon substrate, thereby reducing losses due to effects like eddy current.

In order to verify this, two transformers were fabricated in 90nm CMOS technology. In one prototype, a p-well block layer was put underneath the transformer to make the substrate less conductive ($\rho = 10\Omega - cm$). In the other prototype, the p-well block layer was removed, thereby making the substrate more conductive and hence more lossy. However, measurements (Fig. 2.15) revealed that there was no difference in the insertion loss (S_{21}) and impedance transformation between the two versions, thereby demonstrating the lower sensitivity of transformers to substrate loss.
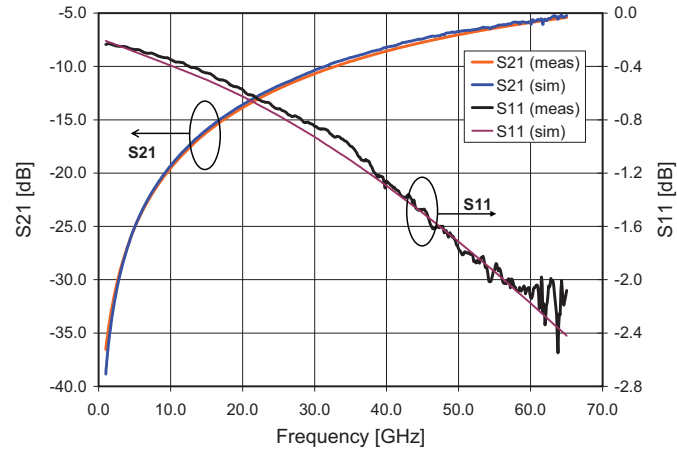
Figure 2.13: Simulated and measured s-parameters for a vertical transformer.
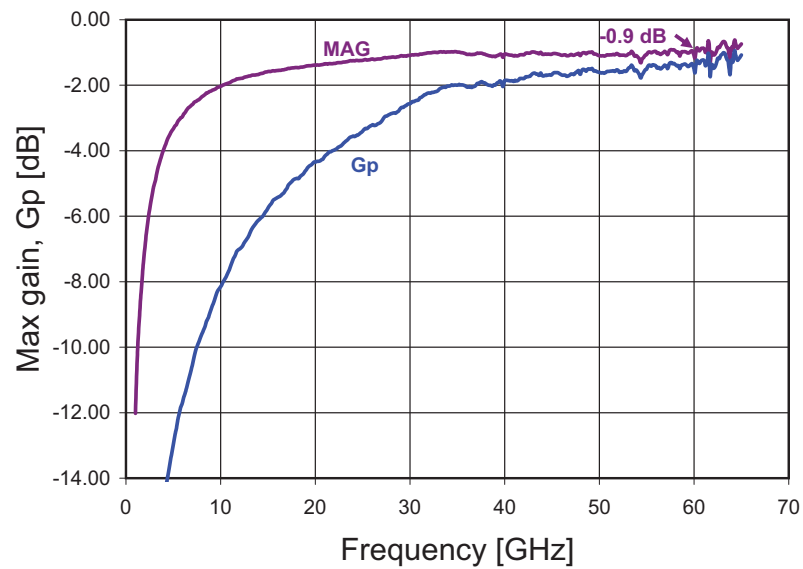
Figure 2.14: Measured insertion loss for a 1:1 vertical transformer.
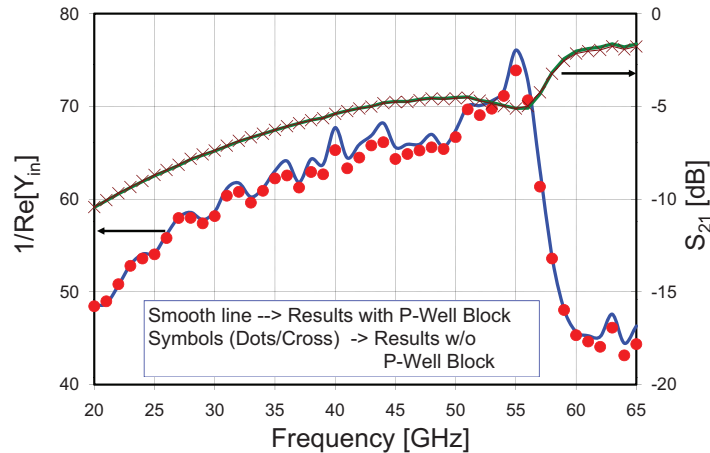
Figure 2.15: Dependence of transformer performance on substrate parameters.

2.3.1 Lumped Modeling of mm-wave Transformers

In the previous sub-section, the feasibility of transformers at 60GHz has been established. However, for useful circuit designs, transformers need to be integrated with active elements like transistors. In order to facilitate transformer-based circuit design, a simple model of a transformer, which can accurately predict the performance up to mm-wave frequencies, is needed.

In theory, an ideal 1:n transformer will transform a load resistor of value R_L to R_L/n^2, seen from the primary side. However, because of the presence of leakage inductances, parasitic resistors as well as inter-winding coupling capacitor, the impedance transformation at mm-wave frequencies is far from ideal. Fig. 2.16(a) shows the real part of the input impedance at 60GHz for different transformer sizes. As we see from the plot, even though the transformer is always 1:1 and the load resistor (R_L) and frequency are held fixed, the transformed impedance is a function of the actual size of the transformer. Similarly, Fig. 2.16(b) reveals that the impedance transformation even for a fixed size transformer is frequency-dependent. In order to predict the transformer size needed, we need a simple model that can capture these effects.

To facilitate transformer optimization, a lumped-element model has been built (Fig. 2.17). It is an extension of a conventional symmetric 2π model to mm-wave frequencies [16]. The model parameters are based on a mix of analytical equations and fitting parameters [17]. The dc inductance value of the primary/secondary windings has been calculated by approximating the sides of the spirals by symmetrical current sheets of equivalent current densities. For a given shape, a winding is completely specified by the number of turns (N), the turn width (w), the turn-to-turn spacing (s), and any one of the following: the outer diameter (d_{out}), the inner diameter (d_{in}), the average diameter $(0.5(d_{out} + d_{out}))$, or the fill ratio, defined as $\rho = (d_{out} - d_{in})/(d_{out} + d_{in})$. The expression for winding inductance is given

(a)

(b)
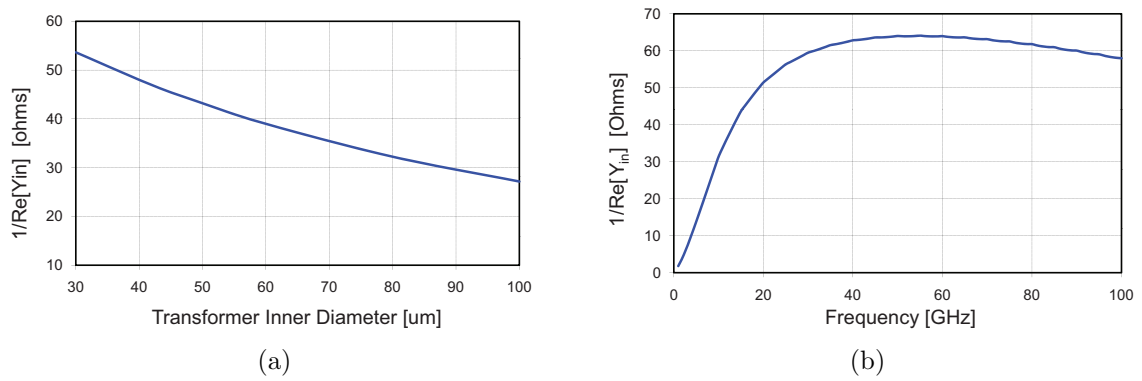
Figure 2.16: Input admittance of a transformer for a) different winding diameters at 60GHz b) a winding diameter of $42\mu m$ at different frequencies.
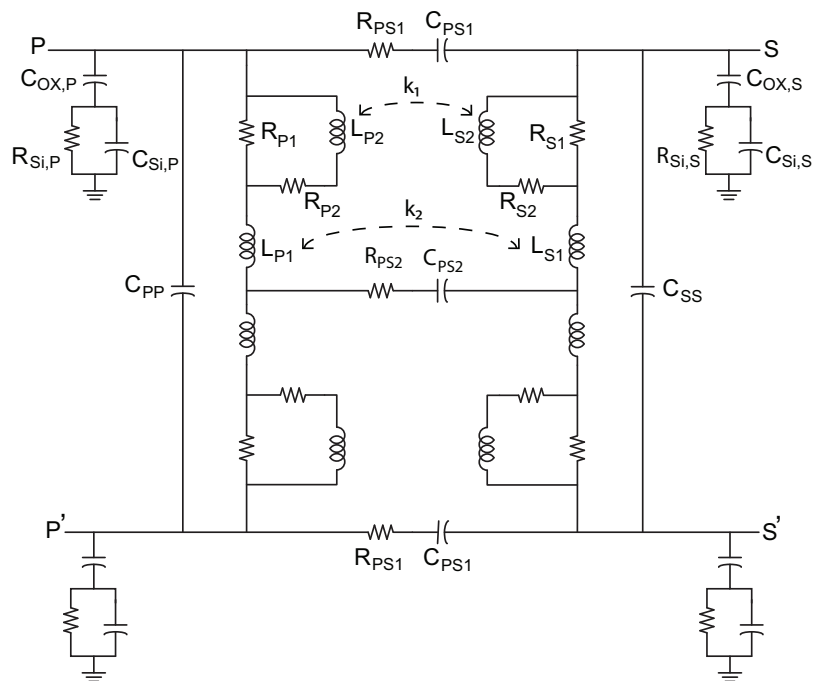
Figure 2.17: Symmetric 2π model of transformer

by [41]

$$L_{DC} = \frac{\mu N^2 d_{avg} c_1}{2}(ln(c_2/\rho) + c_3\rho + c_4\rho^2) \tag{2.23}$$

where c_1, c_2, c_3 and c_4 are equal to 1.07, 2.29, 0 and 0.19 respectively for an octagonal winding. The dc resistance (R_{DC}) can be calculated from the trace length and metal resistivity. At low frequencies, the primary/secondary winding can be represented as a single inductor with a series resistance to capture the quality factor (Q). However, as we move to higher frequencies, the loss of the winding increases due to the skin and proximity effects. In order to capture this frequency dependence, a ladder network comprising of L_{P1}, L_{P2}, R_{P1} and R_{P2} is used on the primary side and a similar network on the secondary side. In reality, a more distributed approach of modeling the skin and proximity effects using multiple ladder networks yields better results. However, to retain simplicity, we chose one section only since the error incurred by doing that was less than 10%.

From the ladder network, the dc inductance and resistance values of the primary winding can be calculated as [16]

$$L_{DC} = L_{P1} + \left(\frac{R_{P1}}{R_{P1} + R_{P2}}\right)^2 \cdot L_{P2} \tag{2.24}$$

$$R_{DC} = \frac{R_{P1}R_{P2}}{R_{P1} + R_{P2}} \tag{2.25}$$

Using Eq. 2.23-2.25, we fit the component values of the elements of the ladder network.

The substrate network consists of oxide capacitance (C_{OX}) and substrate capacitance and resistance (C_{SUB}, R_{SUB}) whose values have been calculated using process parameters and geometry of the windings. However, it is important to note that since we are utilizing a vertical transformer built out of the top two metal layers, the top winding has a different oxide capacitance than the bottom one, and in fact is shielded by the bottom winding. To account for this, asymmetric substrate networks have been used for the primary and secondary windings. Finally, to correctly capture the self-resonance frequency of the transformer, the winding self-capacitance (C_{PP} and C_{SS}) and inter-winding capacitance (C_{PS}) have been included in the model and have been calculated using parallel plate capacitance equations and adding fringing capacitance to it.

Fig. 2.18 shows a comparison of the Z and S-parameters obtained from the model (for a $42\mu m$ transformer) to the 3D electromagnetic simulation results using HFSS. From the graphs, we see that there is good agreement between the results even up to 80GHz. As a metric of the performance of the transformer, we show a comparison between simulated and modeled G_p (power gain) of the transformer (Fig. 2.18(d)). A good match between the two shows that the model used here is indeed physical.
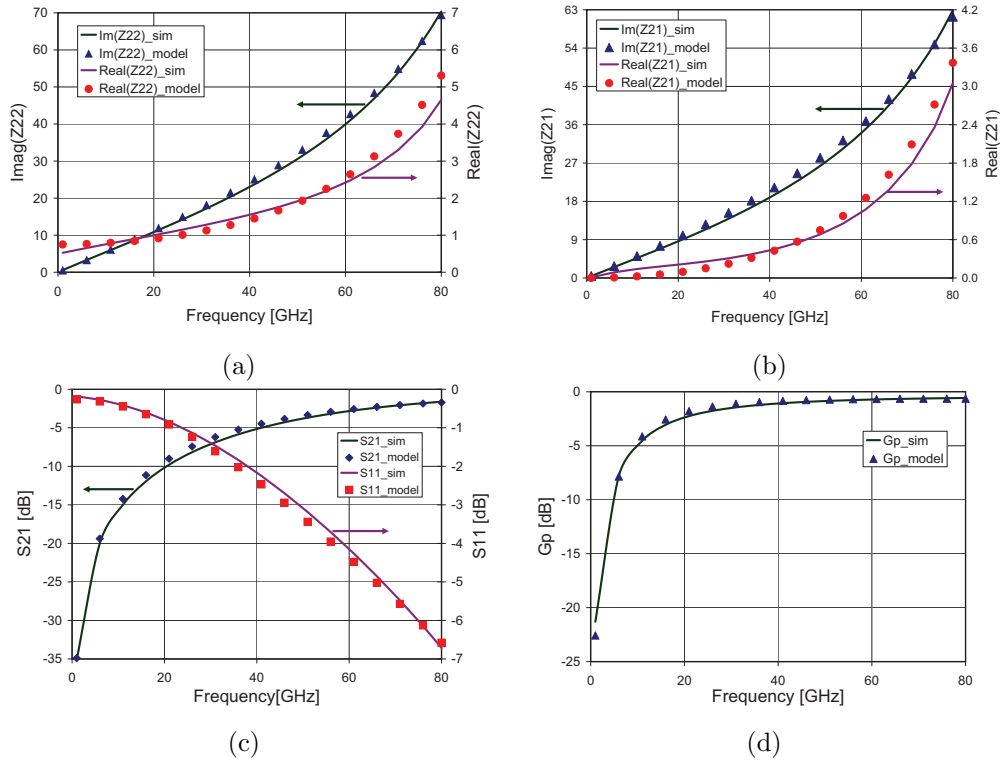
Figure 2.18: Comparison between simulated and model results a) Z_{22} b) Z_{21} c) S_{11}/S_{21} and d) G_p.

2.3.2 Distributed Modeling of Transformers

The previous section demonstrated that a lumped element approximation of an on-chip transformer is valid even at the mm-wave frequencies. However, the model developed above involves a large number of parameters, some of which can be calculated accurately from geometry, but many are fitting parameters to match the modeled results to measurements. Consequently, it is very difficult to make such a model truly scalable with transformer size (winding diameter) and hence difficult to use efficiently during the design phase. We have proposed therefore a simplified distributed model of a transformer [42], which is length-scalable and involves far fewer parameters. However, as will be demonstrated next, it has high accuracy in predicting the input impedance and insertion loss up to mm-wave frequencies.

The fundamental observation that leads to this model is that because of the presence of significant inter-winding capacitance, the energy transfer from the primary to the secondary winding is not only through magnetic field coupling (inductive), but also through electric field (capacitive). Hence, as shown in Fig. 2.19, the transformer can be viewed as a broadside-coupled differential transmission-line (t-line). The parameters needed to charac-
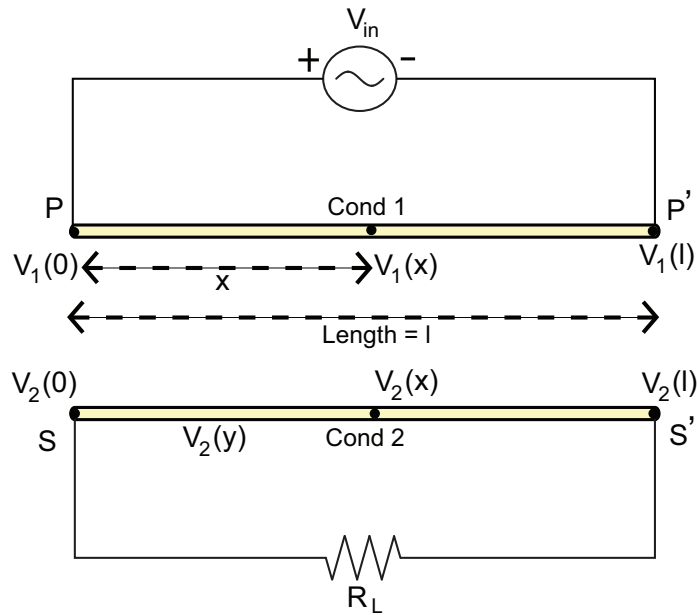
Figure 2.19: Distributed modeling of a transformer as coupled transmission-lines.

terize a transmission line are its characteristic impedance (Z_0) and its complex propagation constant γ ($\gamma = \alpha + jk$), where α is the attenuation constant and k is the propagation factor. Thus, if we are able to represent a transformer as a transmission-line, we require far fewer parameters to capture its behavior accurately [43].

Fig. 2.20 shows the simulated odd-mode characteristic impedance (Z_0) and attenuation factor (α) over frequency for a vertical transformer with a trace width of 5μm and winding diameter of 30μm. The flatness in Z_0 over the mm-wave frequency range shows that it is indeed possible to model the transformer as a transmission line.

However, in representing a transformer as a transmission line, proper attention needs to be paid to how it is excited and the resulting modes of operation [42]. Referring to Fig. 2.19, if the structure would have been used as a normal transmission line in a differential circuit, ports P and S would have been excited differentially and ports P' and S' would have been terminated in the load resistor R_L. However, in a transformer, we represent the primary winding as one conductor of the t-line and the secondary winding underneath it is the second conductor of the coupled-pair t-line. Hence, when the primary winding is excited differentially, we are applying that excitation across ports P and P', while the secondary winding (ports S and S') is terminated in R_L. This means that even though the transformer is excited differentially, with respect to the transmission line (ports P and S) both odd and even modes of propagation are excited. This is even more true since in a power amplifier, the output transformer is also utilized as a balun and hence one of the output ports is grounded. Thus, in order to capture the complete behavior, we need to take into account both the odd
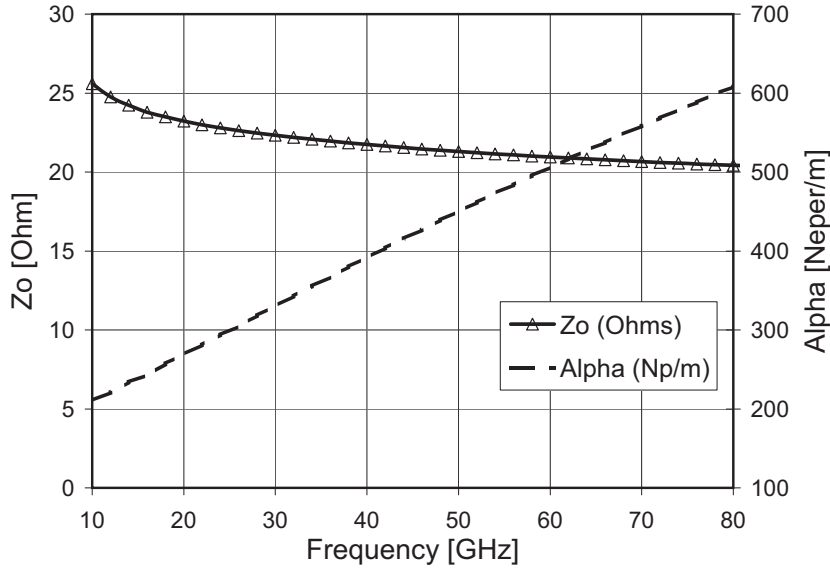
Figure 2.20: Characteristic impedance (Z_0) and loss factor (α) of a 30μm diameter transformer, excited as a transmission line.

(differential) mode and even (common) mode parameters (Z_0, α and k) of the transmission line.

A simplified set of equations that govern this modeling are shown below [44]. Here, V_{d+} and V_{d-} represent the odd-mode forward and reverse travelling waves, while V_{c+} and V_{c-} represent the same for the even mode. Z_d and Z_c are the odd-mode and even-mode characteristic impedances respectively. While Z_d is primarily dependent on the spacing between the two conductors (the dielectric thickness between the top two metal layers in our case of vertical transformer), Z_c is mostly determined by the adjacent ground structures. Given a winding geometry (octagonal shape in our design), the length of the line (l in Fig. 2.19) is determined by the winding diameter. For simplicity, the lines have been assumed as lossless ($\alpha = 0$) in these equations. The real loss has been incorporated in the final model and its frequency dependence has been captured by a simple square-root function. The parameters required for the coupled line model have been extracted from HFSS simulation.

$$V_1(x) = V_{d+}e^{-jkx} + V_{d-}e^{jkx} + V_{c+}e^{-jkx} + V_{c-}e^{jkx} \tag{2.26}$$

$$V_2(x) = -V_{d+}e^{-jkx} - V_{d-}e^{jkx} + V_{c+}e^{-jkx} + V_{c-}e^{jkx} \tag{2.27}$$

$$I_1(x) = V_{d+}e^{-jkx}/Z_d + V_{d-}e^{jkx}/(-Z_d) + V_{c+}e^{-jkx}/Z_c + V_{c-}e^{jkx}/(-Z_c) \tag{2.28}$$

$$I_2(x) = -V_{d+}e^{-jkx}/Z_d - V_{d-}e^{jkx}/(-Z_d) + V_{c+}e^{-jkx}/Z_c + V_{c-}e^{jkx}/(-Z_c) \tag{2.29}$$

Imposing current continuity and voltage boundary conditions, we can solve for the input
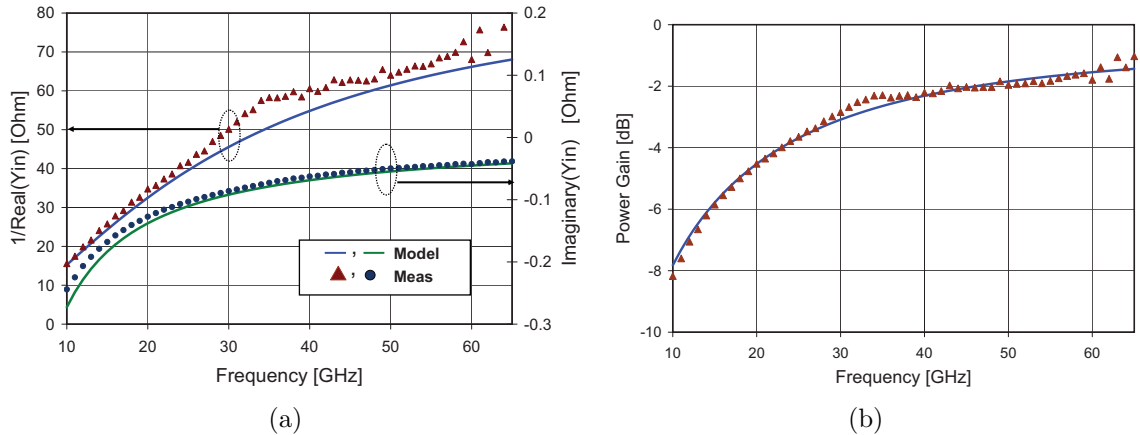
Figure 2.21: Comparison between measured and t-line based model results for a) input admittance (Y_{in}) and b) power gain (G_p).

impedance looking into the primary winding as

$$Z_{in} = (V_1(0) - V_1(l))/I_1(0) \tag{2.30}$$

Using Eq. (2.26) - (2.29) in Eq. 2.30, we get

$$Z_{in} = \left\{ \frac{\sin\left(\frac{kl}{2}\right)\left(4Z_c Z_d \tan\left(\frac{kl}{2}\right) + jR_L(Z_c + Z_d)\right)}{R_L \cos\left(\frac{kl}{2}\right) - j(Z_c + Z_d)\sin\left(\frac{kl}{2}\right)} \right\} \tag{2.31}$$

The above equations (with the odd-mode and even-mode losses) have been implemented using "CLINP" (coupled lossy transmission line) in ADS and the s-parameters predicted by the model match both measurements and full 3-D simulations over a very wide frequency range. Fig. 2.21(a) shows a comparison between the measured and modeled real and imaginary parts of input admittance of the vertical transformer with 30μm diameter, while Fig. 2.21(b) shows that the modeled insertion loss matches measurements closely. The measurements were carried out up to 65 GHz (limited by the available VNA).

A significant advantage of this model is its scalability property. If the trace width of the windings is kept fixed, the characteristic impedance (Z_o), loss coefficient (α) and propagation constant (k) all remain constant. Hence, the only parameter that needs to be altered in the model for a different size transformer is the length of the transmission line, which is directly a function of the winding diameter. Thus, we see that compared to the lumped model, this holds significant advantage, since the performance of a broad family of transformers with the same trace width can be estimated easily without any additional parameter fitting. For transformers with different trace widths, the Z_o and α change, but can be calculated analytically based on the transmission line configuration [45]. Thus, even in these cases,

the model proves to be beneficial, as only two parameters (Z_o and α) need to be made width-dependent, compared to more than 20 parameters in a lumped model.

The scalability of the model has been verified by increasing the winding diameter from 30μm (transformer in Fig. 2.21) to 100μm. As expected, the insertion loss and the s-parameters were well predicted by the model for the new transformer and the only parameter that had to be modified in the model was the t-line length (l). The distributed transformer model proves to be a valuable tool in designing optimized mm-wave circuits, as will be demonstrated in Chapter 3. It should be pointed out that the distributed modeling methodology has been applied to 1:1 transformers only. For higher turns ratio, there exists coupling between multiple primary (or secondary) windings. The extension of such a modeling approach to higher turns ratio transformers could be an interesting extension of this work.

2.4 Transformers and Transformer-Based Power Combiners at RF Frequencies

The design, layout and modeling of transformers developed in the previous section are also valid at RF frequencies. However, at 2.4GHz, the inductance required for proper operation is larger than mm-wave frequencies, which makes the transformer footprint at these frequencies larger. As a consequence, the winding parasitic resistances are higher and in high-power applications, this translates into increased insertion loss. Thus, additional techniques are required to make RF transformers efficient. Furthermore, at 60GHz, the required output power in PA circuits is in the low to moderate range (around 10 dBm). This means that huge impedance transformation ratios are not required and 1:1 transformers often suffice. However in the lower GHz wireless applications like WLAN, the required power levels are much higher ($\geq 20dBm$). This necessitates higher impedance transformation, like 2:1 or 3:1 transformers. Minimizing the insertion loss in such structures is fairly challenging. In this work, two techniques- the parallel-primary 2:1 transformer and transformer-based power combiners- have been developed that can achieve higher impedance transformation ratio with low insertion loss. Transformer based power combiners are particularly attractive since they can realize higher turns ratio by using only 1:1 transformers. The distributed modeling technique can thus be employed to analyze power combiners as well.

2.4.1 A 2.4GHz Parallel-Primary Transformer with Interwound Secondary

Fig. 2.22 shows a conventional layout of a 2:1 transformer, where the secondary winding is wound symmetrically around the primary winding. Such a layout has two distinct shortcomings. First, the width of the primary winding needs to be very large to have low series
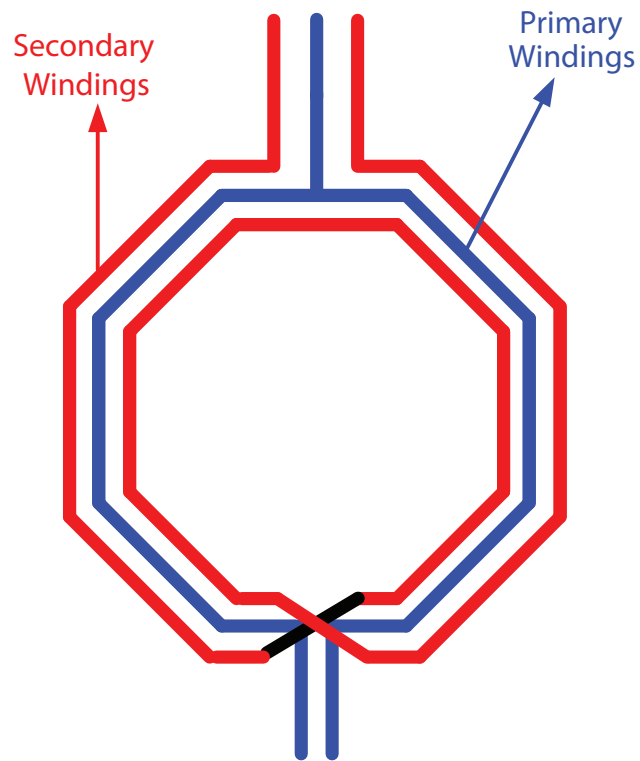
Figure 2.22: Layout of a conventional symmetric 2:1 transformer.
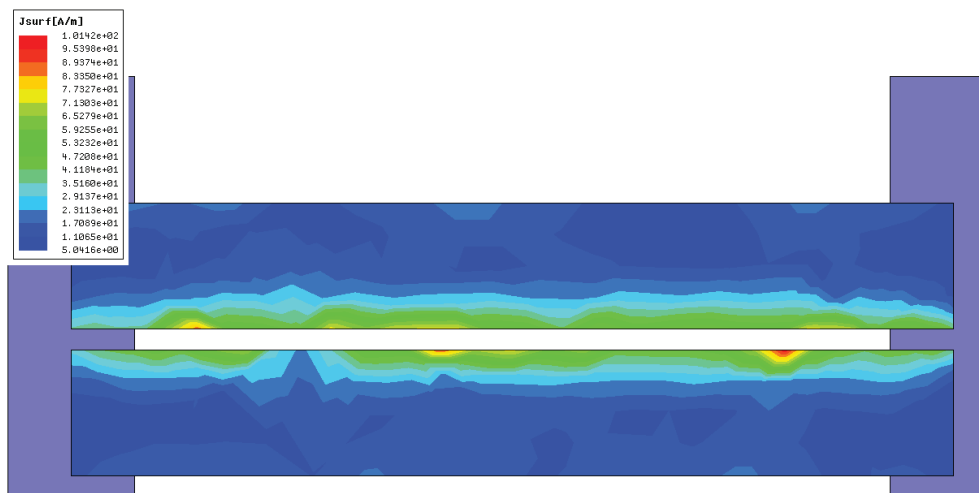
Figure 2.23: Current distribution in two parallel conductors carrying current in opposite directions.
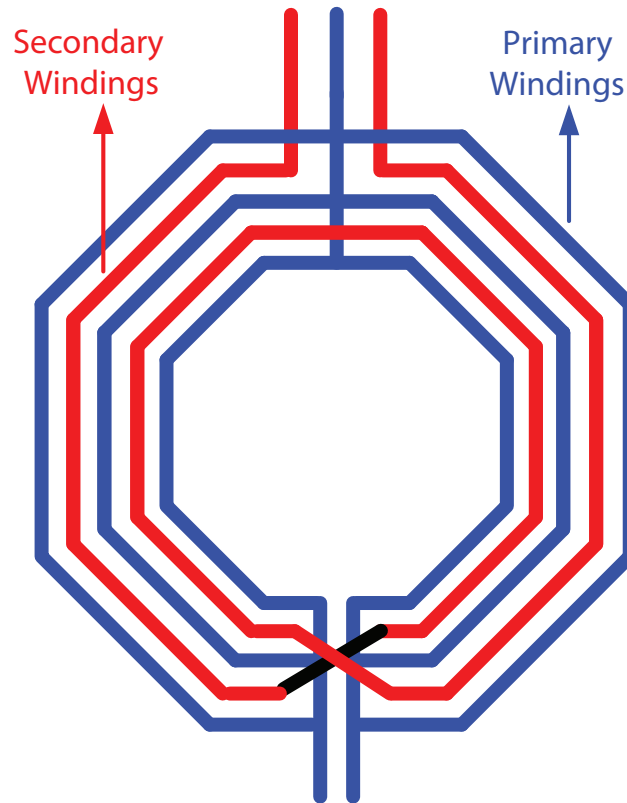
Figure 2.24: Modified layout of a 2:1 transformer with parallel primary.

resistance as well as to satisfy electromigration requirements in high-current applications like PA design. In fine-line CMOS processes, the maximum allowed width of the metal trace is limited to $10 - 12\mu m$, which is often lower than the optimal value based on loss considerations. An increase in trace width would require either a violation of the standard design rules or would require metal slotting, both of which are undesirable. The second disadvantage of the conventional layout stems from the proximity effect. At RF frequencies, when two parallel conductors carry current in opposite direction, the currents tend to get concentrated along the adjacent edges. In order to verify this, a full-wave 3D electromagnetic simulation was carried out in HFSS with two parallel conductors which were excited to carry currents in opposite directions. Fig. 2.23 shows the simulated current distribution on the top surface of the two conductors. As shown in the figure, the current density is indeed higher along the adjacent edges of the parallel conductors. Since the current is not uniformly distributed along the surface of the conductor, it increases the effective AC resistance, which leads to higher insertion loss.

The above shortcomings have been overcome in our proposed layout of parallel-primary transformer, as shown in Fig. 2.24 [46]. In this 1:2 transformer design, we have split up the
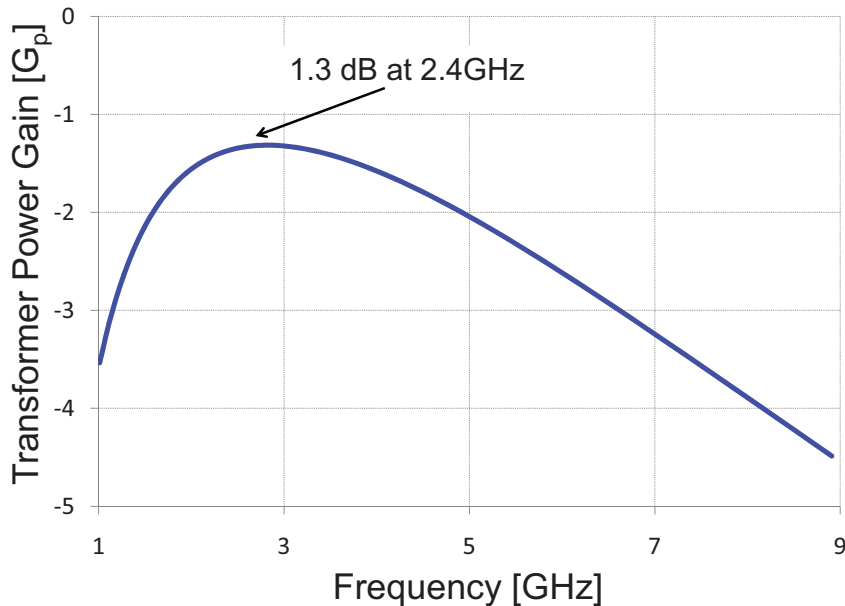
Figure 2.25: Simulated insertion loss of parallel primary transformer.

single primary turn into three parallel primary windings. Thus, the effective DC resistance of the primary winding is cut by a factor of three, compared to the resistance of the maximum width metal allowed by design rules. However, instead of just placing the three primary windings side by side, they are intertwined symmetrically around the two series-connected secondary windings, as shown in Fig. 2.24 [37] [47]. Such an arrangement reduces the loss due to proximity effect. Since each secondary winding now sees a primary winding on either side of it, the current in the winding is better distributed, cutting down the AC resistance. In the layout of this transformer, no ultra thick metal (UTM) or special RF options have been utilized. Fig. 2.25 shows the simulated insertion loss of the transformer around 2.4GHz. The transformer response is fairly broadband, with insertion loss $\leq 1.6dB$ all the way from 1.9-4GHz. The design of a digitally modulated power amplifier, employing this 2:1 transformer will be described in Chapter 5.

2.4.2 Transformer Based Power Combiner Design

In the previous subsection, we introduced an efficient 2:1 transformer, which can transform the antenna impedance to a lower value for higher power generation. While the proposed layout strategy can theoretically be extended to higher turn ratio transformers, the interwinding capacitance and other non-idealities quickly increase the insertion loss. Thus, although in theory, the efficiency of a transformer is independent of impedance transformation ratio, in reality for very high turn ratio, the primary and secondary winding quality factors reduce,
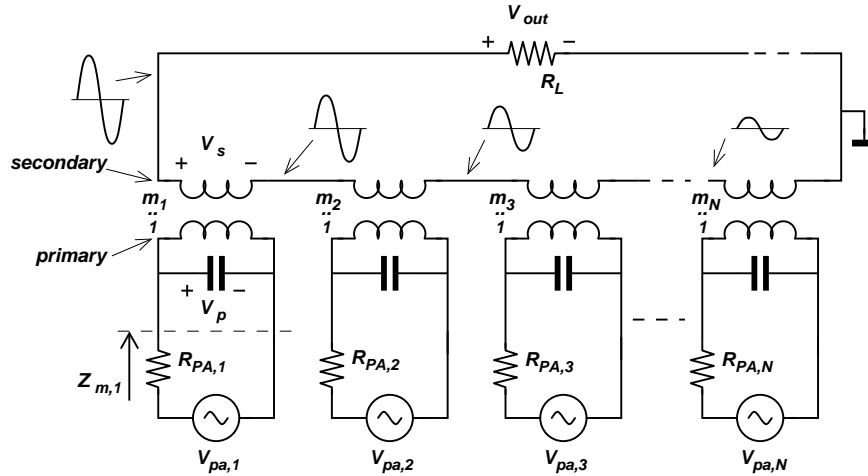
Figure 2.26: Principle of transformer-based power combining.

thereby lowering the power efficiency. In this part, we introduce the concept of transformer-based power combiners as another alternative of generating high output power in low-voltage CMOS technology.

The principle of transformer-based power combining is shown in Fig. 2.26 [7] [20]. Here, the primary windings are driven by independent amplifiers (represented in the figure by their Thevenin equivalent), while their secondaries are connected in series. Thus, the individual amplifiers may be driven by a low power supply, but the voltages add up on the secondary side, generating higher output power. In practice, mostly 1:1 transformers are used as unit elements in the power combiner design ($m_i = 1$, $\forall i$ in Fig. 2.26) to ensure less capacitive coupling and higher quality factor. Under such a condition,

$$V_{out} = V_{p1} + V_{p2} + ... + V_{pN} \tag{2.32}$$

The power delivered to the load will be the sum of the output power delivered by each amplifier minus the power dissipated in the matching network. It is thus of foremost importance to be able to design an efficient power combining network.

The power combiner, in addition to efficiently summing the ac voltages of the individual amplifiers, also performs an impedance transformation. Fig. 2.26 shows N transformers with their secondary windings connected in series. It is important to realize that the secondary winding of each transformer therefore carries the same current V_{out}/R_L. As such, all power amplifiers are coupled to each other. In other words, the impedance seen by each PA is also determined by the output voltage and output impedance of all other amplifiers. In this section, we try to derive a general mathematical expression for the impedance transformation provided by the power combiner and then simplify it under ideal assumptions [48]. To obtain an expression for the impedance Z_m, seen by each PA, one first has to find an expression for the current in the primary winding of the transformers. This current can be obtained by

using the superposition theorem. Assume that the turn ratio of transformer j is $1 : m_j$, and its primary is driven by an amplifier represented by its Thevenin equivalent voltage source $V_{pa,j}$ and output impedance $R_{PA,j}$. Then, the primary current in transformer j, when only considering voltage source $V_{pa,j}$, can be written as

$$i_{p,j,j} = \frac{m_j^2 V_{pa,j}}{R_L + \sum_{i=1}^{N} m_i^2 R_{PA,i}} \tag{2.33}$$

Likewise, the primary current in transformer j when only considering voltage source $V_{pa,k}$ can be written as

$$i_{p,j,k} = \frac{m_j m_k V_{pa,k}}{R_L + \sum_{i=1}^{N} m_i^2 R_{PA,i}} \tag{2.34}$$

Summing all these contributions, according to the superposition theorem, leads to the following expression for the current in the primary winding of transformer j.

$$i_{p,j} = m_j \frac{\sum_{i=1}^{N} m_i V_{pa,i}}{R_L + \sum_{i=1}^{N} m_i^2 R_{PA,i}} \tag{2.35}$$

Since $i_{p,j}$ can also be written as

$$i_{p,j} = \frac{V_{pa,j}}{R_{PA,j} + Z_{m,j}} \tag{2.36}$$

the transformed load impedance, $Z_{m,j}$, as seen by power amplifier j can now be written as

$$Z_{m,j} = \frac{V_{pa,j}}{i_{p,j}} - R_{PA,j} \tag{2.37}$$

$$= \frac{\left(R_L + \sum_{i=1}^{N} m_i^2 \cdot R_{PA,i} \right) \cdot V_{pa,j}}{m_j \cdot \sum_{i=1}^{N} m_i \cdot V_{pa,i}} - R_{PA,j} \tag{2.38}$$

It can clearly be seen that the transformed load impedance $Z_{m,j}$ depends on both the output impedance and output voltage (both magnitude and phase) of each PA, which can be considered as a form of *load-pull*.

When all power amplifiers have the same output impedance and generate the same output voltage, and all the transformers use the same winding ratio m, Eq. 2.37 simplifies to a resistive value of

$$R_m = \frac{R_L}{N \cdot m^2} \tag{2.39}$$

The impedance seen by each amplifier is now determined by two factors only: the turn ratio m of each transformer and the number of parallel stages N. Note that it is independent
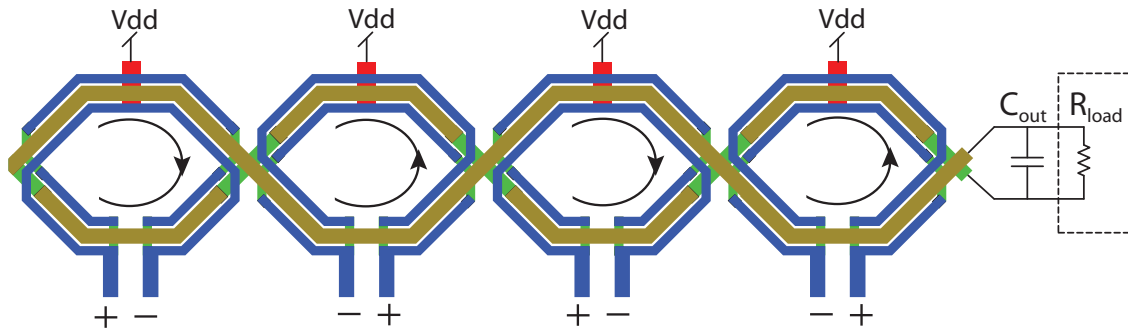
Figure 2.27: Transformer-based "figure 8" power combining network.

of output impedance of the other power amplifiers. Under the same conditions, the total output power delivered to the load equals

$$P_o = N^2 \cdot m^2 \cdot \frac{V_{pa}^2}{2R_L} \tag{2.40}$$

and the impedance transformation ratio is defined as

$$r = \frac{R_L}{R_m} = N \cdot m^2 \tag{2.41}$$

From Eq. (2.40), it can be seen that the output power can be increased either by increasing m or N. On the other hand, Eq. (2.41) shows that the impedance transformation ratio increases only linearly with N but quadratically with m. It was already mentioned that practically a high transformer turn ratio, i.e. a high m, also results in a high insertion loss [7]. Therefore, it is far more efficient to increase the number of stages N, rather than increasing m. Such an approach achieves a high output power with a rather moderate impedance transformation ratio.

"Figure 8" Power Combiner

Various layouts are possible for such transformer-based power combiners. Two efficient layouts, a "figure 8" power combiner [49] and a "dual loop"[35] power combiner will be discussed here. Their applications in PA design will be elaborated in Chapter 4.

The "figure 8" structure is shown in Fig. 2.27. Four independent 1:1 transformers are utilized to design this 4:1 power combiner. This lateral structure places both the primary and secondary windings on the same metal layer. However, an important feature of this layout is the fact that the secondary winding is implemented in *alternating orientation*, resembling an "8". This layout forces the current in two adjacent primary windings to flow in the same direction, thereby minimizing the effects of internal flux cancelation. This results in better coupling and significant efficiency improvement.
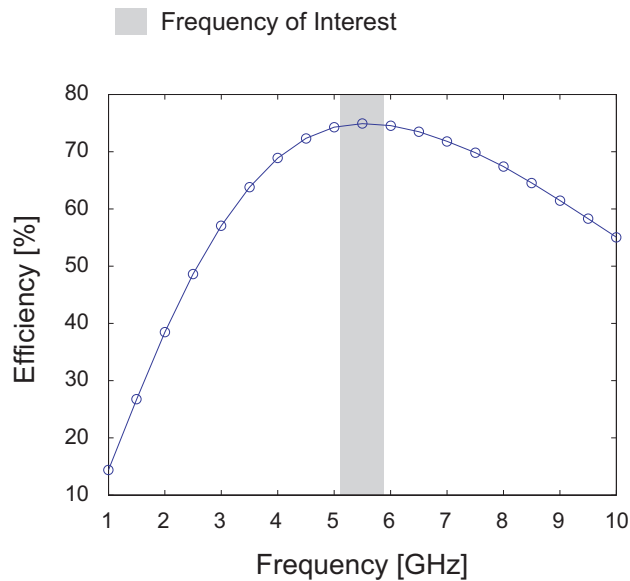
Figure 2.28: Efficiency of the power combining network in the band of interest.

An additional benefit stems from the alternating direction of the secondary windings. The secondary loop is now immune to common mode coupling from a distant source since the incoming magnetic flux induces voltages of opposite polarity across each section of the "figure 8". It is thus advantageous to employ an even number of stages.

A third important feature of the proposed structure is the use of two parallel coils for each primary, as is seen in Fig. 2.27. Because of the presence of primary windings on either side of each secondary winding, the current crowding (proximity) effect discussed earlier is mitigated, similar to the parallel primary 2:1 transformer. Now, the current is spread more uniformly in the secondary, reducing the loss substantially.

The final feature of this combiner is the minimization of eddy current losses in the surrounding metal structures. In the layout of the amplifier, it is convenient to place a ground ring around the inductor to provide low inductance paths to supply rails for the amplifier bypass capacitors. This loop lowers the quality factor of the transformer due to induced eddy currents. The eddy currents arise from the magnetic leakage due to the imperfect coupling between primary and secondary loops. By alternating the orientation of the secondary loops, the eddy currents are suppressed substantially, allowing a more practical layout to be realized. Full wave electromagnetic simulation of the power combiner was performed using Agilent Momentum. Only the top metal layer was used for the windings while the underpass elements were realized with the lower metal layers. The insertion loss of the power combiner is 1.35dB (75% efficient) as shown in Fig. 2.28 and varies by only 0.4% over the band of interest. The proposed combiner is one of the smallest reported (0.65mm × 0.15mm), even compared to a 20GHz combiner [50].
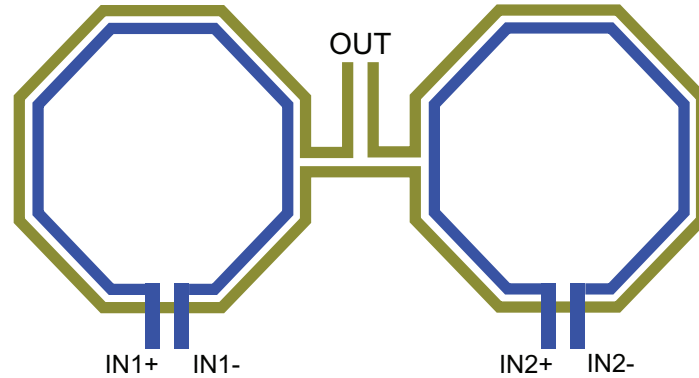
Figure 2.29: A dual-loop transformer power combiner.

Dual-loop power combiner

The "figure 8" power combiner proposed is attractive since it helps to reduce losses associated with negative mutual coupling between adjacent primary leads, thereby improving transformer efficiency. However, in the "figure 8" architecture all the primary windings are not symmetric with respect to the secondary, thereby introducing some amplitude and phase mismatch. The connection to the load impedance is not symmetric with respect to the four windings. This means that due to second order effects, the transformed load impedance will not be exactly identical for all the amplifiers on the primary side. In some applications, this imbalance can cause efficiency loss. In such cases, a dual loop power combiner can be utilized, as shown in Fig. 2.29 [35]. Although more symmetric, it does not have the benefit of alternating orientation of "figure 8". Hence the two primary coils need to be placed a certain distance apart to prevent internal flux cancelation. The insertion loss can be made comparable to the "figure 8" structure, at the expense of more area.

2.5 Summary

Low-loss low-area passive matching networks are critical in building compact, efficient power amplifiers. In this chapter, we have presented the design, layout, simulation and modeling of transformer-based matching networks. It has been shown that for higher impedance transformation ratio, transformers perform better than simple LC matching networks. Even 1:1 transformers are beneficial because they can perform differential-to-single ended conversion and offer ease of biasing. The use of transformers at mm-wave frequencies has been demonstrated and an efficient modeling method has been proposed to enable use of transformers in circuit design. Finally, the transformers have been used to build power combiners that enable high output power generation in low-voltage CMOS processes. The following chapters

will focus on power amplifier and transmitter designs using these transformers and power combiners.

Chapter 3

60GHz Transformer-Coupled Power Amplifiers

In the previous chapter, we have looked into the potential of transformers to perform impedance transformation efficiently even at mm-wave frequencies. In this chapter, we examine the design and layout of active devices at mm-wave frequencies and develop a systematic methodology for designing transformer-coupled 60GHz power amplifiers and transmitters. We begin with a simple link budget calculation to derive the required output power for short range 60GHz communication systems. This is followed by an analysis of power transistors and their characteristics at mm-wave frequencies. We then develop a detailed design methodology for transformer-PA co-design. Two design examples - a two stage PA and a high-gain three stage PA, will be studied. Finally, the integration of the power amplifier into a complete 60GHz transceiver will be described, supported by measurement results.

3.1   60GHz PA Specifications

As discussed in Chapter 1, the most promising application of the wide bandwidth available around 60GHz is in short-range high data-rate communication systems. Many applications require or benefit from such high data-rate communications- e.g., wireless USB, wireless docking stations, video download to a mobile device at a kiosk, and so on. Because of the large available bandwidth, such systems can also use relatively simple modulation schemes like OOK (on-off keying), BPSK or QPSK. These simple modulation schemes can reduce the stringent linearity requirements that low frequency wireless transmitters typically demand in order to achieve high spectral efficiency. Instead, the challenge at 60GHz is more on delivering a certain amount of power in a predictable fashion with as much efficiency and power gain as possible, while ensuring stability over process and temperature.

In order to determine the required output power in such systems and assist in device size selection in designing the transformer-coupled PA, a simple link budget calculation is

Table 3.1: Link budget analysis for 2m communication at 60 GHz

Component	Contribution	Running Total	Comment
Background noise	-174 dBm/Hz	-174 dBm/Hz noise	kT at room temp
Noise BW	93 dB	-81 dBm noise	2 GHz noise bandwidth
Noise figure	10 dB	-71 dBm noise	Typical Rx noise figure
SNR at input	16 dB		BER$< 10^{-15}$ for QPSK
Input power at receiver		-55 dBm signal	(-174+93+10+16)= -55
RX Antenna gain	6 dBi	-61 dBm signal	4-element antenna array
Path loss	74 dB	13 dBm signal	2m communication at 60 GHz
Shadowing margin	10 dB	23 dBm signal	Reflection from a wall etc.
Tx Power (EIRP)		23 dBm	
PA output power (per antenna)		11 dBm	12dB gain from directivity & spatial combining

presented in Table 3.1 [51]. This table assumes that we wish to communicate over 2m distance using a 2 GHz bandwidth (B), that the receiver noise figure (F) is 10 dB and that an antenna array provides a directional gain (D) of 6 dBi on the receive side (equivalent to a four element array). The receiver's noise floor is at -71 dBm for 2 GHz of bandwidth ($kTBF$). Assuming a required SNR of 16 dB at the receiver (for QPSK modulation for BER $< 10^{-15}$) and factoring in the antenna gain, the required power at the receiver is around -61 dBm. Accounting for the path loss and shadowing loss estimates shown in Table 3.1, we can see that an output power of the order of 10-12 dBm is a reasonable choice for the PA.

3.2 Power Transistors at mm-wave Frequencies

In addition to the lossy passive devices, a major challenge towards the realization of a fully integrated high-frequency power amplifier is dealing with the low unity power gain frequency (f_{max}) of the transistors in CMOS process. In MOSFETS, f_{max} is limited primarily by the series gate resistance and the losses in the drain and source connections [51]. Furthermore, the cut-off frequency of a transistor and its breakdown voltage are correlated: a fast transistor

(a)

(b)
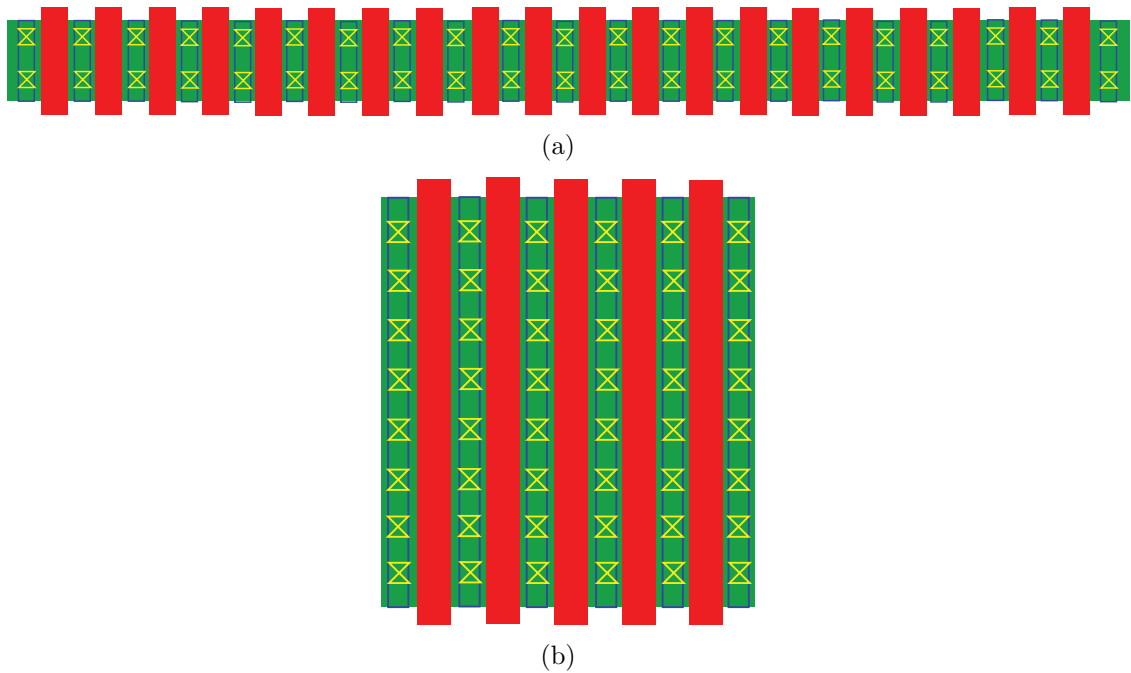
Figure 3.1: Conceptual layout of a wide transistor employing a) multiple fingers (N_F) and smaller finger width (W_F) b) smaller number of fingers and larger finger width.

can not handle high voltages, making high power generation in nanoscale CMOS processes challenging. Delivering high output power in the presence of low supply voltage requires higher output current, necessitating an increase in the total width (W) of the transistor. A larger transistor width can be achieved by either increasing the finger width (W_F), increasing the number of fingers (N_F) or a combination of both. This is conceptually illustrated in Fig. 3.1. However, optimizing either the finger width or the number of fingers leads to a power gain - output power tradeoff at mm-wave frequencies. This trade-off will be studied in this section.

In multi-finger CMOS transistors, a larger finger width results in higher gate resistance (R_g). The gate of each individual finger behaves as a distributed RC network whose total resistance can be approximated by

$$R_g = \frac{R_{poly}W_F}{3n^2L} \tag{3.1}$$

where R_{poly} is the sheet resistance of the polysilicon gate, W_F is the finger width, L is the channel length and $n = 1, 2$ depending upon the number of gate contacts. Higher gate resistance causes increased losses in the transistor, thereby reducing f_{max} and the maximum stable gain (MSG) of the transistor. The unity power gain frequency (f_{max}) can be
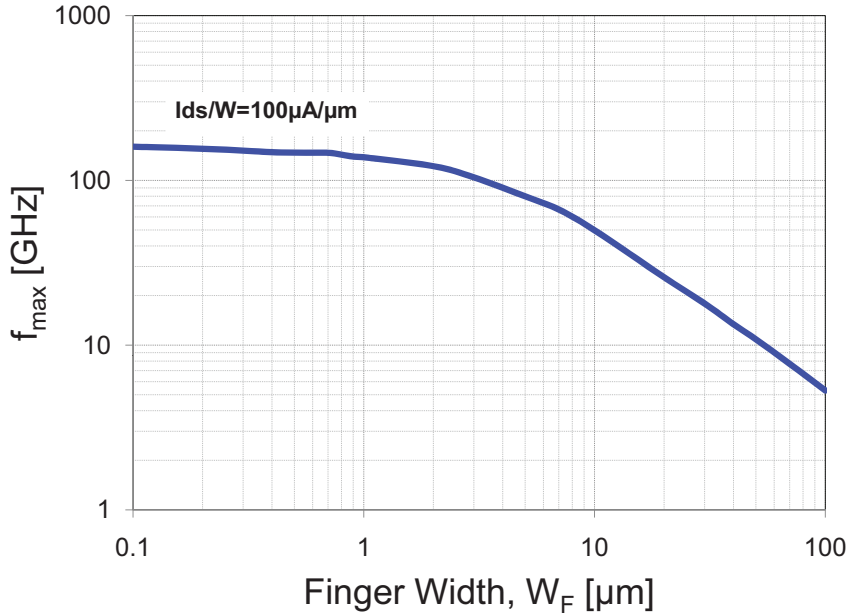
Figure 3.2: Dependence of f_{max} on device finger width at a given current density.

approximated as (neglecting substrate and drain resistance losses) [52]:

$$f_{max} = \frac{f_T}{2\sqrt{R_g(g_m\frac{C_{gd}}{C_{gg}}) + (R_g + r_{ch} + R_s)g_{ds}}} \tag{3.2}$$

where $f_T \approx \frac{g_m}{C_{gg}}$ is the unity current gain frequency, g_m is the transistor's transconductance, $C_{gg} = C_{gs} + C_{gd}$ is its total gate capacitance, g_{ds} is the transistor's output conductance, and R_g, r_{ch} and R_s are the gate, channel and source resistances respectively. Simulations confirm that for a given current density, the transistor f_{max} does decrease as its finger width is increased (Fig. 3.2) [3]. Thus, maximizing f_{max} places an upper bound on the maximum finger width of a transistor and thus on its output power capability. As shown in Fig. 3.2 that f_{max} decreases slowly up to a finger width of $1 - 2\mu m$, but has a much higher slope beyond that.

When operating at lower GHz frequencies, the number of fingers can be increased in order to build a CMOS power device that can deliver higher current. However, this is not the case at mm-wave frequencies. Increasing the total number of fingers with a fixed finger width leads to a large layout. Although the polysilicon gate resistance per finger remains fixed, large tapers are now needed to connect the gates and drains of multiple fingers, introducing additional resistive losses as well as resistive source degeneration, and hence leading to lower f_{max} and MSG.

This issue is highlighted in the layout of a 400 finger transistor layout shown in Fig. 3.3.
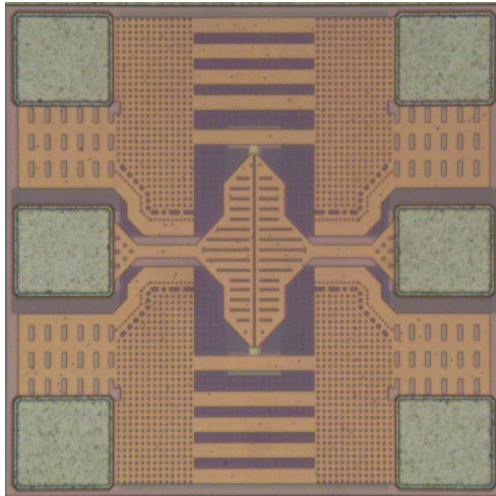
Figure 3.3: Layout of a 400-finger transistor [3].

It has been reported in literature [3] that such a 400 finger device has a gain of only 4.6dB, while the maximum gain available at 60GHz is around 8dB. In the mm-wave domain, where power gain is small, such a reduction cannot be accepted. Considering these limitations, a finger width of $1-2\mu m$ with 60-100 fingers is optimal at 60GHz in 90nm CMOS technology for PA design. Of course, with the reduction of feature size, the f_{max} and MSG increase and layout size reduces, both of which aid mm-wave circuit design.

3.3 Design Methodology for PA with Transformers

The design of the PA is a simultaneous optimization of output power, power gain and efficiency while ensuring stability over all frequencies. In order to simultaneously optimize the PA and transformers, the following design algorithm is employed. Based on output power requirements, an appropriate NMOS device size (W= $80\mu m$ in this design) is selected for the transistor in the output stage. The bias current (22mA) is chosen to ensure that the device is biased near peak f_T current density (0.3 mA/μm) [53]. The finger width ($W_F=$ 1μm) and the number of fingers ($N_F = 80$) have been chosen based on f_{max} considerations, as enumerated in the previous section.

Next, the optimum load impedance must be selected for the output transistor. The load impedance needs to ensure that the chosen device can deliver the required output power with as high efficiency and power gain as possible. To determine this, the contours for constant 1-dB compressed output power and power gain are plotted on a Smith chart, along with the load stability circle (Fig. 3.4).

In plotting the power gain contours on the Smith chart, it is important to pick the most appropriate gain metric. In the microwave community, there exists three different definitions
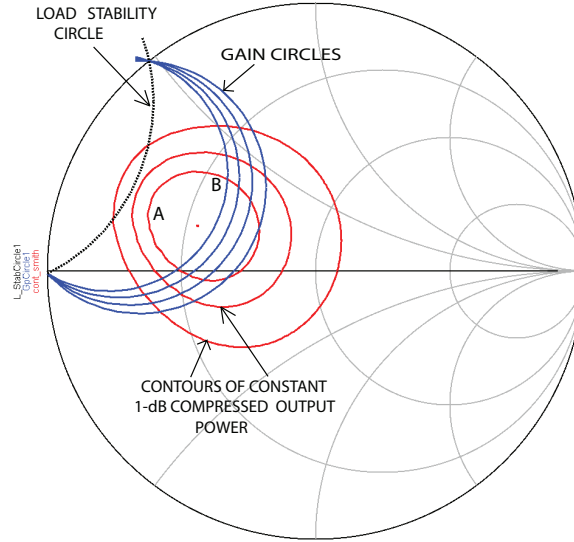
Figure 3.4: Contours of constant 1-dB compressed output power, power gain and load stability circles for an $80\mu m$ NMOS transistor.

of power gain. These are the transducer power gain (G_T), available power gain (G_A) and operating power gain or simply power gain (G_P). They are defined as follows [54] [51]:

$$G_T = \frac{P_L}{P_{AVS}} = \frac{\text{Power delivered to the load}}{\text{Power available from the source}} \tag{3.3}$$

$$G_A = \frac{P_{AVN}}{P_{AVS}} = \frac{\text{Power available from the network}}{\text{Power available from the source}} \tag{3.4}$$

$$G_P = \frac{P_L}{P_{IN}} = \frac{\text{Power delivered to the load}}{\text{Power input to the network}} \tag{3.5}$$

It can be shown that G_T, G_A and G_P become equal and reach their maxima if both the input and output are conjugately matched. However, this is possible only if the device is unconditionally stable, i.e. the stability factor (k) is greater than unity. However, the device is often not unconditionally stable, and so conjugate match may not be possible. In addition, conjugate match may not be optimum for output power and hence a PA is rarely conjugately matched at its output. Of the three gain definitions, operating power gain (G_p) is the one that does not assume conjugate match and hence it is the most suitable metric. This is why the G_p circles were plotted on the Smith chart in Fig. 3.4, along with the output power contours.

In low-frequency PAs, the optimum load impedance is often selected based on output power consideration only, since the device has enough power gain and stability can also be
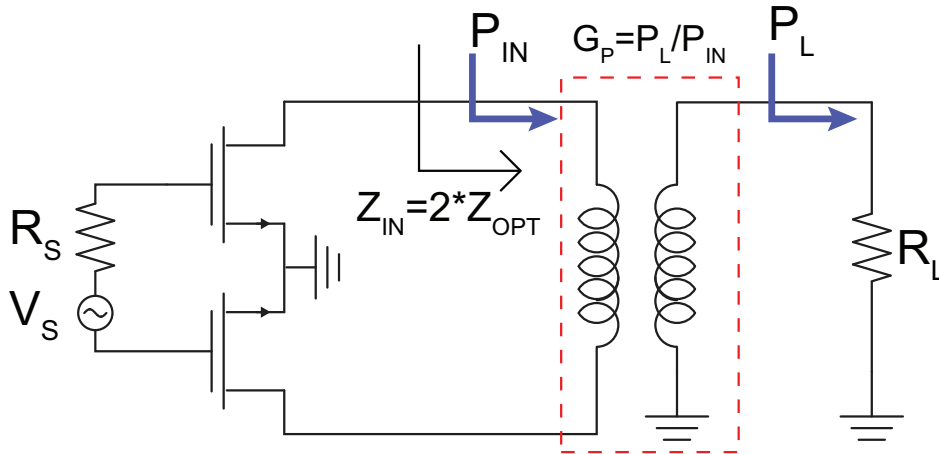
Figure 3.5: Transformer performing impedance transformation.

ensured by introducing resistive gate losses. Such a design approach is not suitable at 60GHz, where the maximum gain of the device is less than 8 dB. Hence, the goal is to choose a point with simultaneously high power and high gain, but as far as possible from the load stability circle. For example, both points A and B in Fig. 3.4 provide almost the same output power, but point A is much closer to the instability zone than point B.

Having selected an optimum impedance (Z_{OPT}), the transformer matching network now needs to be designed, as shown in Fig. 3.5. The diameter and trace width of the output transformer needs to be optimized to efficiently transform the 50Ω load impedance to the chosen value Z_{OPT} for each transistor. It should be noted that since the output stage is differential, the impedance seen by each single-ended transistor will be half the value of the transformed impedance. The pad parasitics are taken into account and pad capacitance is used to tune the transformer secondary inductance. No additional tuning capacitors are used on the secondary winding, which reduces overall loss since MOM capacitors have low quality factor at 60GHz. In choosing the optimum transformer, the distributed model developed in the previous chapter proves helpful, since it can be used to obtain the necessary size and geometry of the desired transformer without resorting to time-consuming EM simulations for every transformer. In this particular design, a 1:1 vertical transformer with a trace width of $8\mu m$ and inner diameter of $42\mu m$ was used for the output network.

Besides presenting Z_{OPT}, the transformer should also have low insertion loss. It is important to note that the minimum insertion loss (MAG) of the transformer, which assumes conjugate matching, is not the appropriate metric here. This is because the load impedance is usually fixed by the antenna or the interface to the filter (50Ω). The more appropriate metric of efficiency is the power gain (G_p, as defined earlier) of the transformer, which takes into account the fact that the load impedance is fixed to 50Ω. G_p does not depend upon the source impedance and thus does not account for the input reflection losses. This is not
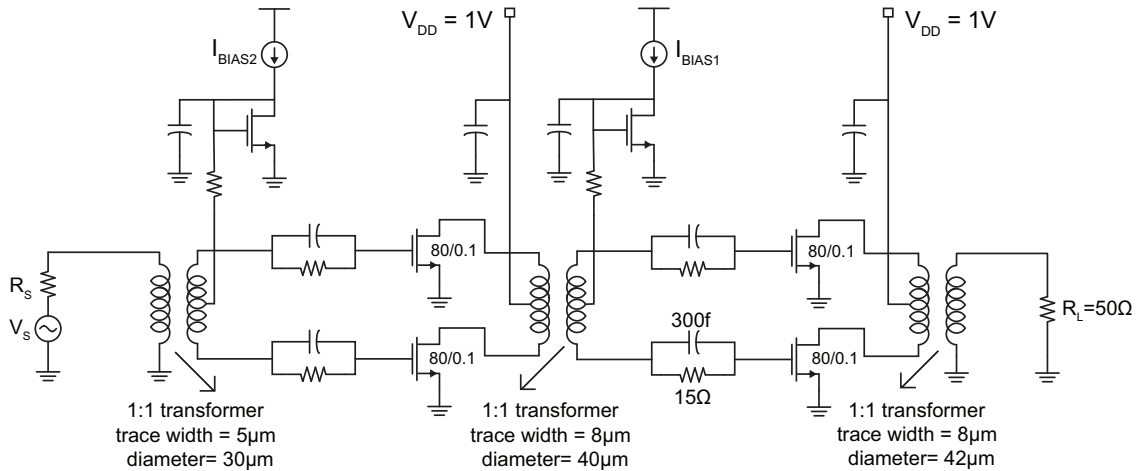
Figure 3.6: Schematic of a transformer-coupled two-stage 60 GHz power amplifier.

a problem, since the effect of that has already been captured when we chose Z_{OPT} to provide a certain power gain and output power. In fact, if we count that reflection loss in the transformer efficiency metric again, we would be double counting it. The design goal is to maximize transformer G_p, while simultaneously ensuring its input impedance (Z_{IN}) is close to $2 * Z_{OPT}$.

A similar procedure can be adopted for the design of the inter-stage transformer and driver stage. The key difference in the driver design is that the required output power level is much lower and more attention needs to be paid to power gain. Care should be taken to avoid compression in the driver stage when the output begins to saturate.

Following this systematic methodology, a two-stage transformer-coupled power amplifier has been designed in 90nm CMOS process. The implemented two-stage PA schematic is shown in Fig. 3.6 [38]. It consists of two differential amplifier stages and optimized transformers for input, inter-stage, and output matching and differential-to-single-ended conversion. Note that the use of transformers eliminates the need for AC coupling capacitors and RF chokes while differential operation reduces the amount of bypass capacitance needed. In this design, for an output power >10 dBm, two 80μm devices have been used and their output power is combined using the transformer in a differential configuration.

3.4   PA Stability Considerations

Stability is a prime concern in PA design. Both small-signal and signal-dependent stability needs to be ensured over process and temperature. As previously discussed, small signal stability is ensured by choosing a load impedance (Z_L) away from the load stability circle. Similarly, the source stability should also be checked. These two load and source stabilities can be tested together using the μ and μ' parameters. Values of μ and μ' greater than

unity at all frequencies ensures unconditional stability [54]. While the same information can be inferred from k-factor simulations, μ-factor is preferred since its numeric value gives an indication of the distance of the chosen impedance value from the instability zone. Thus higher value of μ means its farther away from the instability circle and hence more stable.

In power amplifiers, two-port s-parameter simulations are done and then either k or μ, μ' values are calculated to check stability. However, in a differential power amplifier, such a simulation only captures differential mode stability; common-mode stability is equally important, but is often neglected. In this work, a simple mathematical model is developed that can help to understand the sources of instability and how to mitigate it [48]. The method will be applied to analyze common-mode stability in a transformer-coupled power amplifier, but the analysis method is general and can be applied to analyze differential-mode stability as well.

At this point, a question may arise: if we have taken care to ensure that our load and source impedances are away from the instability zone on the Smith chart, why do we again need to perform stability simulations/analysis? The answer lies in the fact that the Smith chart based optimizations were done for the fundamental frequency (60GHz in this design) only. The output power and power gain at that frequency is of interest and hence load-pull simulations were carried out at the fundamental frequency. However, oscillations can occur at any frequency, which may or may not be even harmonically related to the carrier frequency. That is why making sure that Z_L and Z_S are safely away from load and source instability circles at 60GHz is a necessary but not sufficient test - the stability factor must be simulated at all frequencies. In this regard, the mathematical model developed is also useful, since it can predict the possible oscillation frequencies and thereby provide insights into what methods may be suitable for quenching them.

The stability analysis is carried out for a linear power amplifier with on-chip bypass capacitors, integrated transformer, and bond-wire connections [48]. The details of the analysis are presented in Appendix A. The analysis predicts that for reasonable values of bond-wire inductances and bypass capacitances there is a possibility of a low-frequency common-mode oscillation.

In addition to the possibility of common-mode oscillation, simulations also show the possibility of a low frequency differential oscillation in the two-stage PA. Fig. 3.7 shows the simulated μ factor for the 60GHz PA. Between 32-42GHz, the μ factor is less than unity, indicating possibility of instability. This happens primarily because the transistor gain increases at lower frequencies but the losses of the broadband transformer do not go up significantly. This is a pretty common situation in PA design, where there is possibility of both common-mode as well as differential mode instability. Note that $\mu < 1$ only indicates that there is a possibility of oscillation. Whether oscillation really occurs depends upon the load and source termination values. However, since the antenna impedance changes in a mobile environment, it is safer to make the PA unconditionally stable (i.e. $\mu > 1$ at all frequencies).

In order to prevent any oscillation, a dissipative component must be added to the PA.
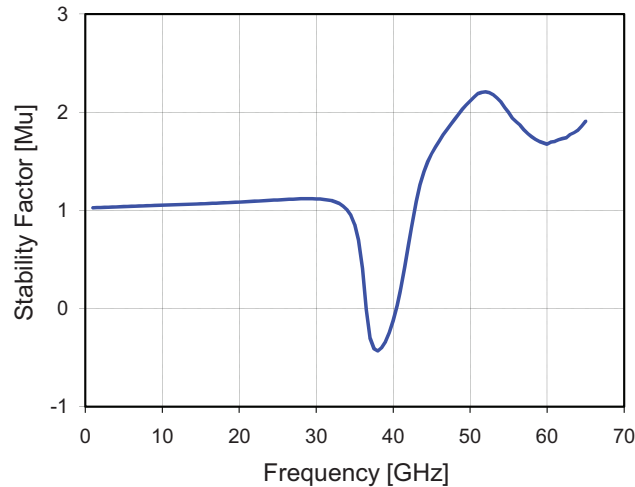
Figure 3.7: Simulated μ factor for the 60GHz differential PA.

In most designs, a simple resistor is added in series with the gate to reduce power gain and make the PA stable. However, often times like in this PA, the possible oscillation frequencies are lower than the desired carrier frequency. In such cases, an RC stabilization network [55] can be added at the gate of each transistor,, as shown in (Fig. 3.6). This network is sized to ensure resistive loss below 40 GHz, without significantly affecting the power gain at 60 GHz. Fig. 3.8 shows how the stability factor (μ) is improved by adding the RC pair. The simulated reduction in 60GHz power gain caused by this RC pair is less than 0.5dB. This shows the importance of being able to predict the oscillation frequencies systematically, since intelligent solutions can be applied which can quench the oscillation without impacting the desired performance.

3.5 Two-stage PA Simulation and Measurement Results

The transformer-coupled two-stage PA with its stabilizing network has been fabricated in a 90nm 7M-1P digital CMOS process. The chip micrograph is shown in Fig. 3.9 [38]. The small die size of 660μm x 380μm, including probe-pads, clearly demonstrates the area benefit of using transformers at mm-wave frequencies. The measured S-parameters are shown in Fig. 3.10. At a supply voltage of 1-V, the measured S_{21} at 60GHz is 5.6 dB and has a peak value of 7.7 dB at 48GHz. The power gain at 60GHz is about 1.5 dB lower than simulation and has been traced back to an inaccuracy in the model of the AC coupling capacitor used in the RC pair at the transistor gates for stability. The losses of the input and output GSG pads (1 dB) have not been de-embedded out from simulations or measurements. The 3-dB
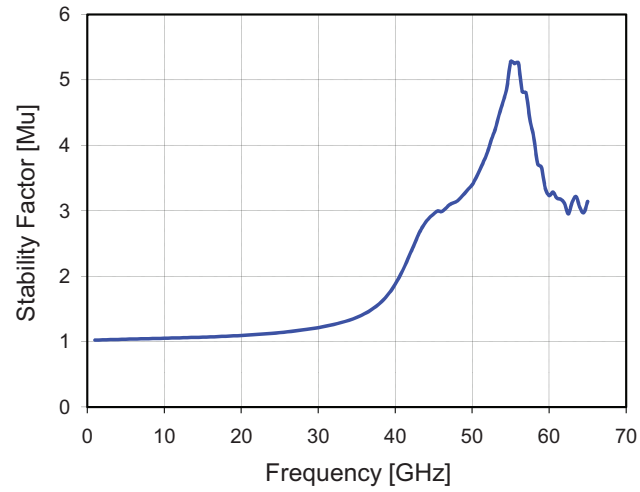
Figure 3.8: Simulated μ factor for the 60GHz differential PA with RC stabilizing network.
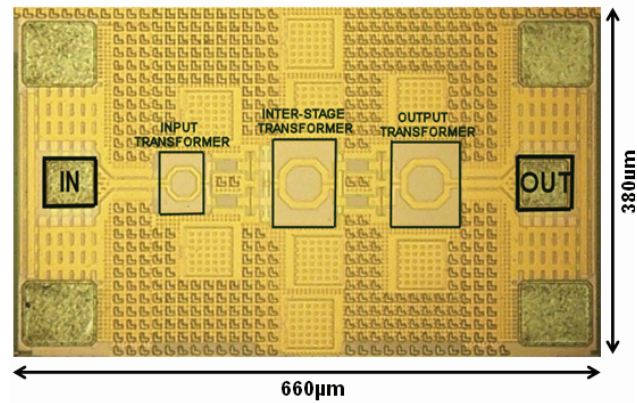
Figure 3.9: Chip micrograph of the transformer-coupled two-stage 60GHz PA.
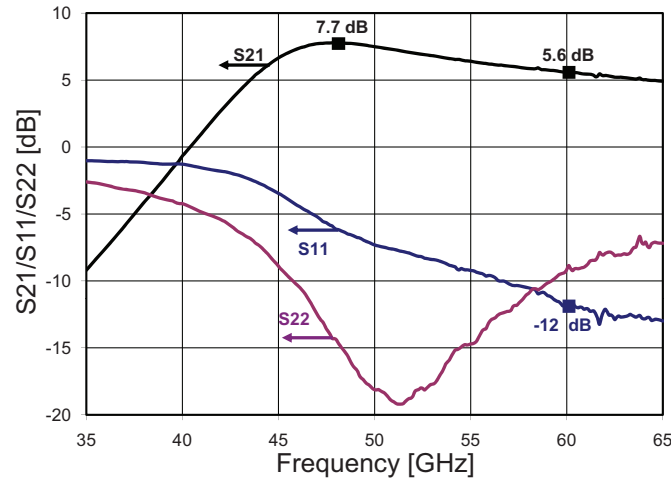
Figure 3.10: Measured small-signal performance of the two-stage 60GHz power amplifier.

bandwidth of the amplifier exceeds 22GHz (43-65GHz; upper point limited by VNA), which is one of the highest reported in literature for a mm-wave PA. The input match is better than -8 dB from 50-65GHz. The amplifier is unconditionally stable at all frequencies as indicated by the measured stability factor (k) in Fig. 3.11.

Next, single-tone output power measurements were carried out. The amplifier large signal performance at 60GHz is shown in Fig. 3.12. The input power is varied and the output power is measured using an Anritsu ML2437 power meter using the SC6230 65GHz power sensor. The linear behavior between the input/output power also indicates stability. Using a 1-V supply, the measured 1-dB compressed output power (P_{-1dB}) is 9 dBm and saturated output power is 12.3 dBm. This is one of the highest that has been reported in literature for a $90nm$ CMOS PA without power combining. This power is also comparable to some SiGe amplifiers operating from higher supply voltages [56] [57].

The efficiency measurements are shown in Fig. 3.13. The measured peak drain efficiency is 32% and peak power-added efficiency (PAE), including DC power consumption of the driver stage, is 8.8%. Most of the degradation in PAE occurs because the gain of the PA is not high. A two-tone intermodulation test has been performed on the PA to test its linearity. Two tones at frequencies of 60 GHz and 60.1 GHz have been combined using an Agilent power splitter/combiner and fed using an external amplifier to the PA input. The output signal is downconverted using an external mixer to bring to the range of the spectrum analyzer. The measured IM3 for different output power levels is shown in Fig. 3.14. It can be seen that for lower output powers the IM3 curve roughly follows a 2:1 slope, as would be expected in a linear amplifier.
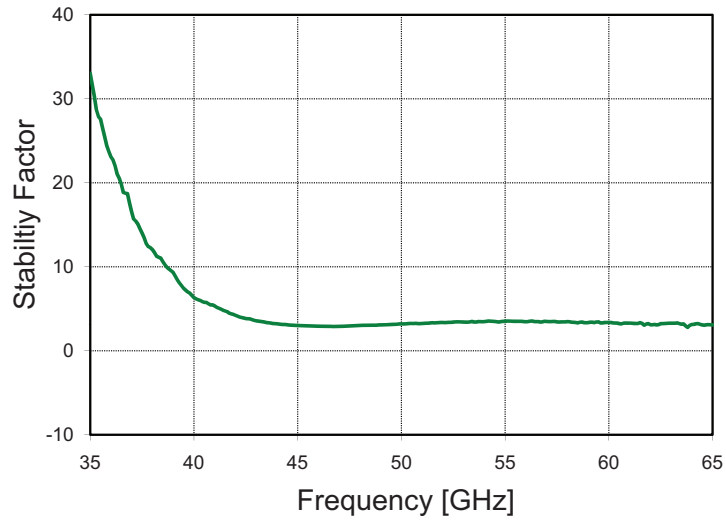
Figure 3.11: Measured small-signal stability factor of the PA.
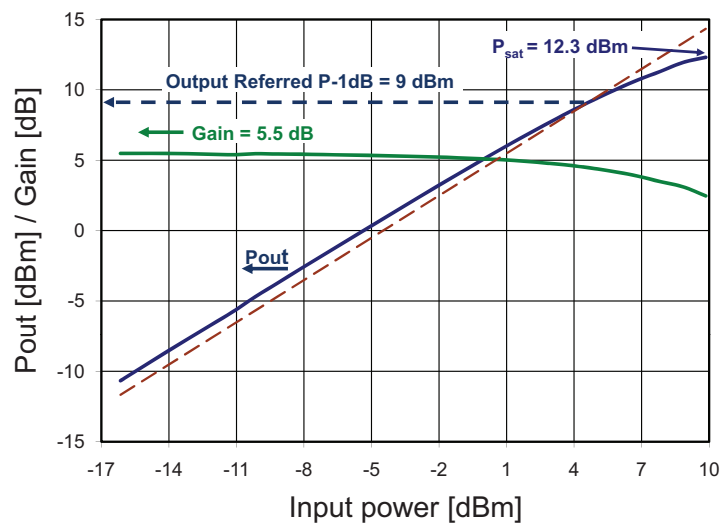
Figure 3.12: Measured gain and output power of the two-stage 60GHz power amplifier.
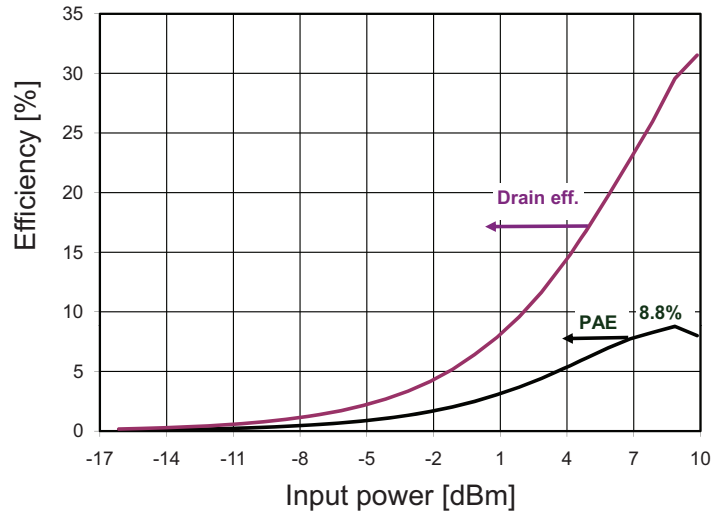
Figure 3.13: Measured efficiency of the two-stage 60GHz power amplifier.
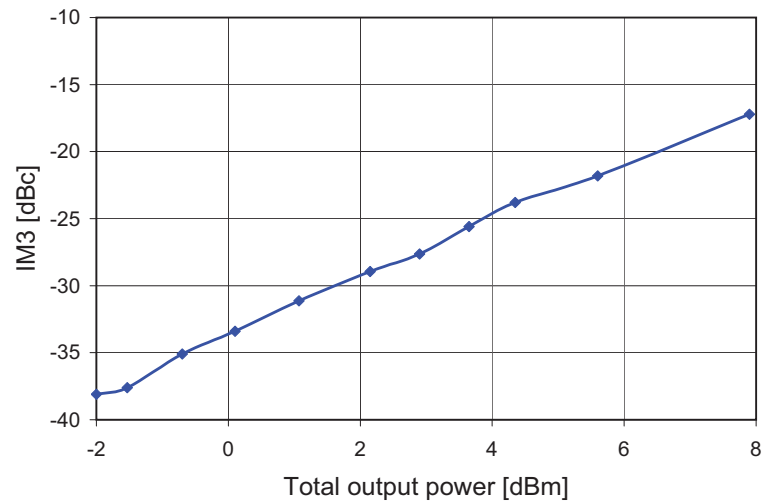
Figure 3.14: Measured IM3 as a function of output power of the two-stage PA.
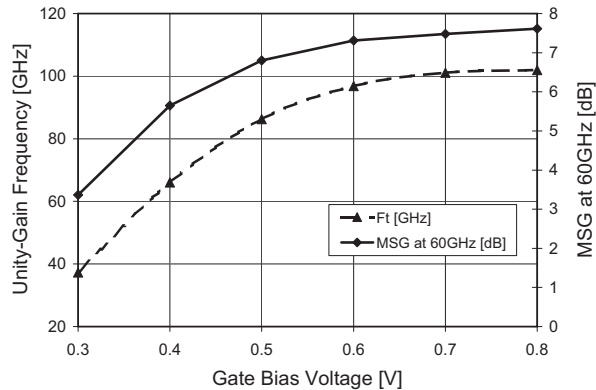
Figure 3.15: Measured f_T and maximum stable gain (MSG) at 60 GHz for an $80\mu m$ transistor.

3.5.1 Class AB Biasing at 60 GHz

Class AB mode of operation (reduced conduction angle operation) is a common method of efficiency enhancement at lower GHz frequencies. At 60GHz, the transistors are biased at around 0.3 mA/μm for optimal cut-off frequency (f_T). This translates into a gate bias voltage very close to the ideal Class A operating regime. In this experiment, we studied the effect of reducing the gate bias voltage and trying to push the output stage into reduced conduction-angle operation. Fig. 3.15 shows the measured f_T of the 80μm common-source transistor and its power gain as a function of the gate bias voltage. From the plot, it is clearly seen that at lower current-density operation, the f_T (and hence the f_{max} of the device) drops rapidly and consequently the power gain of the transistor degrades significantly. At 60GHz, where the maximum transistor gain is only 8 dB, any further reduction in the gain leads to efficiency degradation in one of two ways. If the size of the driver device is fixed, lowering the gain of the output stage forces the driver stage to enter compression together or earlier than the output stage, a phenomenon undesirable in PAs. This leads to a decrease in the total available output power. On the other hand, if the size of the driver stage is increased during design cycle to be able to provide the necessary power to the output stage even under reduced gain conditions, the current consumption of the driver stage will have an appreciable impact on the overall power-added efficiency. Thus efficiency enhancement by reducing conduction angle is not a clear win at these frequencies. The designed two-stage transformer-coupled PA has been tested at various operating points by varying the gate bias of the output stage. Fig. 3.16 shows how the 1-dB and 2-dB compressed output power is changing with bias voltage. As discussed earlier, the reduction in gain at lower bias conditions forces the driver device to compress as well, resulting in a reduction in the overall P_{-1dB} of the PA. The output power is plotted at 2-dB compression point, since for the lower bias, full saturation could not be reached due to input source power limitations. Fig. 3.17
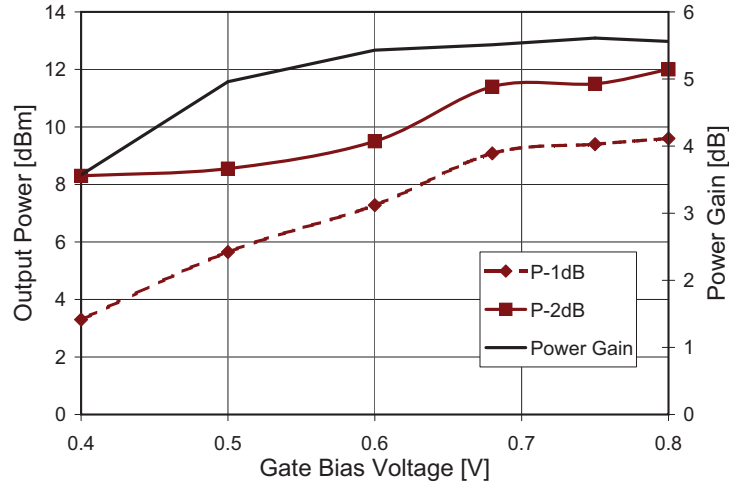
Figure 3.16: Measured output power (P_{-1dB} and P_{-2dB}) and power gain for varying gate bias of the output stage.

shows both the measured drain efficiency at different output power conditions and the PAE. It is clearly seen that at a fixed output power (9 dBm), the drain efficiency increases at lower bias points due to conventional Class AB action; but that enhancement is not reflected in either the drain efficiency at $P_{-1\text{dB}}$ point or in the PAE because of limited gain at mm-wave frequencies. Hence, gain enhancement techniques like internal unilateralization [58] or neutralization might be attractive at mm-wave frequencies to reap the real benefits of Class-AB efficiency enhancement.

3.6 Three Stage 60GHz PA Design and Integrated Transmitter

The designed transformer-coupled power amplifier has been integrated into a complete 60GHz transceiver. In order to be able to integrate such a PA with an on-chip up-conversion mixer, power gain higher than what is provided by the two-stage PA is needed. Hence, a three-stage topology using CAS-CS-CS (cascode-common source-common source) stages was utilized in this design and the schematic is shown in Fig. 3.18 [42]. Compared to the two-stage design, the size of the CS driver transistors has been reduced to 40μm (Fig. 3.18), since it has lower current consumption and higher gain than the 80μm stage. In addition to higher power gain, the load impedance necessary for maximum power and the impedance necessary for maximum gain are in close proximity for the 40μm devices, enabling almost a simultaneous power and gain match. The first driver stage has been implemented using pseudo-differential cascode transistors. At 60GHz, the 40μm cascode amplifier is uncondi-
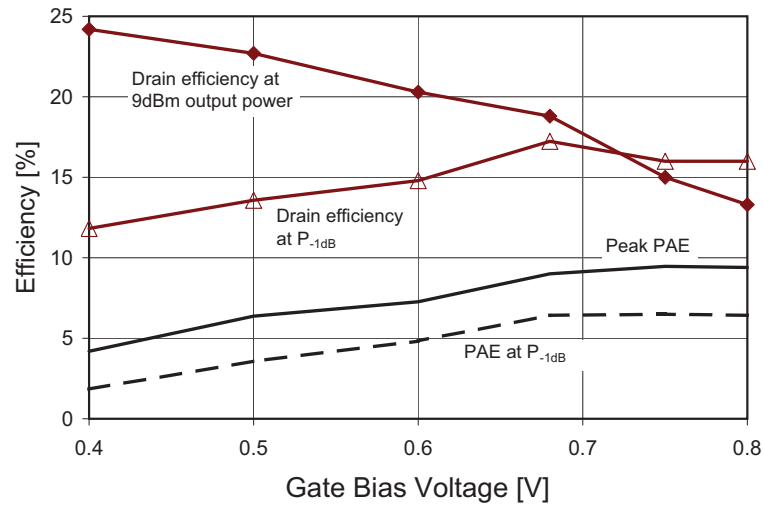
Figure 3.17: Measured drain efficiency and power-added efficiency for varying gate bias of the output stage.
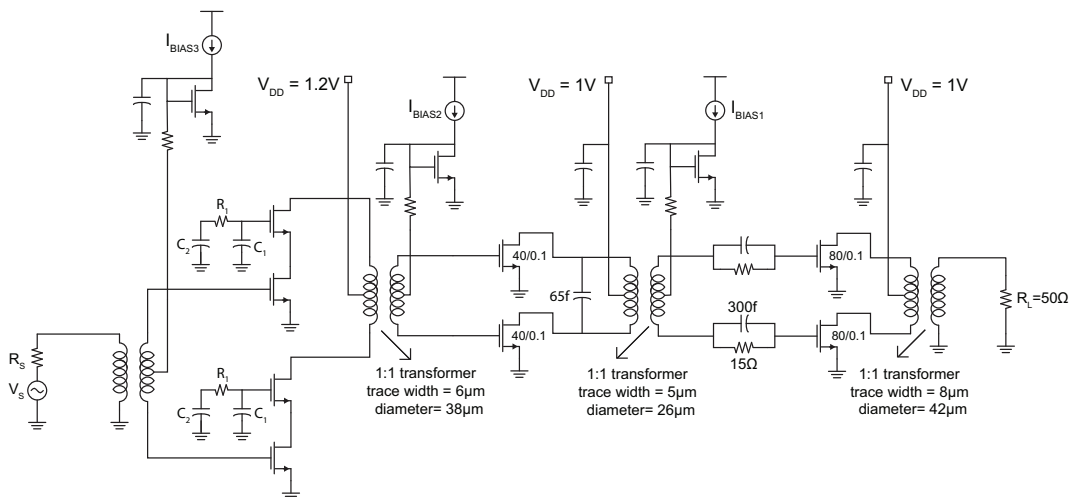
Figure 3.18: Transformer-coupled three-stage PA schematic.
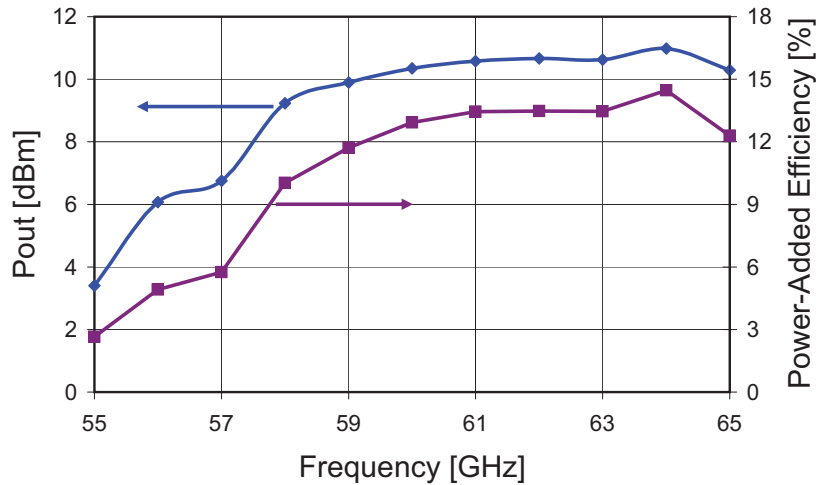
Figure 3.19: Measured output power versus frequency for the three-stage PA.

tionally stable and has slightly higher gain than its corresponding CS counterpart. This unilateral behavior improves the overall stability of the chain and also makes the input impedance of the PA (critical for the mixer) almost independent of any loading variation down the chain. The cascode transistor is operated out of a 1.2-V supply, while the common-source stages use a V_{DD} of 1V for reliability reasons.

3.6.1 Three-stage PA Measurement Results

The PA has been measured using single-tone excitation and the output power and PAE are shown in Fig. 3.19. The peak PAE is greater than 14%, which is more than 70% improvement compared to the two-stage design. This shows the importance of driver design and appropriate passive design to ensure optimum impedance at mm-wave frequencies. The PA has been tested for different supply voltages for extended time (Fig. 3.20) and showed no degradation in performance. The gain of the PA could not be directly measured using s-parameters, since it is driven by a digitally-controlled mixer on-chip. However, from measurement of the output power of the mixer in a separate test structure, the PA power gain has been inferred to be 13.8 dB, which matches simulations well. Compared to the two-stage PA, this gain enhancement comes from fixing the RC-network problem, optimizing driver stage design as well as adding an extra cascode gain stage at the beginning.

3.6.2 PA-Modulator Interface Design

The three-stage PA described above has been integrated into a complete 60GHz transceiver, as shown in Fig. 3.21 [59]. The transmitter design consists of a digital baseband pattern generator, integrated digital-to-analog (DAC)/mixer structure and the three-stage power
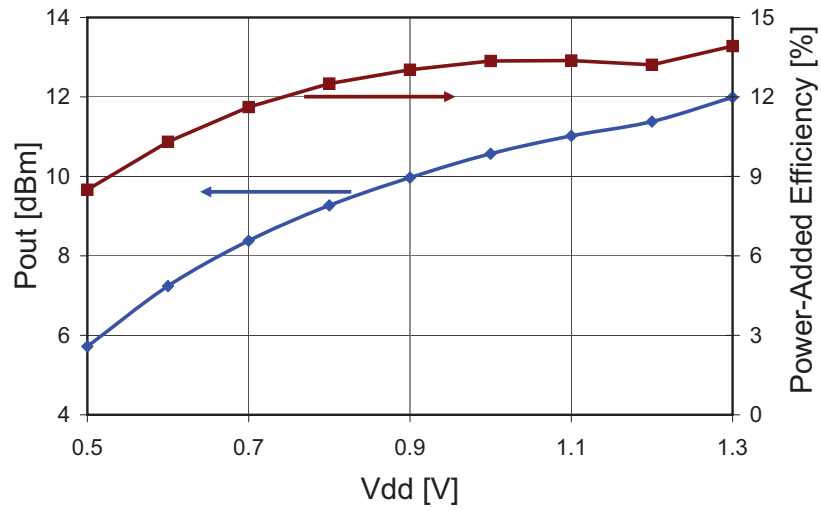
Figure 3.20: Measured PA output power for different supply voltages.
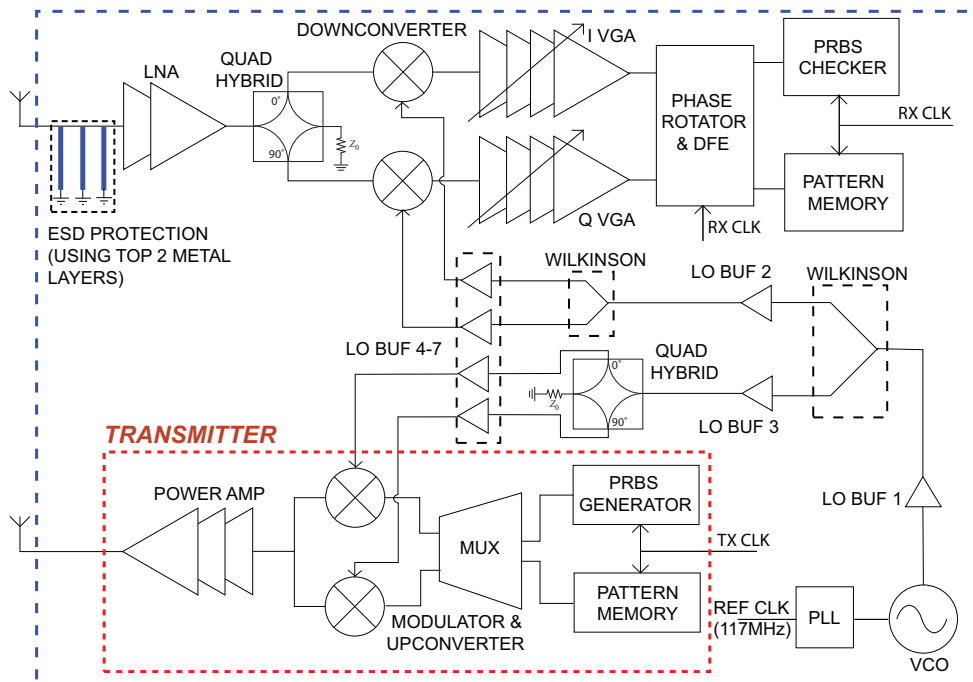
Figure 3.21: Block diagram of the 60GHz integrated transceiver.
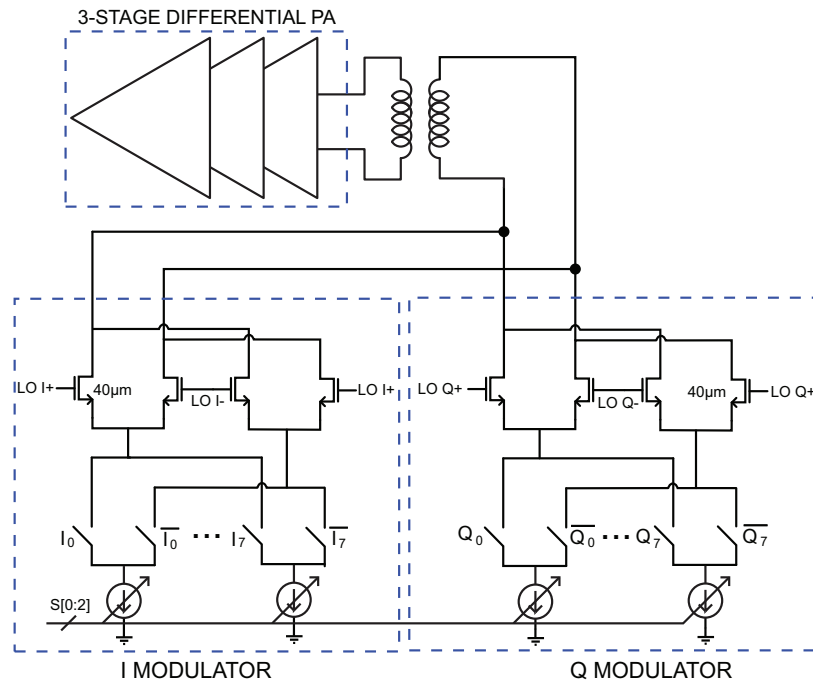
Figure 3.22: Schematic of the quadrature modulator feeding the power amplifier.

amplifier (PA). A key feature of the RF portion of the transmit section (mixer and PA) is that it employs a transformer-based architecture. Compared to conventional transmission-line based matching network design, such a design approach is more compact, reducing silicon area, with insertion loss comparable to transmission-line based designs. The digital data from the on-chip PRBS generator is fed to the modulator, which consists of a fully differential combined DAC-mixer structure (Fig. 3.22). The modulator uses a double-balanced Gilbert quad whose tail current sources are digitally switched by the input data. This structure re-uses the DAC current for the mixer and improves linearity by omitting the transconductance (g_m) stage of a conventional mixer.

In order to achieve high data rates, the transceiver employs QPSK modulation, and hence the outputs of both an in-phase (I) and a quadrature (Q) up-conversion mixer must be combined into a single RF output. A convenient way to achieve this combination is through current-mode summation, achieved by directly connecting the outputs of the two mixers as illustrated in Fig. 3.22. However, current mode summation is challenging at 60GHz due to the low output impedance of the transistors (even when they are in saturation). To achieve low-loss current summation, it would seem that a matching network is required to bring the input impedance of the PA down significantly below the output impedance of the mixer (about 100Ω in this design). However, if the transformation ratio is too high, such a matching network would have significant insertion loss. The design of this matching network therefore inherently involves a trade-off between achieving low-loss current summation while
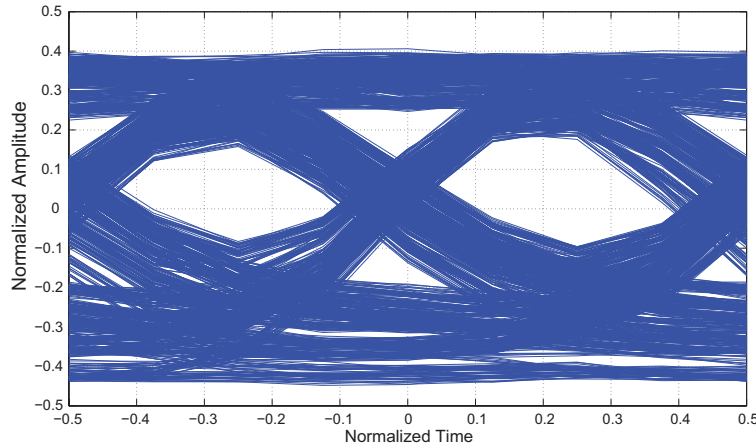
Figure 3.23: Measured eye diagram when transmitting QPSK data.

maintaining low insertion loss, and thus there is an optimum impedance that balances these two loss components and provides maximum power to the PA. Matching the input impedance of the three-stage power amplifier to this optimum value is crucial to minimize loss. This impedance transformation was achieved by the use of a transformer composed of two loop inductors stacked vertically to provide high coupling factor. However, connecting the differential outputs of the I and Q mixers to the input transformer of the PA creates appreciable lead inductances. If unaccounted for, these leads would cause unwanted impedance transformation and thereby increase combination loss. In order to avoid this, these leads have been modeled as edge-coupled differential striplines and absorbed into the impedance matching network design. The modulator is designed to provide $-2dBm$ output power and has a measured power consumption of 19.2 mW under nominal settings. The combined output from the quadrature mixer is fed to the transformer-coupled, three-stage pseudo-differential power amplifier (Fig. 3.22).

The entire transmitter chain was tested using QPSK modulated signals created by on-chip PRBS generators. The 60GHz output from the transmitter was down-converted using an external fourth-harmonic mixer and the eye diagram was observed on an oscilloscope. Fig. 3.23 shows the eye diagram when transmitting 5Gbps on the I-channel. Thus, with the same data rate being transmitted on the Q channel, this would be equivalent to 10Gbps QPSK transmission.

One of the advantages of a transformer-based design for the PA is that it also offers inherent ESD protection. At 60GHz, conventional ESD protection diodes can have significant losses, degrading the output power in the case of a transmitter, or the noise figure in the case of a receiver. However, at 60GHz, the size of transformers required is fairly small, with winding diameter mostly between $30 - 70\mu m$. ESD transients typically last on the order of a few μs (MHz frequencies). Therefore, at those low frequencies, the secondary winding of

the transformer looks like a short circuit that shunts the ESD current to ground. Since the magnetic coupling of the 60GHz transformer reduces to very low values at MHz frequencies, there is very little voltage transfer to the primary side and hence the active devices do not see the voltage or current spikes. However, the main limitation of the transformer when used as ESD protection is the transient current handling capacity of the metal traces, which in this design are $8\mu m$ wide. Nonetheless, ESD testing of the transmitter indicates that the transformer provides adequate protection against a 400V machine model (MM) event. The complete transceiver with integrated transmitter, receiver, LO generation and baseband achieved a data-rate of 4Gbps over a 1m wireless channel.

3.7   Summary

High frequency efficient power amplifier design in low-voltage CMOS technology represents a significant challenge. This work develops a systematic design methodology for high-power transformer-coupled 60GHz power amplifiers. The choice of active devices, load impedances and transformer matching networks has been explained in details. A simple yet accurate insight into stability analysis and simulations has been provided. These methods have been followed to demonstrate a 1-Volt 90nm PA with more than 12dBm of output power and 14% PAE, which is one of the highest reported in literature. This is also the first demonstration of a 60GHz CMOS power amplifier design with on-chip transformers only. The PA has been integrated into a complete direct-conversion transceiver which can transmit 4Gbps of QPSK data over 1m of wireless channel at 60GHz.

Chapter 4

Linear RF Power Amplifiers Supporting Multi-Level Modulation

The design methodology for mm-wave transformer-coupled power amplifier discussed in the previous chapter is also valid at RF frequencies. It is true that the transistors have much higher available gain at RF frequencies and thus some aspects of the design like ensuring stability become easier as some gain can be traded-off, unlike at mm-wave frequencies. The models for active devices are also well-established at these RF frequencies, making the design more predictable than at mm-wave frequencies. However, there are two aspects which require significantly more attention at RF frequencies: high output power, and high linearity requirements. The targeted range for such RF systems is of the order of tens to hundreds of meters. Hence, in contrast to mm-wave power amplifiers, such systems need to generate higher output power, often exceeding 20dBm. In addition, as described in Chapter 1, there is only a modest amount of bandwidth available at RF frequencies. Hence, to produce high data-rates, wireless systems tend to pack more bits/Hz using higher order modulation. Furthermore, orthogonal frequency division multiplexing (OFDM) is frequently used in such wireless standards to counteract the channel imperfections. Combined, these increase the peak-to-average ratio (PAPR) of the signals significantly. Processing such high dynamic range signals efficiently while preserving signal integrity places a significant burden on the PA.

Higher impedance transformation is required to achieve high output power in low-voltage CMOS technology. The concept of transformer-based power combining has been outlined in Chapter 2. The distributed active transformer (DAT) was proposed in [7] to create an efficient power combining structure. Recent work [20] [49] used modified versions of the DAT in which individual power amplifiers can be turned off in low-power mode. Nevertheless, significant attention has not been paid to achieving watt-level output power with very high linearity. In this chapter, we therefore examine the design of a very high output-power transformer-coupled RF PA with stringent linearity requirements. A power combiner architetcure, similar to [20] in concept but using a modified layout for enhanced linearity,
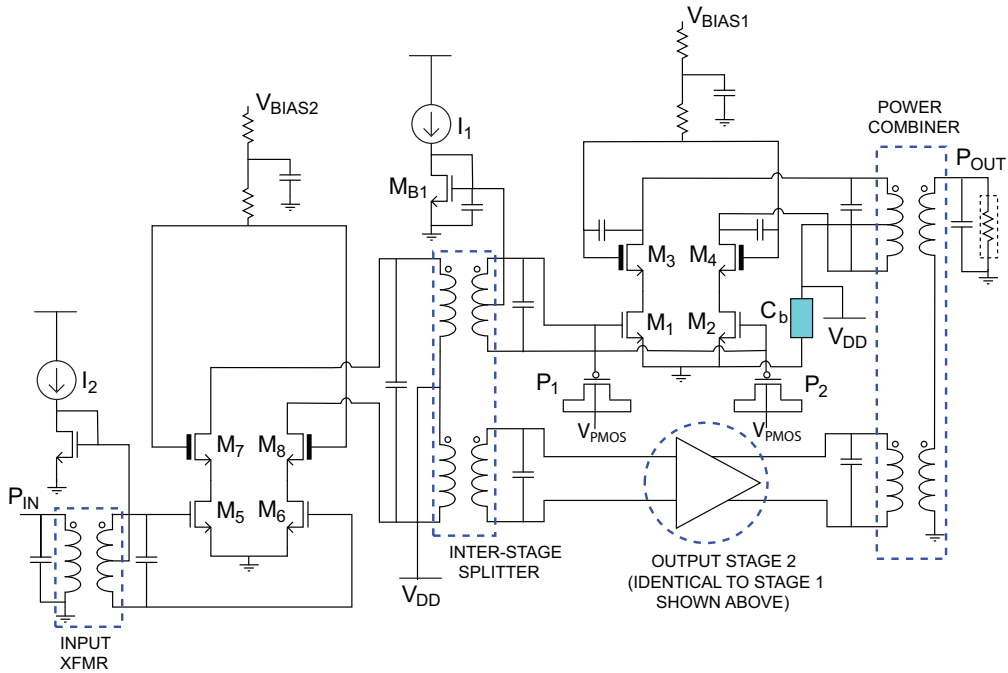
Figure 4.1: Schematic of the two-stage transformer-coupled 2.4GHz PA.

will be utilized. We will analyze which linearity metrics affect signal integrity in terms of output mask requirements and EVM, and how linearity can be improved without necessarily sacrificing efficiency.

4.1 Two-Stage High-Power Highly-Linear PA Design

The PA in this work is a two-stage all transformer-coupled design, as shown in Fig. 4.1 [60]. An input transformer converts the single-ended signal into differential form, while performing the input matching as well. A forked architecture is employed, whereby a single driver stage is utilized and its output is fed to the two output cores using a transformer-based signal splitter. It splits the driver output voltage into two, while simultaneously performing the necessary inter-stage matching as well. The two output stages are combined using a transformer-based power combiner.

Transistors M_1-M_4 (in Fig. 4.1) form the output stage. The two amplifiers driving the two primary windings of the power combiner are identical, and only one is elaborated in Fig. 4.1. When delivering saturated output power in excess of 1W, the drain voltages of M_3/M_4 can swing very high. Hence, a cascode configuration is adopted, using thin-oxide ($90nm$) transistor for the common-source device, and high-voltage thick-oxide transistor ($0.35\mu m$) for the common-gate device. In order to ensure reliability (breakdown prevention as well
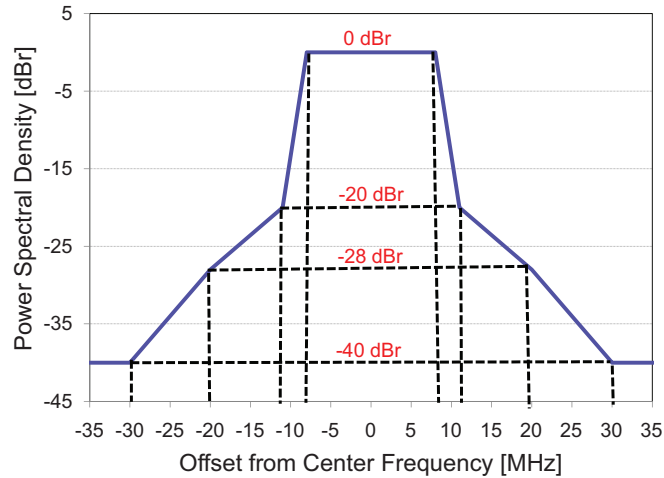
Figure 4.2: Spectral mask specified by the 802.11g standard.

as hot carrier effect mitigation) when delivering saturated output power, a dynamic bias network is employed on the cascode gate [61]. However, compared to a traditional resistive feedback network, a capacitive feedback network is used. This allows the gate of the cascode transistors M_3/M_4 to be biased independent of the supply voltage. This is critical, since a high DC voltage at the gates of M_3/M_4 will translate into higher potential at the drain of the thin-oxide devices (M_1/M_2), possibly stressing them. In addition, the feedback capacitors can be conveniently absorbed into the tuning capacitors needed on the transformer primary.

It is worthwhile to note that differential operation provides benefit in relation to this dynamic gate biasing scheme. Due to the differential operation, under small-signal excitation when the drain voltages of M_3 and M_4 are equal in magnitude and opposite in phase, their gate voltages do not move (i.e., the gates are at a virtual ground). Hence there is no significant impact on the small-signal gain of the amplifier due to this dynamic gate bias. However, as the drain swing increases and the drain voltages become asymmetric due to Class AB operation, the cascode gate voltage starts tracking the drain voltage, preventing excess voltage across the gate oxide. Extended simulations showed that all voltages remain safely below the reliability limits, even across PVT variations.

4.1.1 PA Linearity Issues

It has been already emphasized that PA linearity is one of the most important requirements for a wireless system that employs complex modulation like WLAN, WiMAX, LTE, etc. All radio systems are required to induce the minimum possible interference to other users. Hence, they must keep their transmissions within the bandwidth allocated and maintain negligible energy leakage outside the band.

PA non-linearity causes spectral regrowth and adjacent channel power leakage. Wireless standards generally put an upper limit to the allowable leakage by specifying a spectral mask. As an example, Fig. 4.2 shows the spectral mask specified in the 802.11g WLAN standard. The graph illustrates the maximum allowed power density, normalized to the power density within the signal band, as a function of the frequency offset from the carrier.

In addition to spectral mask, the PA must be linear enough to preserve the in-band signal information. This is usually guaranteed by meeting a certain EVM (error vector magnitude) specification. In digital communications, transmitted signals are often coded into a constellation in the in-phase (I)/quadrature (Q) plane. However, non-idealities of the transmitter, including nonlinearity in the power amplifier, cause the constellation points in the output to deviate from their original locations. Defining an error vector as a vector in the I/Q plane from the ideal constellation point to the actual point, the error vector magnitude (EVM) is thus the ratio of the root mean square (RMS) power of the error vector to the RMS signal power.

The EVM required by different wireless communication standards depends on the density and shape of the constellation. Error-correcting codes can compensate for some of the received errors due to RX non-idealities, and thus the required EVM also depends on the code rate and the targeted data rate. As an example, WiMAX requires a transmitter EVM better than -23dB while transmitting 16-QAM data.

Amplitude and phase non-linearity are the two most important issues in implementing a PA with stringent EVM and spectral mask requirements. Amplitude linearity- also referred to as the AM-AM response- simply means that the PA output swing increases linearly with the input. Conventional wisdom in linear PA design is to use a class A amplifier in which the transistor is always on. This implementation relies on the active device being a linear transconductor and backing off from the maximum output swing until the linearity requirements are met. However, in nanoscale CMOS technology, the output impedance of the transistors is relatively row - i.e., $g_m r_o = 5 - 10$. Furthermore, as the output swing of the PA increases, the transistor output impedance drops further, leading to gain compression even relatively far away from the peak output power. An amplifier with such a soft compressive gain characteristic, as shown in Fig. 4.3 and typical in a Class-A PA, needs a large back-off (often exceeding $10dB$) from the maximum output voltage swing in order to meet its linearity requirements. This often yields very poor efficiency - less than 5% in some cases where modulated signal with high peak-to-average ratio is used.

Thus, although the common belief is that Class A is the best for linearity since the transistor never shuts off, due to the finite and voltage-dependent output impedance, this is not really the case in nanoscale CMOS.

In the presence of a maximum achievable output swing (limited by V_{DD}), the most optimal AM-AM response for EVM and spectral purity is a brick-wall in gain vs. input power (P_{in}). This means that as the output power increases, the PA gain remains constant up to P_{sat} and then the gain drops sharply to zero. A PA with this "hard clipped" behavior requires the minimum back-off from peak power to meet EVM requirements. Of course, such a brick-wall
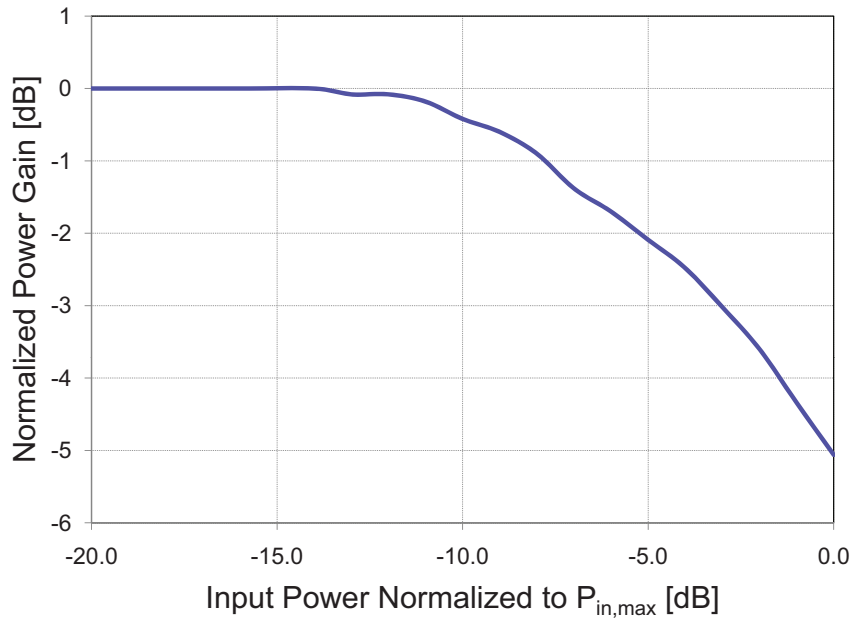
Figure 4.3: Soft compressive PA gain response.

characteristic cannot be produced by a real amplifier. However, if we can keep the ripple on the gain (both droop and overshoot) to about $0.5dB$ up to the required output power, then the EVM requirements can be satisfied with almost the same back-off as an ideal clipper. In order to synthesize such a response, some sort of gain expansion is needed to counteract the gain compression of Class-A amplifiers. In this context, Class B power amplifiers that operate by increasing the average drain current when the input power increases- therefore increasing g_m and causing gain expansion- are attractive. Thus, if the gate bias voltage is chosen so as to balance the gain compression and expansion within $0.5dB$, then a nearly-flat gain response can be synthesized (as confirmed in Fig. 4.4).

Note that Class-AB amplifiers are well-known in literature, but are typically thought of as being a compromise between efficiency and linearity. However, by choosing the optimal bias point, an appropriate AM-AM response can be engineered, which will ensure linearity (i.e. required back-off) better than a Class-A amplifier in terms of EVM.

Although a Class-AB amplifier with an optimal bias voltage can be engineered to have near-optimal AM-AM response, the amplifier has more phase distortion- or more precisely amplitude-dependent phase (AM-PM) distortion- than a Class A PA. This AM-PM distortion can also be a major contributor to EVM.

One of the main contributors to AM-PM distortion is the non-linear gate capacitance of the output transistor. Shown in Fig. 4.5 are plots of the simulated NMOS device gate capacitances as a function of gate-source voltage. It is clearly seen that as the device transitions from the "off" (below threshold) state to the "on" (above threshold) state, the C_{gs}
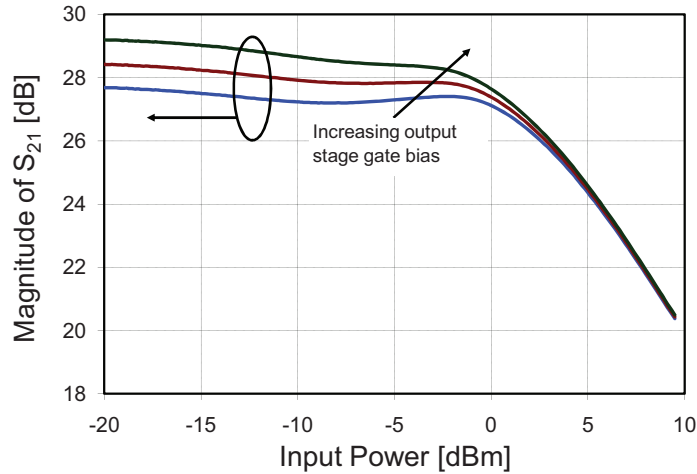
Figure 4.4: Measured AM-AM response for different gate bias voltages at 2.4GHz.

varies significantly. While C_{gs} as plotted includes both the intrinsic and extrinsic contributions, most of the variations happen because of the change in intrinsic capacitance with bias. This variation is particularly germane for Class-AB operation, because the transition in the capacitance occurs at the device's threshold voltage, close to where it is typically biased.

To reduce this AM-PM distortion, a PMOS based capacitive compensation technique has been used in this design [62]. Plotted in Fig. 4.5 is the variation in gate capacitance of a PMOS transistor with changing gate-source voltage. As can be seen from the plot, the variation is complementary to that of the NMOS device. Therefore, it is possible to compensate or partially cancel the non-linearity of the NMOS C_{gs} with the aid of the PMOS device.

In our design, since fixed capacitors were anyhow needed to tune the winding inductance of the inter-stage transformer connected to the NMOS gate, some part of that fixed capacitance has been replaced by PMOS capacitance to compensate for phase distortion. The size and bias voltage of the PMOS devices (P_1/P_2 in Fig. 4.1) has been chosen to cancel the non-linear variation in C_{gs} of the NMOS transistors. The net effective gate capacitance is shown in Fig. 4.5 and shows much lesser percentage variation over the same range of gate-source drive. Note that although there is an optimum PMOS bias voltage for ideal compensation, the improvement in phase linearity obtained is not very sensitive to an accurate value of that voltage, making such a compensation scheme robust.

Fig. 4.6 shows the measured phase variation at the PA output at 2.4GHz for increasing input drive and for different bias conditions. Although pushing the devices more into the Class-AB region increases the phase distortion, this compensation scheme keeps the variation in the output phase within 5° up to the targeted average power for all bias voltages (Fig. 4.6). The outlined methodolgy thus allows synthesis of optimal AM-AM response without increased AM-PM distortion.
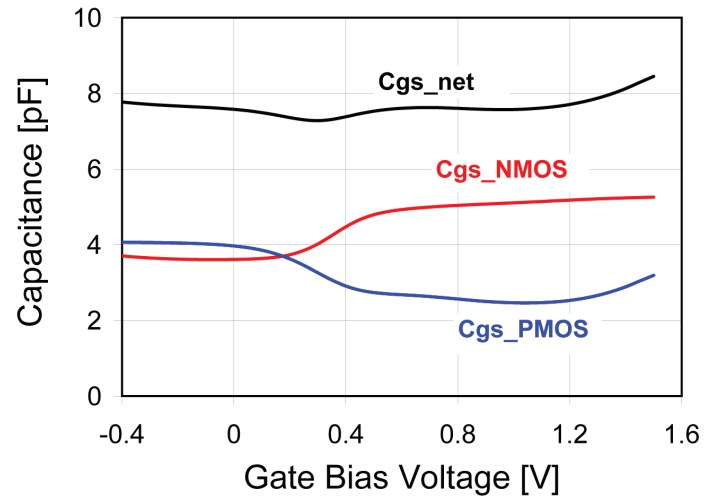
Figure 4.5: Simulated gate capacitance of NMOS, PMOS and total sum device.
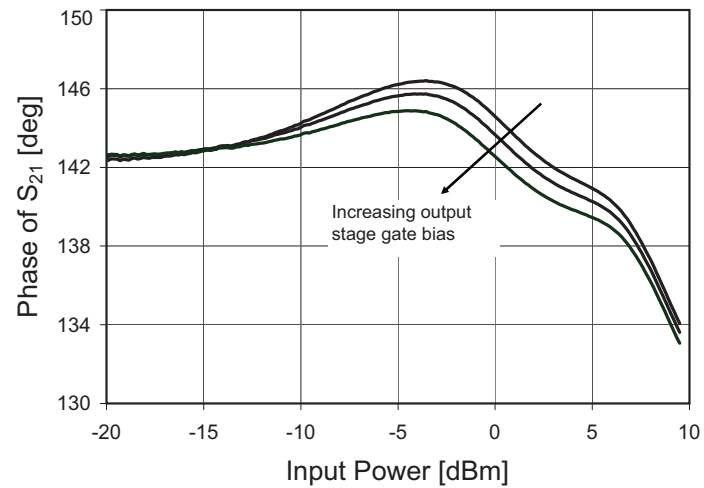
Figure 4.6: Measured variation in phase of S_{21} of the PA as a function of input power at 2.4GHz with PMOS-based compensation scheme.
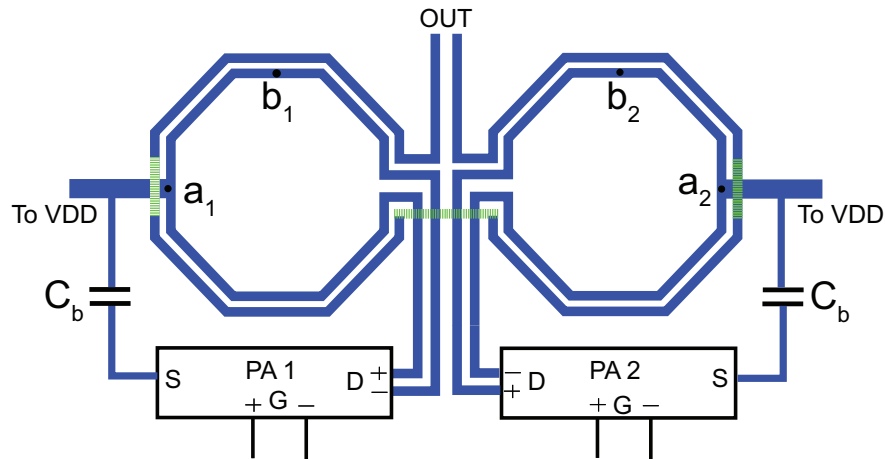
Figure 4.7: Modified layout of a 2:1 power combiner.

4.1.2 Modified Transformer-Based Power Combiner Layout

In the previous sub-section, we dealt with the issue of designing a highly linear power amplifier. In order to also generate a high output power, a 2:1 transformer based power combiner has been employed. Such power combiners have been described in details in Chapter 2. However, for better linearity, a layout modification (Fig. 4.7) is proposed in this power combiner. A notable feature of this layout is the orientation of the windings. Compared to a standard power combiner layout [20], the primary windings have been rotated 90°. In a standard layout, the center-tap would have been taken from the node labeled b_1 (or b_2). However, in this rotated layout, the center-tap node will be a_1/a_2. When the device is aligned as in Fig. 4.7, a low-inductance bypass path (C_b) between center-tap (a_1/a_2) and the source (S) of the differential devices is possible. This was beneficial because higher lead inductance with the bypass capacitance causes significant second harmonic ripple at the center-tap node at higher powers, degrading the amplifier linearity and hence the error-vector magnitude. The proposed power combiner layout helps to alleviate this issue by keeping the lead length associated with the bypass connection fairly short.

In addition to summing the voltages of the two differential PAs, the power combiner also transforms the 50Ω single-ended load impedance to a differential load impedance which is ideally 25 ohms for each of the two power amplifier stages. Custom finger (metal-oxide-metal) capacitors have been employed on both the primary and secondary side of the transformer to tune to the desired center frequency of 2.4GHz. Care has been taken to utilize higher finger spacing for the output capacitor, so that it can withstand the large voltage swing reliably. The power combiner has a simulated (using Agilent Momentum EM simulator) insertion loss of 1.25dB at 2.4GHz and varies by less than 0.15dB over the whole band of interest from 2.2-2.7GHz.

It was noted that excess inductance with the bypass capacitance can cause supply ripple

at twice the signal frequency $(2*f_o)$, degrading linearity. However, as much as bypassing the second harmonic current appropriately is important, it is equally necessary to bypass the even-order low-frequency (at modulation frequencies, 0-10MHz) signal [37]. Varying reactive impedance at baseband not only introduces distortion, but also introduces asymmetry between the lower and upper sideband distortion [63]. This makes one side of the output spectrum the limiting factor in setting the maximum output power at which the mask can be met. In this respect, transformers offer significant advantages. The transformer-based matching network produces very low impedances at the modulation frequencies (0-10 MHz). The use of transformer center-taps for DC biasing together with low-impedance current mirrors ($MB1$ in Fig. 4.1) helps to keep the asymmetry small.

4.1.3 Bypass Network and Common-Mode Stability
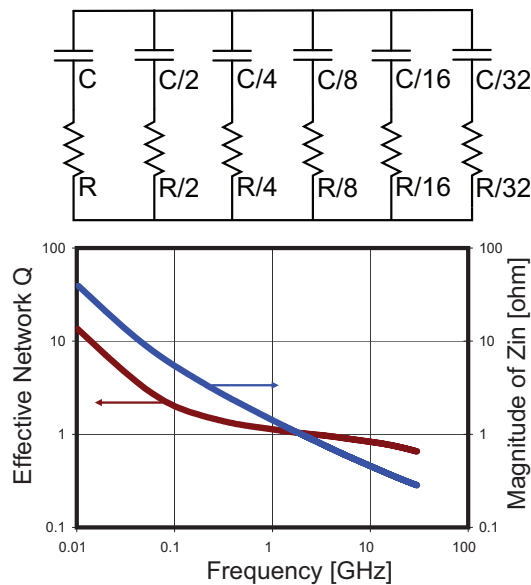
Figure 4.8: Bypass network and its frequency response for C=100pF and R=10Ω.

For low distortion and high efficiency, the second harmonic ($2f_0$) current at the output transformer's center tap needs to be properly bypassed to the transistor source. However, the combination of bypass capacitors with ground and supply inductances can produce a positive feedback path, resulting in common-mode oscillations, as was elaborated for the 60GHz power amplifier in the previous chapter. There are various ways in which such instability can be mitigated. Conventional techniques like adding gate resistors will affect the differential RF gain as well, reducing the power-added efficiency. A series resistor can be added with the bypass capacitor to de-Q it, however this will raise the impedance of the bypass path and cause voltage ripple, as stated above. Thus we need a bypass network which will have

low impedance at twice the signal frequency (for efficient bypassing), and simultaneously a low Q over a broad range of frequencies to quench any possibility of oscillations.

In this work, a staggered-RC bypass network has been proposed to meet these requirements (Fig. 4.8). Here, the values of the successive RC pairs have been scaled by a tapering factor, so that the zeros are offset from each other. The network provides a low Q over a broad frequency range, thus avoiding any high-Q shunt resonances between the capacitor bank and the package wire-bond. This technique has the virtue that it maintains stability regardless of the precise parasitic L-C resonance frequency, while providing an impedance which continues to decrease with increasing frequency. The simulated frequency response and Q-factor of the distributed bypass network are shown in Fig. 4.8. In layout, the parasitic inductances of traces in series with the capacitors modifies the exact shape of the response, however up to most practical values of frequency (where the transistor has high gain) the difference is not significant.

4.2   Measurement Results
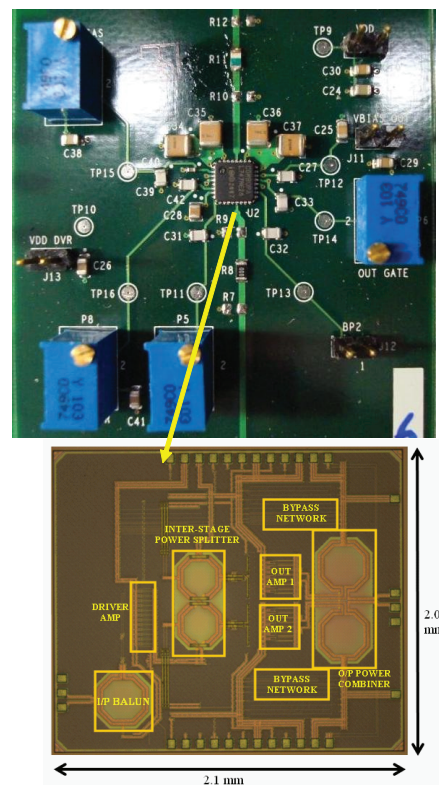
Figure 4.9: Test FR4 board with the packaged chip and die micrograph of the PA.
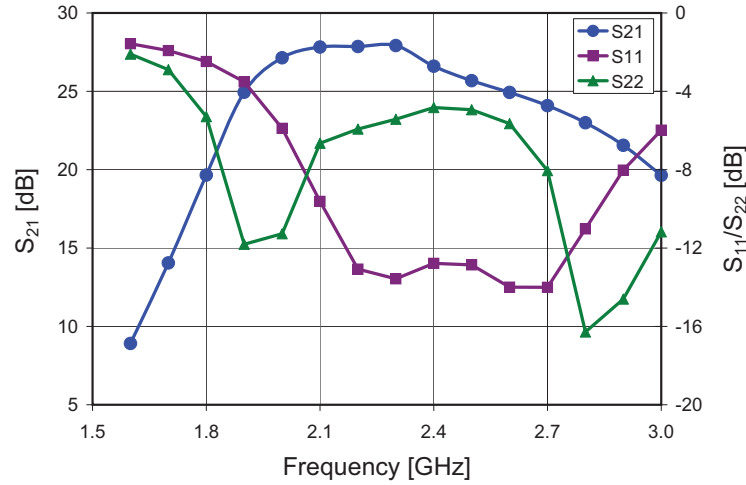
Figure 4.10: Measured small-signal performance of the PA.

The linear power amplifier with transformer power-combiner was designed and fabricated in TSMC 90nm CMOS technology. The chip was packaged using a 32-pin MLF package and mounted on a FR4 board (Fig. 4.9). The die micrograph is also shown in Fig. 4.9.

Small signal measurements (Fig. 4.10) were performed on the PA at a supply voltage of 3.3V. The amplifier has a peak gain of 28dB at 2.3GHz with about 600MHz of 3-dB gain bandwidth. The input match is better than -12dB over the whole band of interest.

Next large-signal measurements were performed using a CW-tone at 2.3GHz. The output was checked on a spectrum analyzer to confirm that there were no spurious oscillations. The measured Psat is 30.1dBm, and P_{-1dB} is 28dBm. The P_{-1dB} can be varied from about 27-28.5dBm by adjusting the gate bias of the output stage: note that the effect of gate bias on AM-AM response was already shown in Fig. 4.4. The peak drain efficiency is 36%, and the peak PAE is 33% (Fig. 4.11).

Next, in order to test the linearity of the PA, a two-tone test was performed where two sinusoids around the center frequency with a certain frequency offset were injected into the PA. The output spectrum for a tone-spacing of 1MHz is shown in Fig. 4.12. As shown in figure, the IM3 at 23dBm output power is better than -34dBc. IM_3 better then -30dBc had been estimated in system simulations to be sufficient to meet spectral mask in a modulated test. An important feature to note in the spectrum is the balance between the upper and lower distortion products. Due to the low AM-PM distortion in the PA and the low-impedance at the bias nodes created by the transformer, the asymmetry between distortion sidebands is small (less than $2dB$).

The PA has been tested for more than 200 hours continuously while putting out saturated output power in the CW mode. The degradation in output power over time is less than 0.8dB
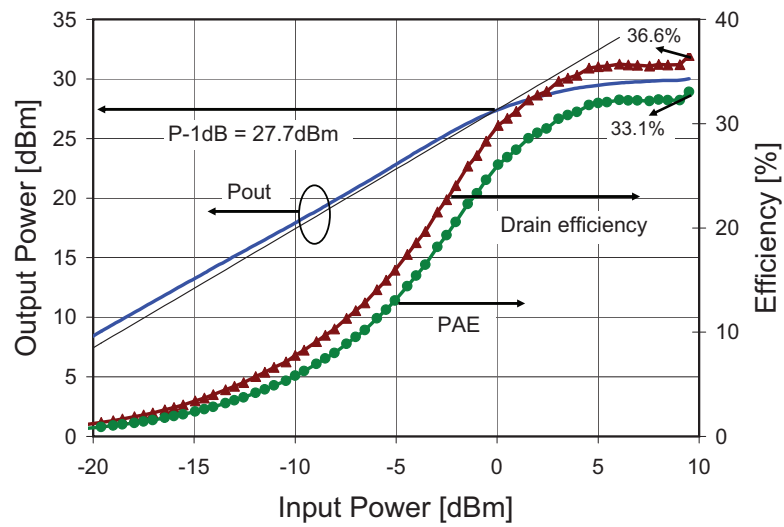
Figure 4.11: Measured large-signal performance of the PA.
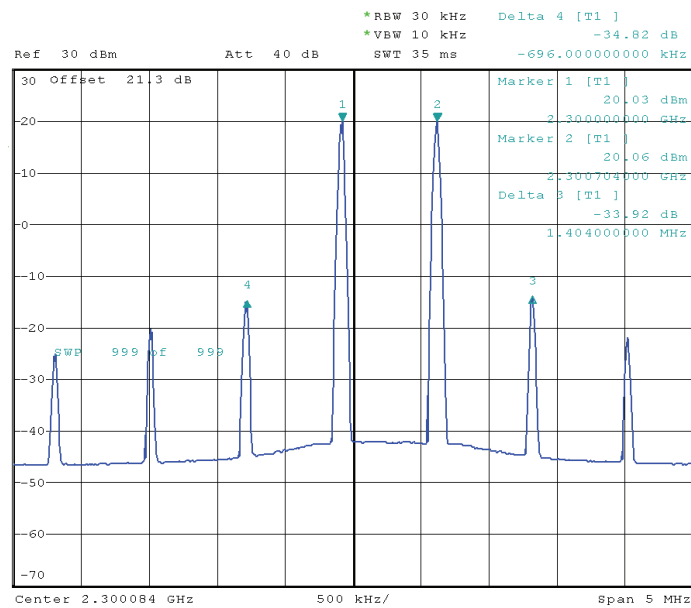
Figure 4.12: Intermodulation distortion in a two-tone test with 1MHz tone spacing.
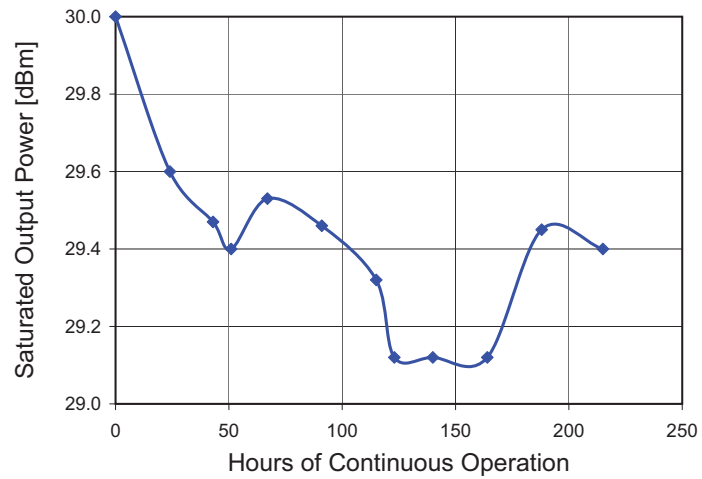
Figure 4.13: Saturated output power of the PA as a function of time.
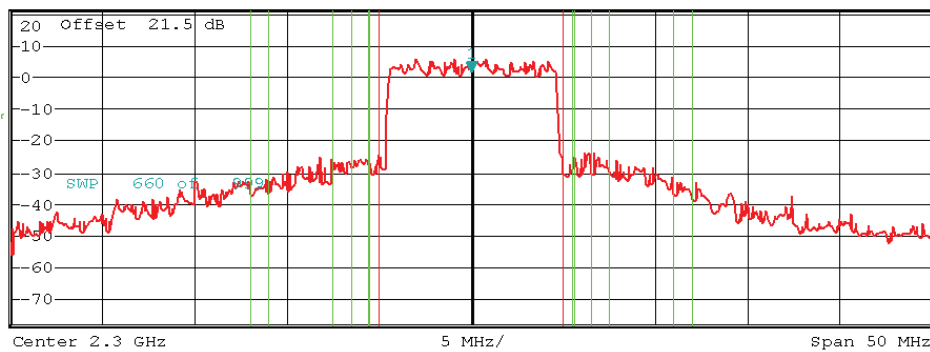
Figure 4.14: Output spectrum of the PA excited with 802.16e mobile WiMax signal.
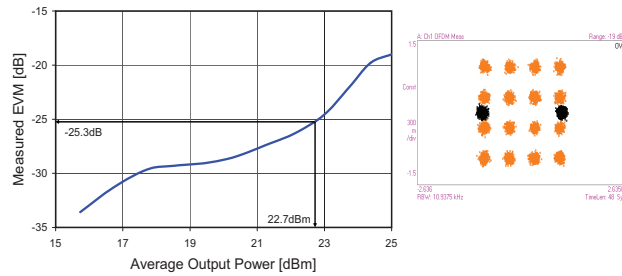
Figure 4.15: Measured EVM as a function of average output power.

after 200 hours of operation (Fig. 4.13). In practice the PA hits the peak power very rarely, and so this translates into several years of operation. The PA has been tested under severe mismatch (8:1) as well as elevated supply voltage (4.4V) and shows no stability or reliability problems.

Finally, the PA was tested with complex modulation. As an example, the 802.16e mobile WiMAX standard has been chosen. It employs 1024 carrier OFDM signaling with a modulation bandwidth of 10MHz. Each carrier can be modulated using BPSK, QPSK, or QAM. Such a standard indeed requires high linearity and is thus a good test of the design. When excited by the WiMAX signal, the PA delivers 22.7dBm of average output power at 14.2% drain efficiency and 12.4% PAE, while being fully compliant with the spectral mask requirements. The increased flatness in the AM-AM and AM-PM response enables the PA to meet the mask with only 7.4dB back-off from P_{sat}. A ideal PA with a "brick-wall" gain characteristic would have required about $6.5dB$ back-off from the peak power to meet the WiMAX mask. The linearization technique has thus allowed the PA to meet the mask with only 0.9dB more back-off. Fig. 4.14 shows the output spectrum for 22.7dBm average output power.

The measured EVM for 16/64 QAM OFDM input is better than -25dB at the average output power and meets WiMAX requirements. The EVM as a function of average output power is plotted in Fig 4.15. In addition to the spectral mask, the FCC also specifies the power that can be put out at harmonic frequencies. Due to the use of a tuned transformer, the harmonic distortion levels were safely below the requirements, without the use of any external band-pass filtering.

4.3 Transformer Power Back-Off

The designed PA delivers about 23dBm of average power and 30dBm of peak output power. Such high output power is needed for long-range applications like WiMAX, LTE, etc. However, the PA used here has an interesting property that it can be also be used for slightly lower power applications like WiFi. Normally, if the PA output power is reduced, its efficiency drops almost linearly with the output amplitude. However, by virtue of the 2:1 power
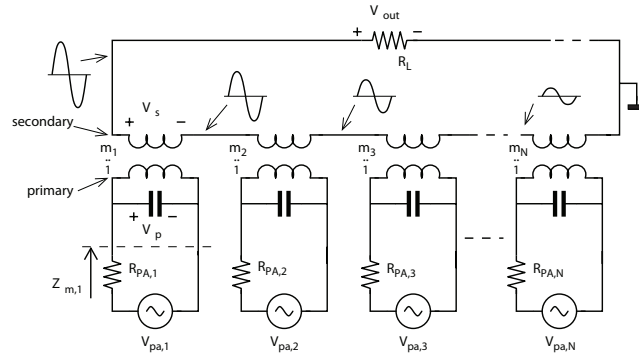
Figure 4.16: Transformer-based power combiner.
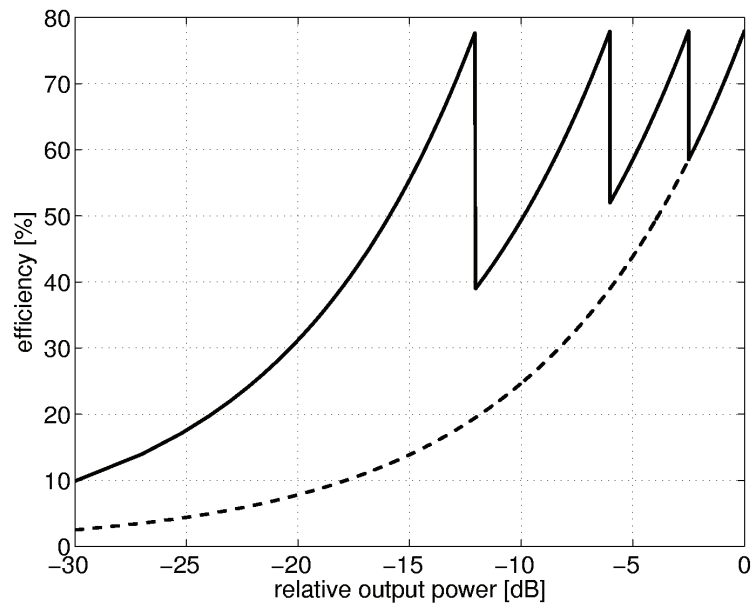
Figure 4.17: Efficiency enhancement with power back-off.

combiner, we will show that we can create a dual-mode PA where it is almost equally efficient in both the modes despite the variation in output power. In particular, since the PA contains 2 (or more generally N) independent amplifier stages combined through transformers, some of these stages may be turned off at lower average power to improve efficiency [20].

Consider a topology consisting of N identical stages coupled using N 1:1 transformers (Fig. 4.16). When all N stages are active, the transformed differential impedance seen by each stage is $R_m = R_L/N$ and the gain of each stage is $g_m R_L/N$. If V_i is the input voltage, the differential output voltage of each stage V_p and the total output voltage (V_{out}) across the load are given by [48]

$$V_p = \frac{g_m R_L}{N} \times V_i \tag{4.1}$$

$$V_{out} = N \times V_p = g_m R_L V_i \tag{4.2}$$

This is the case when peak output power is being generated and all the amplifiers are designed to operate at maximum efficiency at this power level. Now suppose that the input voltage reduces from V_i to $V_i/2$ and half of the parallel stages are turned off. The impedance transformation ratio changes from $1 : N$ to $1 : (N/2)$ and each stage sees an effective resistance $R_m = 2R_L/N$. The output voltage of each stage and the total output voltage are now given by

$$V_p = \frac{g_m R_L}{N/2} \times \frac{V_i}{2} = \frac{g_m R_L}{N} \times V_i \tag{4.3}$$

$$V_{out} = N \times V_p = g_m R_L V_i \tag{4.4}$$

Thus we see that in spite of turning off half of the stages (6dB power back-off), because of dynamic load variation, the output swing of each active stage (V_p) is still the same and hence it operates at peak efficiency even at power back-off. The same is true regardless of the number of stages turned off and allows different power back-offs while maintaining high efficiency. Fig. 4.17 shows the theoretical PA efficiency vs. output power for the case of N=4.

In the design presented in this work, since we have two amplifiers being power-combined, we can turn one off at around 6dB back-off to transmit 17dBm average output power instead of the full value of 23dBm without significant efficiency degradation. This can allow the backed-off PA to be re-used as a WiFi PA, since the power level required for WiFi is around 15-18dBm [64]. In order to ensure that the primary of the PA that is off does not add extra loss, it is necessary to short it when that PA stage is turned off. This has been achieved in this design using thick-oxide NMOS transistors as switches, as shown in Fig. 4.19.

The PA was measured in the low-power mode by turning off one of the stages. As can be clearly seen from Fig. 4.18, if the output power would have been reduced simply by reducing input drive (and keeping both the stages on), the efficiency would be degraded to 6% at 18dBm average output power. However, by turning off one of the stages, it is
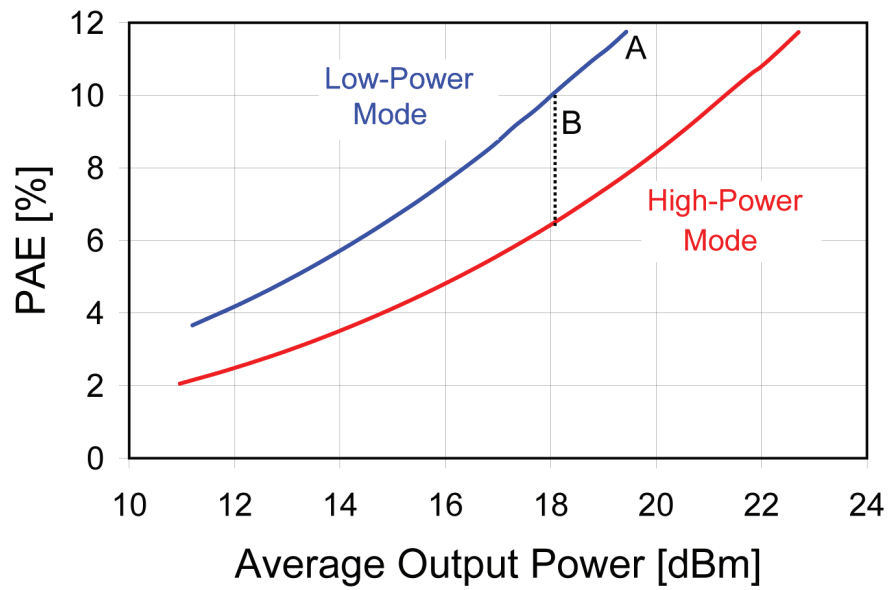
Figure 4.18: Efficiency of the PA in low-power and high-power mode.
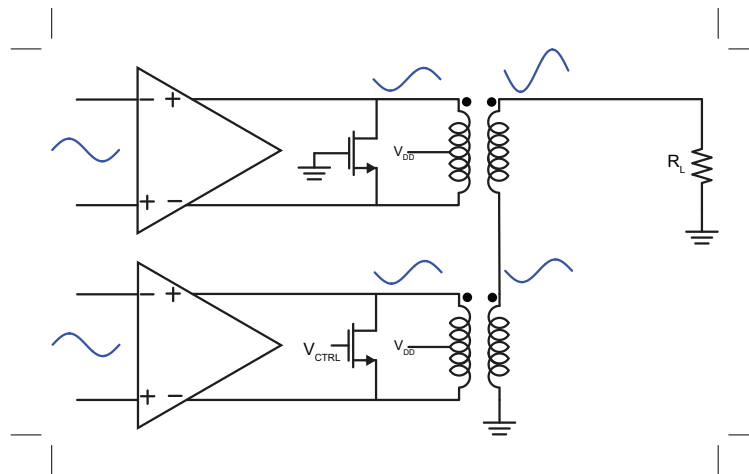
Figure 4.19: Implementation of power back-off with shorting switches.

possible to improve the backed-off efficiency to about 10.4%, which is very close to the peak PAE obtained at 22.7dBm output power (12.4%). The PA has also been tested with 802.16e WiMAX signal and 802.11g WiFi signal in this low-power mode and satisfies both the EVM and spectral mask requirements. This provides a methodology to create a dual-mode integrated WiFi/WiMax PA in CMOS technology.

4.4   Summary

In this chapter, we have demonstrated a highly linear power amplifier in nanoscale CMOS technology that can deliver high output power. Transformers have been extensively used as matching networks in this design, as they are compact and can perform impedance transformation efficiently. The choice of optimal bias to produce flat AM-AM response, along with an AM-PM compensation, can create highly linear PAs even in fine-line CMOS processes. The linearity of the PA has been verified using both single-tone and high dynamic range modulated signals. In the next chapter, we will examine what are the shortcomings of such an analog power amplifier and if there exists any better ways to do a holisitc design of a transmitter without sacrificing any performance.

Chapter 5

A 2.4GHz Mixed-Signal Polar Power Amplifier

In the previous chapter, we described a high-power linear CMOS power amplifier, employing a transformer-based on-chip power combiner. Such linear power amplifiers constitute the last building block in a conventional transmit (TX) chain, as shown in Fig. 5.1. In such a cartesian, direct-conversion transmitter, the baseband digital I and Q data are converted to the analog domain using digital-to-analog converters (DACs). The analog I and Q signals are low-pass filtered and then up-converted to the carrier frequency using a quadrature mixer. The outputs of the I and Q mixers are combined together, often in the current domain, and fed to a multi-stage power amplifier. The PA amplifies the signal and delivers the desired output power to an external load (filter or antenna) with as much efficiency as possible.

The RF signal resulting from combination after the quadrature modulators (V_{RF}) can be expressed as

$$V_{RF} = I \cdot \cos \omega t + Q \cdot \sin \omega t \tag{5.1}$$

V_{RF} has both amplitude and phase information, as is common in wireless systems using complex modulation like 16/64/256 QAM. Since the input signal to the power amplifier has a varying envelope, a linear power amplifier (Classes A/AB/B), like the one described in
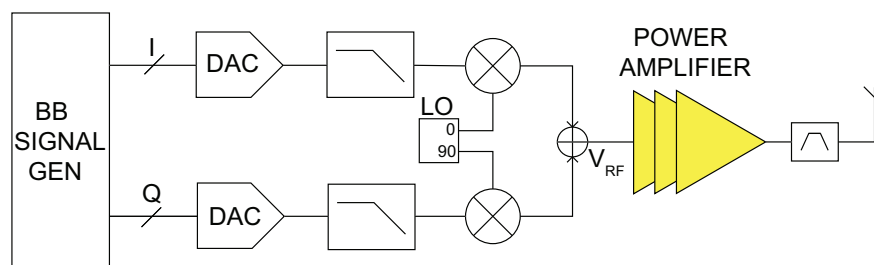
Figure 5.1: Block diagram of a conventional wireless transmitter.

Chapter 4, needs to be used. However, it is well-known in literature that a linear power amplifier has lower power efficiency than switching power amplifiers (Classes D/E/F). However, being switches, such PAs cannot transmit amplitude modulation and hence cannot be employed in wireless systems where the signal is not constant-envelope. Thus there exists a linearity-efficiency tradeoff, well-known to the PA design community. As outlined in Chapter 1, various architectures have been proposed in literature to break this trade-off.

The polar architecture has emerged as a promising system-level approach for the realization of a flexible, efficient TX architecture. In a polar system, the baseband data is represented as amplitide (A) and phase (ϕ), instead of the conventional I and Q, as shown in Eq. 5.2.

$$A = \sqrt{I^2 + Q^2} \tag{5.2}$$

$$\phi = \tan^{-1}\left(\frac{Q}{I}\right) \tag{5.3}$$

The phase information (ϕ) is superimposed on a carrier to create a phase-modulated RF signal (ϕ_{RF}). Since ϕ_{RF} is constant-envelope, a switch-mode PA, having high efficiency, can be employed. In a classic implementation, the amplitude modulation is reconstructed by varying the supply voltage of a saturated PA. This is achieved by using dedicated supply modulators, either switch-mode or linear regulators. However, such supply modulators have an efficiency-bandwidth tradeoff. Switching regulators have high efficiency but low bandwidth. In order to process wideband envelope signals, linear regulators are often used, which have lower efficiency. Thus the net efficiency of the supply-modulated PA is often low in wideband polar systems, even though the core non-linear PA block has high efficiency.

In addition to the efficiency-linearity tradeoff inherent in conventional analog transmitters, such analog systems also require a large number of passive matching networks. As we have seen in Chapter 2, such matching networks are always lossy, reducing system efficiency. In addition, in the lower GHz frequency, integrated inductors, capacitors and transformers tend to be bulky, increasing silicon area and system cost. With the continuous downscaling of CMOS technolgy, the transistor feature size decreases every generation, making digital circuits smaller in area. However, analog and RF circuits do not scale well with technology, since the area of the passive components at a given frequency is almost constant over technology nodes. Thus a solution which employs fewer passives, has smaller form-factor and is scalable to newer technology nodes would be very attractive.

A digitally-modulated polar power amplifier has the potential of allowing such a scalable solution, while simultaneously breaking the efficiency-bandwidth trade-off found in analog supply modulators. In addition, it can conserve power by merging more functionalities into one block, a feature very beneficial in mobile, portable applications, where battery life is the key. The principle of such a mixed-signal solution is shown in Fig. 5.2 [31] [32]. The amplitude (A) signal is in the digital domain, while the phase bits are used to generate a phase-modulated RF signal (ϕ_{RF}). In such a system, the core PA is sliced into several
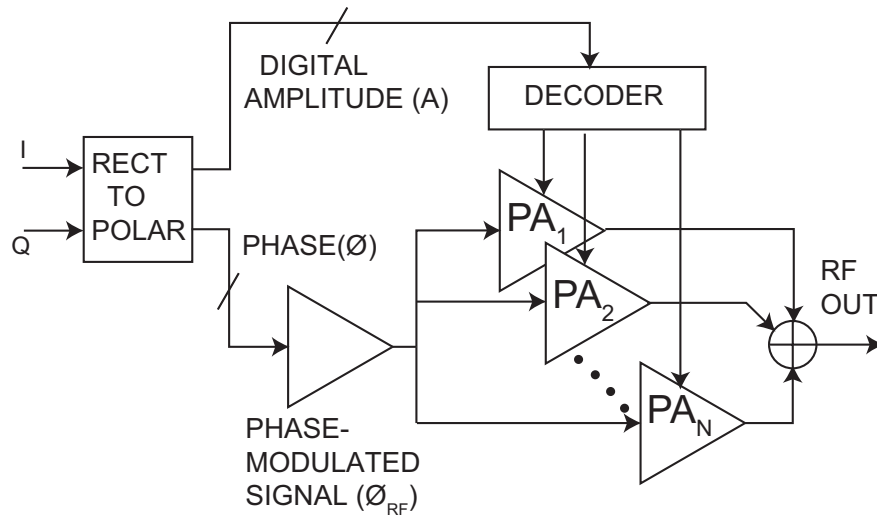
Figure 5.2: A digitally-modulated polar power amplifier.

weighted unit cells (like in a DAC), which are turned on or off by the amplitude control word (A). In this way, the amplitude information is transmitted. By employing such a digital switching technique, the analog circuit complexity is reduced, since wide-bandwidth supply modulators are avoided and the circuit is more directly interfaced to the digital baseband. In addition, because of the digital nature of ϕ_{RF}, simple inverters can be used as drivers, eliminating the need of inter-stage matching networks. The output matching network will still be needed for optimizing power and efficiency. Such a mixed-signal transmitter thus has the potential for multi-mode operation. Finally, the power amplifier, in such a system, absorbs the functionality of a DAC as well, thereby lowering required power consumption.

As outlined in Chapter 1, there has been some work published in literature that have investigated such digitally modulated architectures. However, there has been little work done on a detailed study of this digital polar architecture with emphasis on an efficient, high-power, fully-integrated PA design. This thesis investigates the design of such a mixed-signal transmitter architecture with major focus on an efficient PA design. We will show in this thesis how an efficient yet compact switching PA can be designed and how it can be operated as an RF DAC. Systems level considerations like appropriate choice of the amplitude and phase resolution, clocking rates and filtering requirements will be discussed in details. The proposed design methodology will be verified through measurement results performed on a 65nm CMOS prototype.
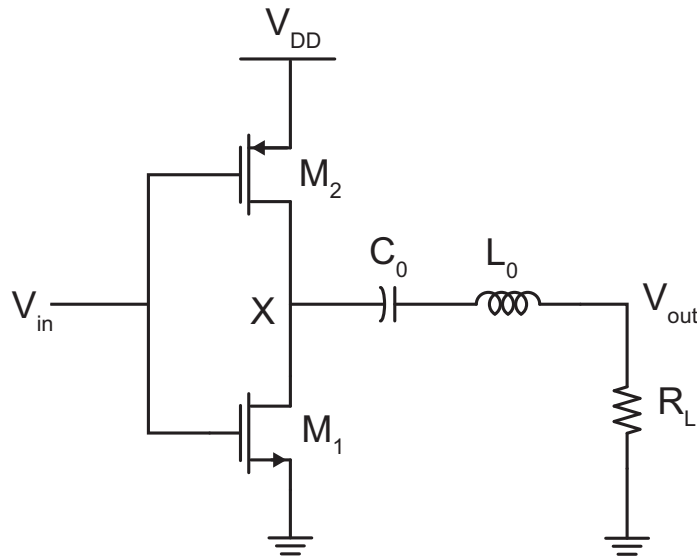
Figure 5.3: The Class D power amplifier.

5.1 Inverse Class-D PA as Unit Cell

As mentioned earlier, a switching PA topology can be utilized in this polar architecture since the input RF signal is constant envelope. In a switching amplifier, as the name suggests, the transistor acts as a switch rather than as a linearly-controlled current source. The most common switching PA topologies are Classes D and E.

The basic Class D power amplifier is depicted in Fig. 5.3. The input signal (V_{in}) and the drain voltage (V_X) are both square waves. If this square wave was applied directly to the load resistor R_L, significant harmonic power would be wasted. Hence, a resonant tank consisting of C_0 and L_0 is inserted in series with R_L in order to allow only sinusoidal current through R_L. Figure 5.4 depicts the drain voltage and drain current waveform of the ideal Class D amplifier. The drain voltage will be a square wave with an amplitude of $V_{DD}/2$ and a DC value of $V_{DD}/2$. The first harmonic of that square wave has an amplitude of $V_{DD} \cdot 2/\pi$ and therefore the output current has an amplitude of [65]

$$I_o = \frac{(2/\pi)V_{DD}}{R_L} \tag{5.4}$$

and the output power will be equal to

$$P_o = \frac{2}{\pi^2} \cdot \frac{V_{DD}^2}{R_L} \tag{5.5}$$

There is ideally no overlap between the current and voltage waveforms, thereby in principle achieving 100% efficiency.
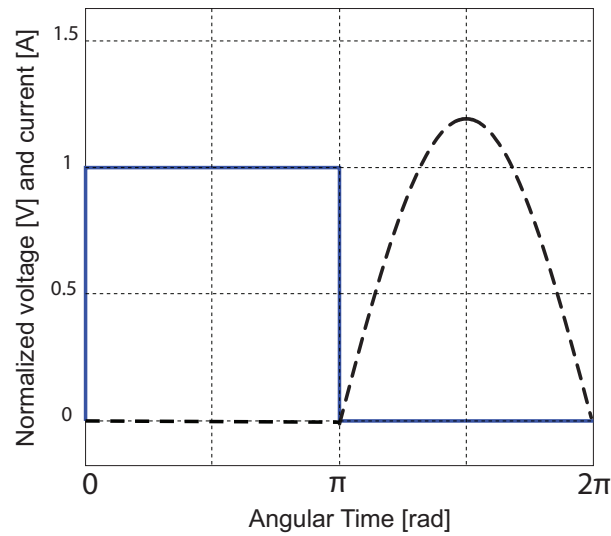
Figure 5.4: Current and voltage waveforms in the Class D PA.

However, for GHz-range applications, Class D amplifier has some drawbacks. The drain capacitor of the transistor is not a part of the matching network and needs to be charged and discharged every cycle, leading to a CV^2f power dissipation. Another disadvantage of the conventional Class D PA is the that it requires the use of a PMOS device, which has higher on-resistance than its NMOS counterpart. Typically, to reduce the on-resistance of the PMOS switch, its size needs to be increased two to three times, thereby increasing input capacitance significantly and making the driver design more challenging.
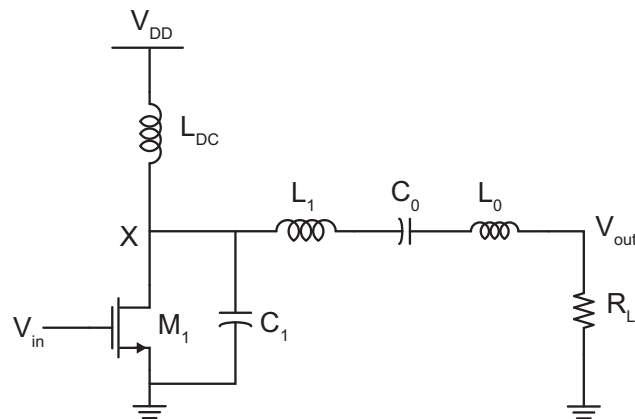
Figure 5.5: The Class E power amplifier.

Class E amplifiers have become popular since they can overcome the CV^2f loss associated with the drain parasitics of the transistor. Like Class D, Class E is also capable of achieving

100% efficiency. The basic circuit of a Class E amplifier is shown in Fig. 5.5. Due to the series resonant tank (L_o and C_o) the output voltage will be sinusoidal, and no harmonic power will be wasted. The main advantage of Class E design is that when the switch is closed, the voltage across it is zero and therefore the drain parasitic capacitance of the switch need not be discharged. This is referred to as *zero voltage switching* (ZVS). In fact, the parasitic drain-source capacitance of the switch can be absorbed into C_1 in Fig. 5.5. The Class E theory, as proposed by Sokal and Sokal in 1975 [66], requires that both the switch voltage and its first derivative are zero (ZVS and $dZVS$) when the switch closes. The requirement for a zero first derivative is not absolutely necessary to obtain 100% efficiency. However, this property makes the amplifier less sensitive to component variations.

In spite of being an attractive architecture, Class E suffers from some practical issues. Ideal Class-E operation requires a large choke (L_{DC}) which is difficult to achieve in an integrated CMOS design. In addition, it requires multiple passive components, whose finite quality factor quickly degrades the achievable efficiency. Multiple passives also mean increased silicon area and higher cost, which are undesirable for a portable solution. Finally, the amplifier uses a series inductor, which is not readily compatible with baluns or transformers that are anyhow needed for differential PAs to interface to a single-ended antenna.

As a compromise between efficiency, area and ease of interfacing, we propose the use of a current-mode inverse Class-D (D^{-1}) PA. Such an amplifier is illustrated in Fig. 5.6. This amplifier is the dual of a conventional Class D PA [67] and uses only one parallel LC tank that can be conveniently absorbed into the output transformer network. In a Class D^{-1} PA, the current through the transistor is a square wave, whereas the output voltage is sinusoidal (due to the parallel resonant network). The drain voltage (normalized to V_{DD}) and current (normalized to I_D) of one of the transistors is illustrated in Fig. 5.7. Similar to other switching power amplifiers, the overlap between high voltage and high current is avoided, thereby achieving 100% ideal drain efficiency. In addition, when the transistor turns on, the voltage across the transistor is zero. Thus, the output capacitance discharge problem is eliminated, just like in Class E amplifier. The current-mode Class D can thus avoid losses by ensuring ZVS, but it cannot ensure $dZVS$ like Class E. However, the passive network complexity is greatly reduced in this topology. The parallel LC tank required in this design (L and C in Fig. 5.6) is realized using an on-chip transformer, which simultaneously performs impedance transformation and differential-to-single ended conversion. The architecture of transformer-coupled inverse-D PA is thus very compact and amenable to integration in nanoscale CMOS processes.

5.1.1 Practical Design Considerations

In practical operation, several factors distort the ideal voltage/current waveforms and tend to degrade the PA efficiency. These non-idealities also impose design trade-offs which are described below.
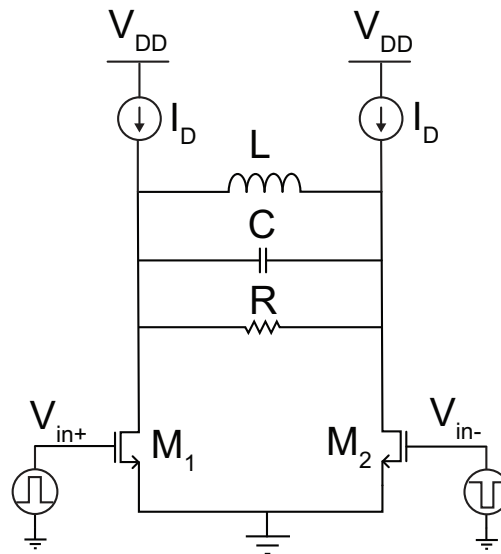
Figure 5.6: The inverse Class-D (D^{-1}) power amplifier.
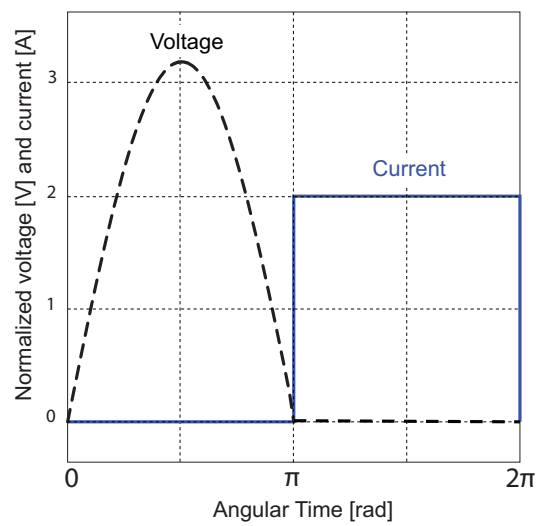
Figure 5.7: Current and voltage waveforms in Class D^{-1} PA.
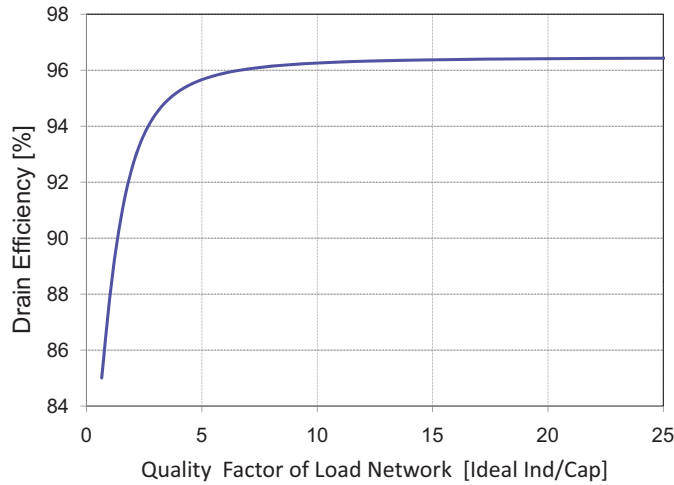
Figure 5.8: Efficiency of an ideal Class-D^{-1} PA as a function of the loaded quality factor of the RLC network. Here, the inductor is assumed lossless, and $Q = \omega \cdot R \cdot C$.

- Higher order odd harmonic current leakage into the load resistor (R in Fig. 5.6) results in efficiency loss. The drain current is ideally a square wave having only odd harmonics, while the drain voltage is a half-sinusoid. This means that the shunt capacitor C in the LC resonator is intended to provide a short circuit for the higher order odd harmonic currents. Of course in practice there are higher order leakage currents flowing through the load that will reduce the efficiency since some power is wasted in the harmonics. This leakage depends upon the loaded quality factor (Q) of the resonator given by $Q = \omega \cdot R \cdot C$. The efficiency ($\eta$), assuming that everything else is ideal and this higher harmonic leakage is the only loss mechanism, can be derived as [68]

$$\eta = \frac{1}{1 + \displaystyle\sum_{n=1}^{\infty} \frac{1}{(2n+1)^2} \cdot \frac{1}{1 + ((2n+1) \cdot Q)^2}} \tag{5.6}$$

The above equation shows that a higher Q is desirable for increasing the efficiency. With lower Q, the voltage waveform has more distortion. This problem is definitely exacerbated in low-voltage CMOS process. In order to generate high output power while using lower supply voltage, lower R is needed, which degrades the network quality factor. The above result would justify a choice of a higher C and lower L (for a given R) in order to increase the loaded quality factor. In fact, if the matching network were lossless, simulations confirm that the efficiency increases as the parallel capacitance (C) value is increased (R is held fixed at 14Ω), while simultaneously adjusting the inductor value to resonate at the fundamental frequency. This result is shown in Fig. 5.8.

In reality, the transformer efficiency places a conflicting requirement that does not
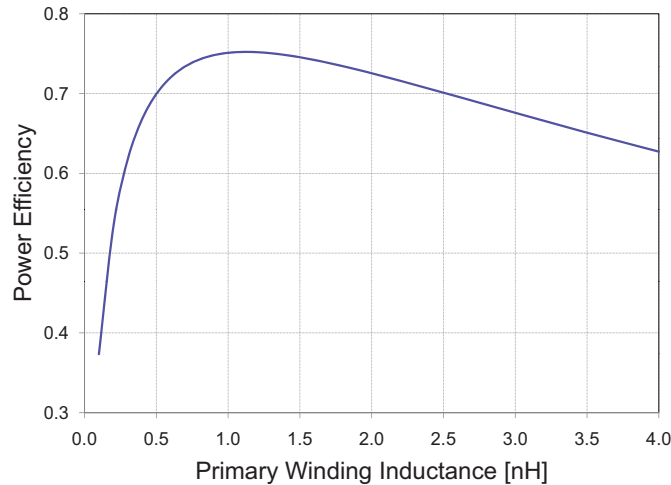
Figure 5.9: Efficiency of a 2:1 transformer at 2.4GHz as a function of primary winding inductance.

allow the capacitance to be increased arbitrarily. A transformer is utilized to perform the necessary impedance transformation (50Ω to R) and also provides the necessary inductance value (L). It has been shown in Chapter 2 that there exists an optimum inductance value that maximizes the transformer efficiency at a given frequency [7]. This trend is shown again in Fig. 5.9. As is clearly seen from this result, the optimum primary winding inductance for the chosen transformation ratio (2:1) and assumed winding quality factor ($Q = 10$) is around $1\ nH$. Deviating largely from this optimum value will lead to a reduction in transformer efficiency and hence there exists optimum values of L and C which will maximize the net PA efficiency. In this design, an inductance of $700pH$ was chosen, which made the parallel capacitance roughly $4pF$.

- One of the major loss mechanisms in such switching power amplifiers is the transistor on-resistance. Fig. 5.10 shows the simulated efficiency degradation of the PA when all other elements are ideal and the only non-ideality is the switch on-resistance. The switch resistance for a transistor can be approximated as

$$R_{on} = \frac{1}{\mu_n C_{ox} \frac{W}{L}(V_{GS} - V_T)} \tag{5.7}$$

where μ_n, C_{ox} are process parameters, W and L are the device width and length, V_{GS} is the switch gate-source voltage, and V_T is the threshold voltage. Clearly, the simplest way to reduce the on-resistance is to increase the device width W.

However, a real switch has parasitic drain capacitance (C_{par}). Increasing the device width increases C_{par}, which alters the ideal even harmonic impedances. In an ideal
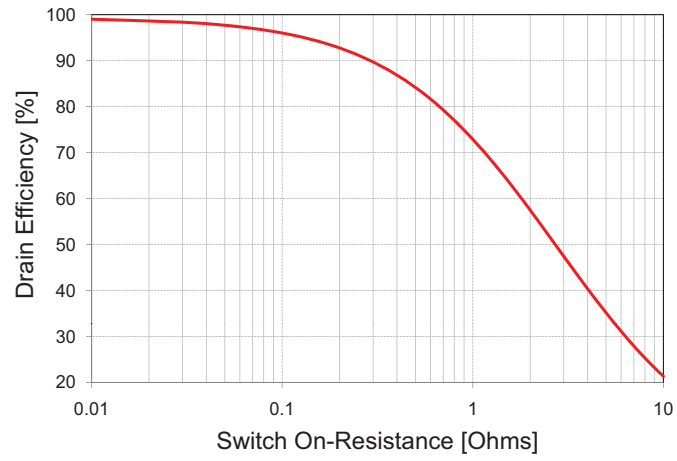
Figure 5.10: Variation in PA efficiency as a function of the switch resistance.
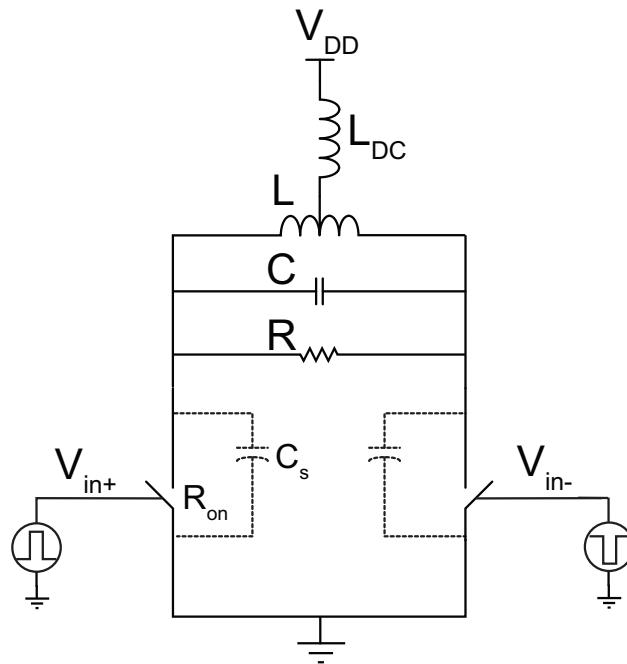
Figure 5.11: Simulation setup of an ideal inverse-D PA.
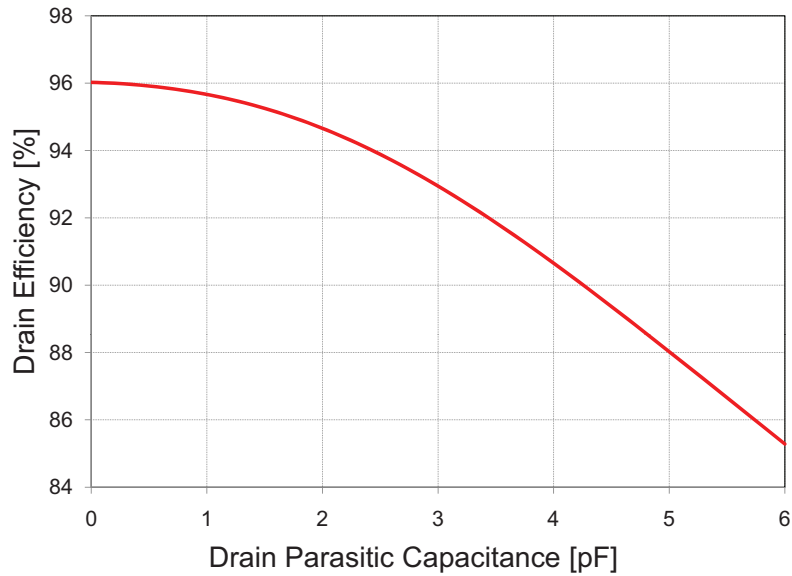
Figure 5.12: Degradation in drain efficiency due to increased device parasitic.

inverse-D PA, the current is a perfect square wave, which means it has all odd harmonics but no even harmonic component. Thus, the impedance presented by the matching network in Fig. 5.6 at even harmonics should be infinite. However, the transistor drain parasitic makes this assumption invalid at RF frequencies. For fundamental mode, C_{par} can be absorbed into the matching network capacitor (C). However, since C_{par} is single ended, it creates a path for even harmonic leakage. This distorts the current waveform, increasing the overlap between transistor voltage and current waveforms and lowering the efficiency of the PA.

In order to verify this, an otherwise ideal PA set-up was chosen, as shown in Fig. 5.11. The transistor on-resistance (R_{on}) was fixed at 0.1Ω, so that its impact is minimal. The total tuning capacitance was fixed at $10pF$ and was split between the differential tuning capacitor (C) and two single-ended capacitors (C_s). The value of C_s is swept, while changing C simultaneously to keep the tank resonant at fundamental frequency. Fig. 5.12 shows that everything else being fixed, the addition of single-ended drain capacitance degrades the drain efficiency by lowering the even-order harmonic impedance. The conflicting requirements of R_{on} and parasitic capacitance results in an optimum device width that balances these effects. In this work, a device width of $2.5mm$ was chosen.

- In our original discussion of current-mode inverse Class-D PA, as outlined in Fig. 5.6, ideal current sources (having infinite output impedance) were assumed. However, in a real implementation, the use of active current sources reduces the allowable swing.
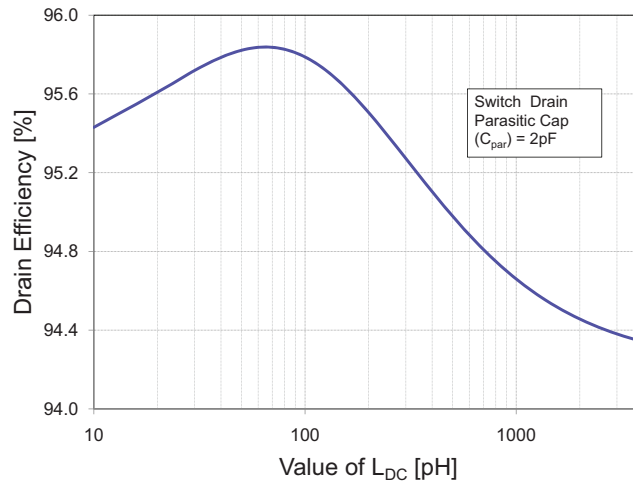
Figure 5.13: Dependence of drain efficiency on DC feed inductance value, in presence of finite drain parasitic capacitance.

Hence a DC-feed inductor (L_{DC}) is used, as shown in Fig. 5.11. It might seem that a large inductance is needed to increase the even harmonic impedance.

However, the scenario changes because of the presence of finite drain parasitic (C_{par}). The even order harmonic currents now have two paths - one capacitive (through C_{par}) and other inductive (through L and L_{DC}). Since one is capacitive and one inductive, it opens up the possibility of choosing L_{DC} appropriately to make the second harmonic impedance approach infinity. Fig. 5.13 shows that indeed such an optimum exists. Though the optimum is shallow and depends upon transistor parameters, for most practical values the optimum value of L_{DC} is lower than $500pH$, meaning a separate inductor is not needed for this purpose.

Taking all the above factors into consideration, an inverse-D PA was designed. A cascode topology was adopted in this design for reliability. Operating from a 1V supply, this PA achieves an intrinsic drain efficiency of 77% (lossless matching network). Fig. 5.14 shows the simulated drain efficiency as the parallel tuning capacitor is varied. The output power plot is shown in Fig. 5.15. The peak output power from the transistors is around 24dBm. The simulated voltage and current waveforms is shown in Fig. 5.16. Clearly, the waveforms are no longer ideal square waves or half-sinusoids. However, over a period, the overlap between high voltage and high current has been minimized greatly.

5.1.2 Matching Network for Class-D^{-1} PA

As mentioned previously, the output matching is performed by a 2:1 transformer. A parallel primary transformer, as described in details in Chapter 2, has been utilized here. The
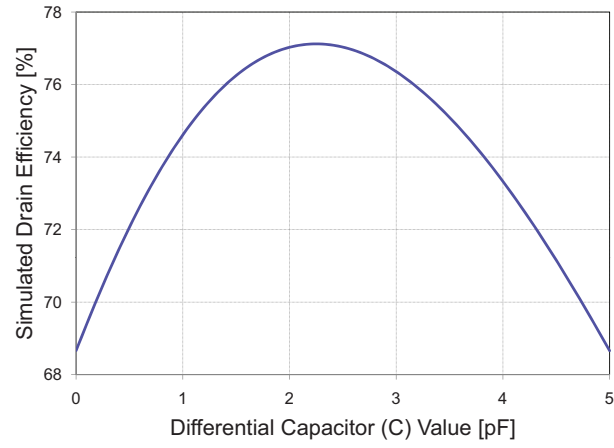
Figure 5.14: Simulated drain efficiency of Class-D^{-1} PA as a function of the parallel tuning capacitor.
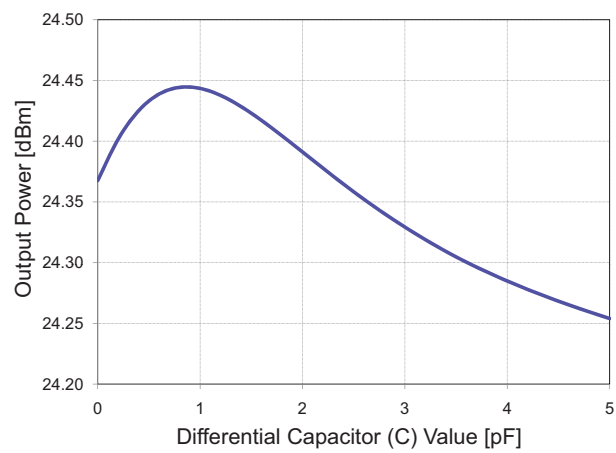
Figure 5.15: Simulated output power of Class-D^{-1} PA as a function of the parallel tuning capacitor.
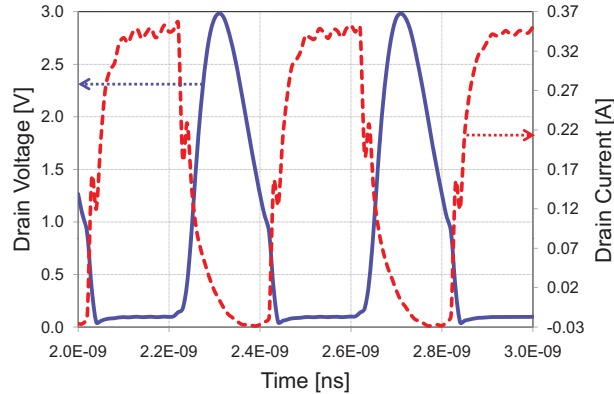
Figure 5.16: Simulated drain voltage and current waveforms in CMOS Class-D^{-1} PA.

schematic of the final transformer-coupled current-mode Class-D^{-1} PA is shown in Fig. 5.17 [46]. The effective L of the transformer and the parallel C values are selected to attain ZVS switching. After post-layout extraction, the PA with on-chip transformer has a simulated peak efficiency of 48% (Instrinsic drain efficiency = 77%, transformer efficiency = 74%, Loss due to DC I*R drop in PA output lines, transformer and supply connections accounting for rest). No ultra-thick metal (UTM) has been used for transformer layout or for supply routing; simulations reveal that if such an UTM were available the net PA efficiency would be increased to around 55%.

In the design of the transformer-coupled power amplifier, an important issue that needs attention is ground bounce. Ground bounce which is caused by higher order harmonic currents flowing off-chip through external bond wires and then returning on-chip through ground bond wires is a well-known phenomenon. On-chip bypass capacitors on supply nodes are typically used to provide a low impedance path for RF currents to circulate. However, in such transformer-coupled designs, there exists another mechanism that can cause ground bounce, as shown in Fig. 5.18. As discussed in Chapter 2, in any real implentation of a transformer, there exists inter-winding capacitance (C_w in Fig. 5.18) between the primary and secondary windings. The inter-winding capacitance between transformer windings causes current injection from the secondary to the primary winding. Since the ground references on the primary and secondary side are usually not connected on-chip to raise the even-mode impedance, this leakage current must return to the secondary port by flowing off-chip, and hence can cause substantial ground bounce due to the presence of bond-wire and PCB trace inductances. To reduce the inter-winding capacitance, a lateral transformer layout has been chosen, where the primary and secondary windings are in the same metal layer. To further reduce the impact of this ground bounce on the performance of the transmitter, a low-swing differential receiver with high common-mode rejection is designed to receive the
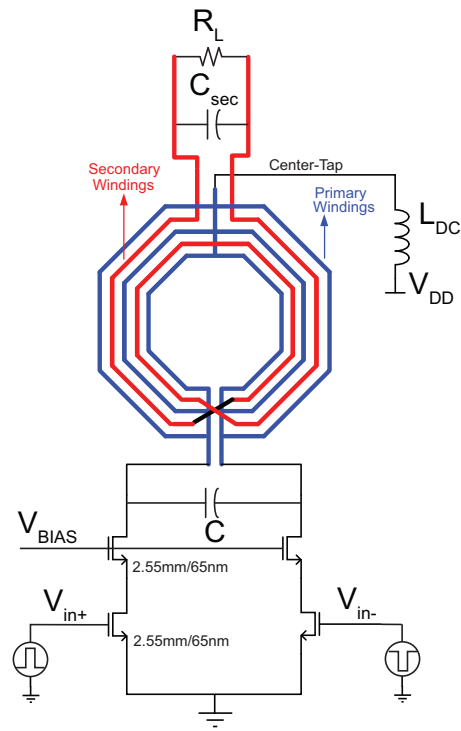
Figure 5.17: Schematic of the transformer-coupled inverse-D PA.
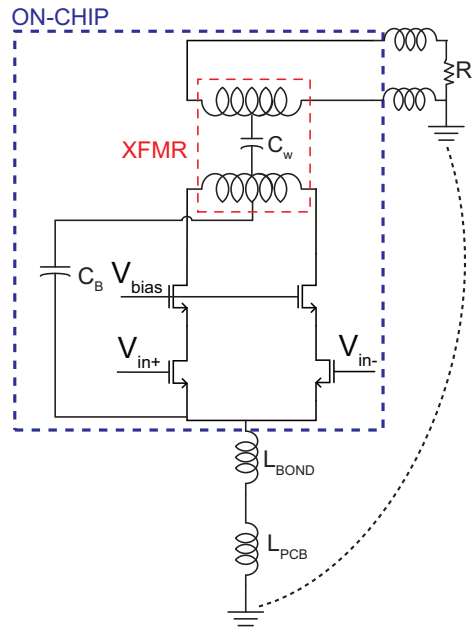
Figure 5.18: Ground bounce in a transformer-coupled PA.

phase-modulated RF signal, which is generated off-chip in this prototype.

5.2 PA as an RF-DAC

In the previous section, we dealt with the design of an efficient, compact PA that can be used as the core of a mixed-signal polar system, as described in Fig. 5.2. This PA will actually be used as an RF-DAC, where the unit cells are switched in and out, based on the amplitude codeword. In order to be able to design the system, we need to decide the needed resolution on the amplitude and phase paths, as well as the appropriate clocking rates.

5.2.1 Targeted Standard: 802.11g

The target application for this prototype was the 802.11g Wireless Local Area Networking (WLAN) standard. The IEEE 802.11g standard was released in 2003 with the vision of improving the maximum data rate of the IEEE 802.11b standard (11-Mb/s) by incorporating modulation similar to that used in IEEE 802.11a systems. Thus, IEEE 802.11g systems offer the same maximum data rate (54-Mb/s) as IEEE 802.11a systems, but at the lower carrier frequency of 2.4GHz. In order to accommodate high data rates and a large number of users in the limited available frequency spectrum, IEEE 802.11a and g systems employ a spectrally efficient modulation named orthogonal frequency division multiplexing (OFDM). In 802.11g, each OFDM symbol consists of 52 subcarriers, of which four are set aside as
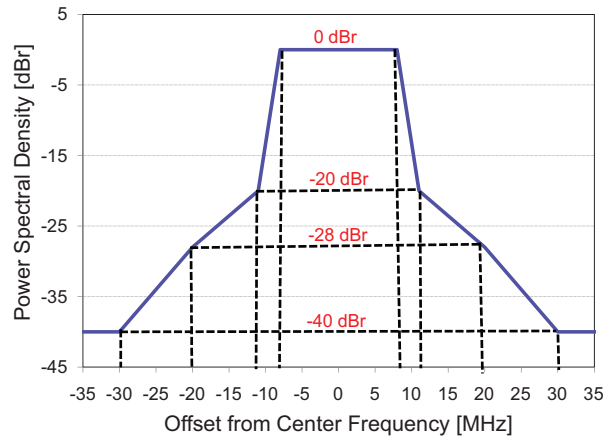
Figure 5.19: WLAN 802.11g transmit spectral mask.

Table 5.1: EVM requirements of WLAN 802.11g standard.

Modulation	Data Rate (Mbit/s)	Min. EVM (dB)
BPSK	6	-5
BPSK	9	-8
QPSK	12	-10
QPSK	18	-13
16-QAM	24	-16
16-QAM	36	-19
64-QAM	48	-22
64-QAM	54	-25

pilot tones and the remaining 48 carry data. The subcarriers are each modulated using conventional quadrature amplitude modulation (QAM), namely, BPSK, QPSK, 16QAM, or 64QAM. The center frequency (subcarrier 0) is left as NULL. The subcarrier frequency spacing is $0.3125MHz$. Therefore, the total occupied bandwidth is $16.6MHz$ ($0.3125MHz \times 52$). The total allocated bandwidth is $20MHz$. As described in the previous chapter, every wireless standard specifies a spectral mask and EVM requirement. The spectral mask for IEEE 802.11g is shown once again in Fig. 5.19. In addition to spectral mask, transmitters also need to achieve a certain error vector magnitude (EVM). Table 5.1 lists the EVM requirements of the IEEE 802.11g standard for different modulation schemes and data rates. As can be seen from the table, the strictest requirement is for 64QAM signals that carry data at the rate of 54 Mb/s. The digitally modulated polar transmitter thus needs to meet the EVM and spectral mask when transmitting 54Mbps, 64QAM data.

Figure 5.20: EVM of an ideal digitally-modulated transmitter while transmitting 802.11g 64QAM data. The variation in EVM is shown for two PAPR cases- 6 dB and 9 dB.

5.2.2 Amplitude and Phase Resolution

The amplitude and phase resolution of the system is determined by two major considerations - signal integrity (EVM, transmit mask) requirements and out-of-band noise floor. Out-of-band noise in particular requires increasing attention due to co-existence considerations. A system may incorporate multiple radios supporting different standards. A scenario may be envsioned where a WiFi radio is co-existing with a cellular band (say GSM) radio. WiFi is a time division duplex (TDD) system. Hence, its own receiver is off, when the transmitter is on. However, the receiver of a GSM-1900 radio may be on. If the noise of the 2.4GHz WiFi radio at 1900MHz is high, it may reduce the sensitivity of the GSM receiver. The problem is particularly germane to such RF-DACs since the output noise is dominated by quantization noise. Hence, a significantly high resolution on the amplitude and phase path is required to keep the out-of-band noise low.

In order to determine the required resolution, system level simulations were carried out using the ADS Ptolemy simulator. Fig. 5.20 shows how the EVM varies as the number of bits on the amplitude path is increased, while the phase path resoltuion is held fixed at 10 bits. The simulation is performed for two different peak-to-average power (PAPR) ratios - 6dB and 9dB.. Theoretically, 802.11g signal has 52 subcarriers and a maximum PAPR of 17dB. However, such a peak is very rare and the PA average efficiency would degrade if such a high PAPR is allowed. Hence, the PAPR is reduced in the baseband before the amplitude and phase signals are generated. From Fig. 5.20, we see that for typical values of PAPR, the number of bits required by an ideal transmitter to meet the WLAN EVM requirements is about 4. Of course, non-linearity and other non-idealities in the transmitter will degrade the effective number of bits (ENOB). Still, a 6 bit implementation has enough margin, even

Figure 5.21: Variation in the noise density at 200MHz offset from the carrier as a function of the amplitude resolution (the phase path resolution fixed at 10-bits).

for 6-dB PAPR case.

Unfortunately the out-of-band noise floor is rather high for this relatively low resolution. Fig. 5.21 shows how the out-of-band noise varies as the amplitude resolution is increased (assuming a fixed oversampling ratio of 50, which will be discussed later). In this work, we targeted out-of-band noise better than -140dBc/Hz without any on-chip or external filtering. Hence, 8 bits of resolution were chosen on the amplitude path. Lower noise floor would, however, necessitate higher resolution or higher oversampling.

The phase resolution choise is also impacted by similar considerations of EVM and out-of-band noise. Fig. 5.22 shows the impact of phase resolution on the EVM for two different amplitude resolutions and PAPR. It is seen that for a given amplitude resolution and PAPR, EVM improvement is marginal if the phase resolution is increased beyond 8 bits. However, Fig. 5.23 shows how the out-of-band noise continues to go down with increased resolution. In order to achieve the targeted noise performance, 10 bits of resolution on the phase path and 8 bits on amplitude path were chosen.

5.2.3 Power Back-Off and Efficiency

As shown in Fig. 5.2, in such a digitally-modulated polar transmitter, the amplitude information is used to turn on and off unit cells, just like in a DAC. In this work, the unit cells comprise of inverse Class-D PAs, which were described in the previous section. The $2.5mm$ of total transistor width in our design is divided proportionally between the sub-elements. A segmented architecture is employed here, where the top four MSBs are thermometer-coded and the four LSBs are binary coded. Thus we have 15 unit cells, corresponding to the thermometer bits (T_1 - T_{15}) and 4 binary-weighted (B_0 - B_3) cells. These sub-PAs are combined
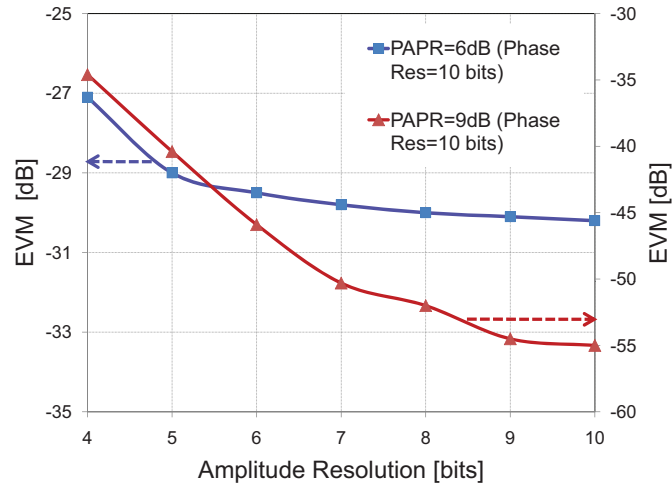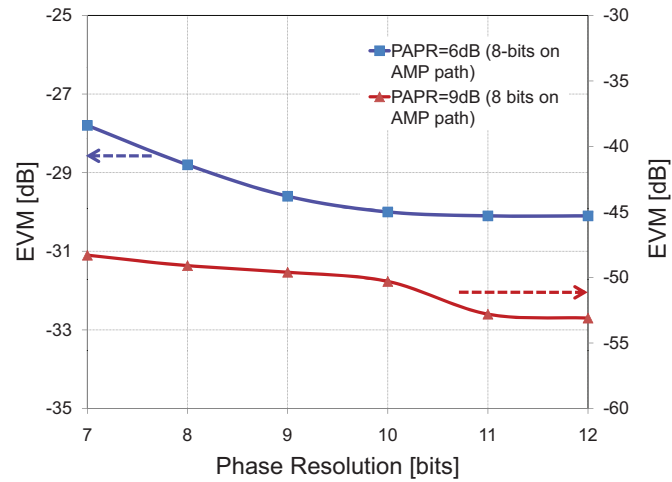
Figure 5.22: EVM of an ideal digitally-modulated transmitter while transmitting 802.11g 60QAM data. The variation in EVM with the phase resolution is shown for two PAPR cases - 6 dB and 9 dB.
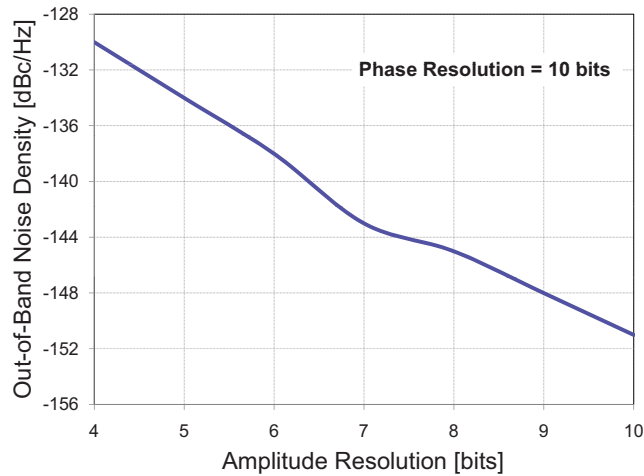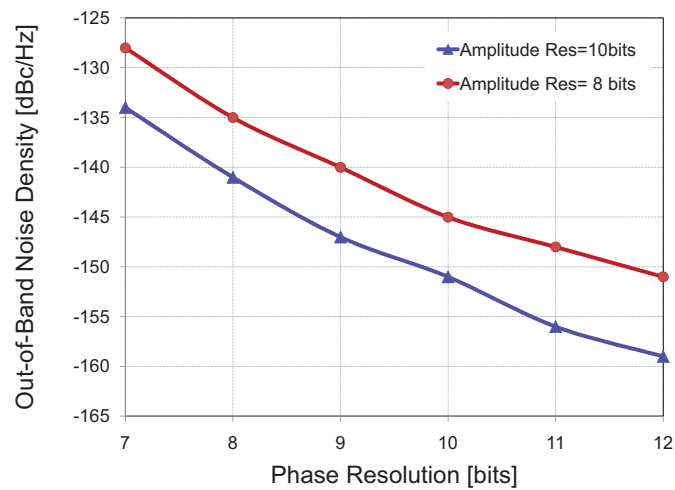

Figure 5.23: Variation in the noise density at 200MHz offset from the carrier as a function of the phase resolution (the amplitude path resolution fixed at 10-bits).
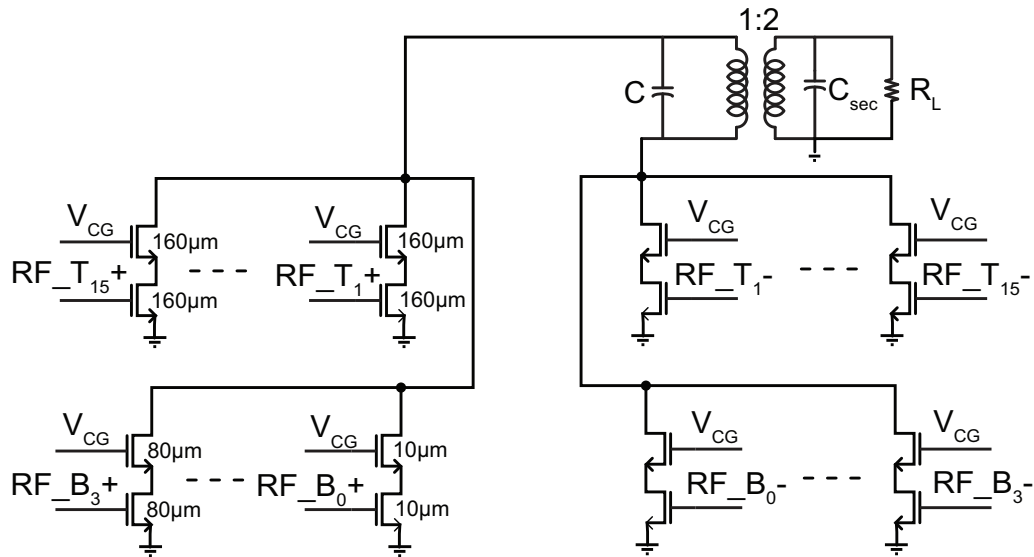
Figure 5.24: RF DAC using Class-D^{-1} PA as unit cells.

in the current domain by simply shorting their drains together in order to perform as an RF-DAC, as shown in Fig. 5.24.

The unit cells in the RF DAC are turned on or off by the amplitude codeword. Previous work had proposed switching the gate of the cascode transistor in the unit cell to turn off the cell. A simplified version of such an arrangement is shown in Fig. 5.25 for a single unit cell. Although using the digital amplitude bit (B_0) to directly control the gates of the cascode transistors is a convenient way to turn cells on and off, it can suffer from reliability issues. Let us assume that the 8-bit amplitude codeword value is 254 (0-255 being the full-scale range). Now, all the unit cells except one LSB cell is on. Since all the cells are on, the drain voltage swing is high, close to 2.4 times V_{DD}. However, the gate of the cascode transistors in the LSB cell is being held at zero (B_0 is low to turn the cell off). As a result, the gate-drain voltage across the transistor can exceed reliability limits.

In order to avoid this issue, a different switching scheme is proposed in this work. Since the phase-modulated signal is constant envelope, the common-source transistors in the unit PA cells can be driven by digital inverters. Furthermore, we can combine the amplitude and phase information using digital NAND gates, as shown in Fig. 5.26. When the amplitude bit (B_0) is high, the phase-modulated RF signals (PHA_p/PHA_n) pass through. When the amplitude bit is low, the unit cell is turned off. As is seen in the figure, this combination using NAND gates is done even before the driver inverters. An advantage of combining before the PA drivers is that the power of not only the PA, but also the drivers back-off gracefully with the amplitude codeword. This improves average transmitter efficiency.

Let us now see that how the efficiency of the PA backs-off with the output amplitude. The drain efficiency is the ratio of the power delivered to the load (P_{out}) to the power drawn
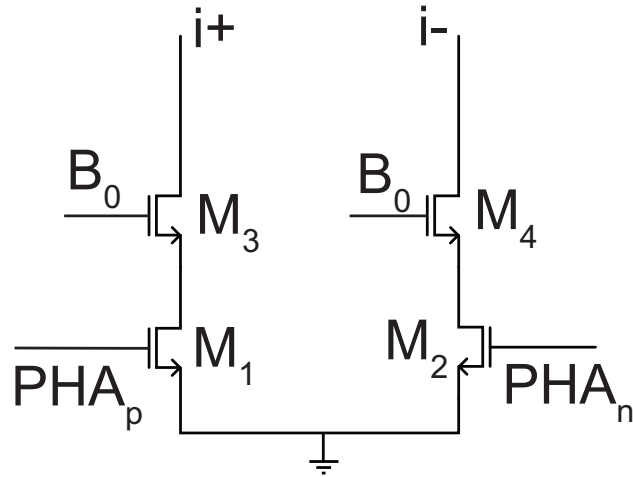
Figure 5.25: Switching an unit cell on/off through the cascode gate.
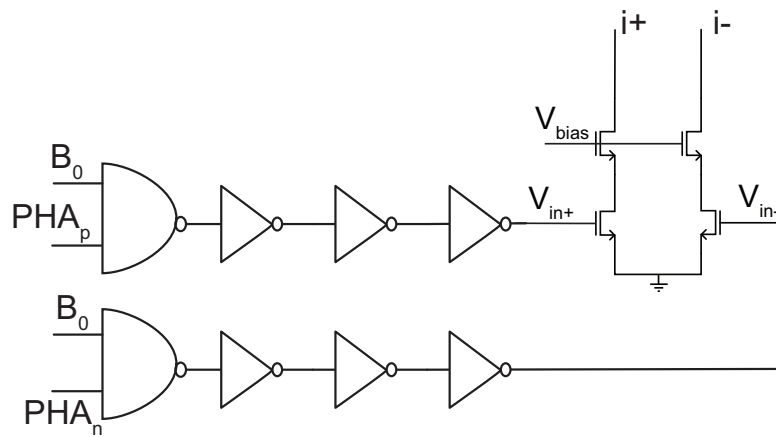
Figure 5.26: Combining amplitude and phase information using digital NAND gates.
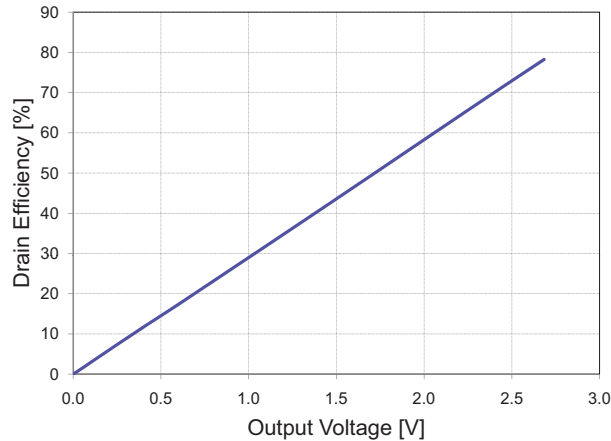
Figure 5.27: Linear back-off of drain efficiency with the output amplitude.

from the supply (P_{DC}):

$$\eta_D = \frac{P_{out}}{P_{DC}} = \frac{V_{out} \cdot I_{out}}{V_{DD} \cdot I_{DC}} \tag{5.8}$$

As the amplitude codeword reduces from its peak value (255 in this design), the number of unit cells that are turned on reduce proportionally. Hence, the output current (I_{out}) reduces. However, since the unit PA does not have any static power consumption, the DC current (I_{DC}) also reduces, proportional to I_{out}. Since the load impedance (R_L) is held fixed, V_{out} reduces proportional to I_{out}. But, since there is no supply modulation, V_{DD} is fixed and does not track V_{out}. Thus, the efficiency reduces linearly with V_{out} or equivalently, proportional to $\sqrt{P_{out}}$. Simulation results confirm such a response (Fig. 5.27). Such an efficiency back-off characteristic is superior to what can be obtained from a Class-A power amplifier, but is similar to a Class-B PA. It is then justified to question the advantage of such a digitally-modulated architecture over a Class-B PA. There are three aspects which make such an architecture beneficial over a Class-B analog power amplifier. First, a Class-B PA can seldom be used in a wireless transmitter due to its high phase non-linearity. A Class AB power amplifier is the more common choice, which has worse efficiency back-off characteristics. Secondly, such a mixed-signal architecture opens up the opportunity of using a switching PA, which has higher efficiency than its linear counterpart. Thus, in spite of having a linear back-off characteristic, the peak point is higher in such PAs. Finally, this type of digital solution absorbs the D/A power into the PA, which helps to boost the overall transmit efficiency. It also has less passive matching networks, smaller area and better scalability than its analog counterparts. These advantages make such a mixed-signal solution attractive in nanoscale CMOS processes.

5.2.4 Amplitude and Phase Linearity

In the previous section, we have seen that the drain efficiency of the digitally-modulated PA backs-off linearly with the output amplitude. But, linearly with the output amplitude does not mean linearly with the amplitude codeword. An ideal digitally-modulated power amplifier should be perfectly linear, which means that the output current should increase linearly with the amplitude codeword. However, the behavior in a real implementation deviates from this since the output impedance of the RF DAC changes as a function of the amplitude code-value. The real part of the output impedance of the transistor array decreases linearly with the code-value. This creates a nonlinearity in the voltage developed across the load resistance. Let r_{ds} and I_{unit} be the output impedance and drain current of an unit cell, R be the transformed load impedance, and N be the number of cells that are switched on. For simplicity, let us assume that the DAC comprises of unit elements only, which means $0 \leq N \leq 255$. Then the output voltage (V_{out}) across the load resistor (R) is given by [33]

$$V_{out} = \frac{N \cdot R_L \cdot I_{unit}}{1 + NR_L/r_{ds}} \tag{5.9}$$

The numerator describes the desired transfer fucntion: the output voltage is proportional to the number of transistors being switched on. The presence of a denominator dependent on N clearly shows the distortion. This code-dependent impedance causes both amplitude (AM-AM) and phase (AM-PM) non-linearity at the output of the PA. Fig. 5.28(a) shows the simulated amplitude response (blue curve). Clearly, as the codeword increases, the net PA output impedance reduces, delivering less power to the load and causing a compressive characteristic. Similarly, a change in output impedance (both real and imaginary parts) causes amplitude dependent phase shift, as shown by the blue solid curve in Fig. 5.28(b). Such non-linearity in amplitude and phase response will cause spectral regrowth, violating the transmit mask requirements. The linearity can be improved by reducing the device sizes, and increasing their on-resistance. However, as discussed in the previous section, this will degrade the PA efficiency rapidly.

In this work, we employ static digital predistortion to remove the effects of these non-linearities. The nonlinear amplitude and phase transfer curves are measured and its inverse function stored in a correction table in the digital baseband. The predistortion scheme is shown in Fig. 5.29. The baseband cordic generator produces the amplitude (A) and phase signals (ϕ). The 8-bit amplitude word is mapped onto a new codeword (A_p) by the look-up table ($LUT1$). LUT1 thus corrects for AM-AM non-linearity. In order to correct for AM-PM distortion, another table ($LUT2$) is employed. Based on the new amplitude codeword (A_p), a phase correction word (ϕ_c) is generated, which is added to the original phase word (ϕ). The resultant word (ϕ_p) is the predistorted phase information.

In this predistortion scheme, two important things need to be considered. The first is that the predistortion scheme is memoryless. This means that the current predistorted word depends upon the current input only, and not on any previous inputs. It has been verified
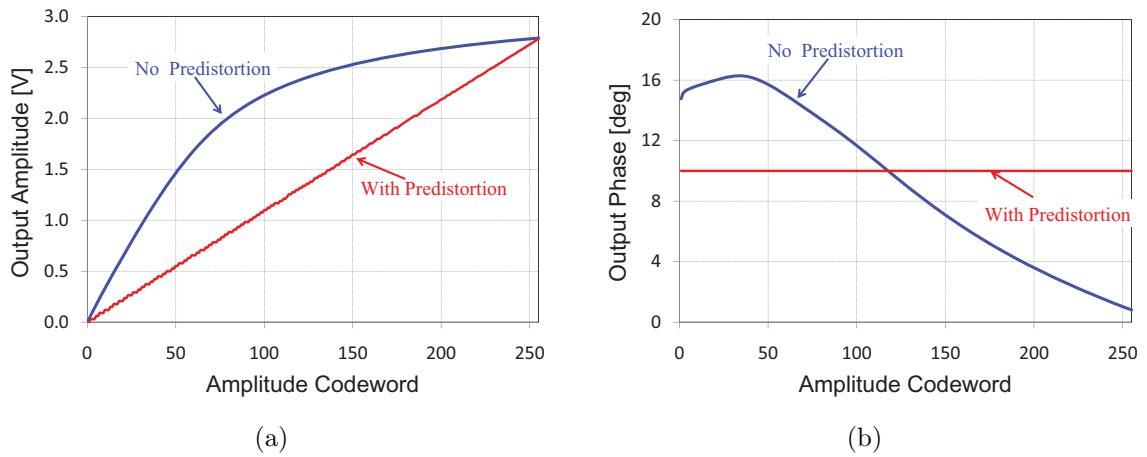
Figure 5.28: Simulated linearity of the PA array: a) AM-AM response b) AM-PM response.
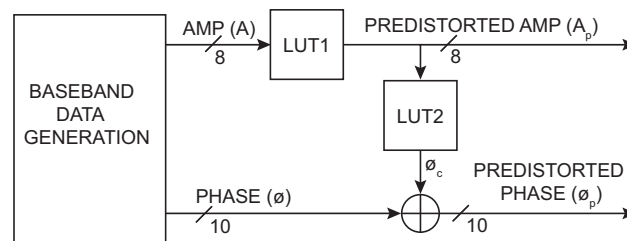
Figure 5.29: Look-up table based static digital predistortion.

through system simulations that the EVM degradation caused by making such a memoryless assumption is around 1 dB. Nevertheless, the low complexity and low power consumption of such a static predistortion scheme proves a win from a system point of view.

The second issue with predistortion is related to the interaction between amplitude and phase correction tables. It can be shown through simulation and measurement that the PA is fairly linear in the phase domain, which means it has no PM-PM distortion. Also, since the phase-modulated signal is constant envelope, the PA does not have PM-AM distortion as well. Under the assumption that the PA has no PM-PM and PM-AM distortion, the predistorted amplitude word (A_p) and phase correction word (ϕ_c) depend upon the original amplitude word (A) only, that is

$$A_p = f(A) \neq f(\phi) \tag{5.10}$$
$$\phi_c = f(A) \neq f(\phi) \tag{5.11}$$

Thus the two look-up tables LUT1 and LUT2 can be independent and have only 255 (corresponding to 8 amplitude bits) entries. However, if the PA had any PM-PM distortion, then ϕ_c in Fig. 5.29 would have been a function of both A and ϕ, which would have increased the size of the LUT enormously. Simulations have been performed to verify the effectiveness of the static predistortion scheme. The linearized amplitude plot is shown in Fig. 5.28(a) (red curve), while the phase response after correction is shown in Fig. 5.28(b). The simulated EVM for 7-dB PAPR is better than -30dB after predistortion, which meets the 802.11g requirements for all data rates and modulation formats.

5.2.5   Clocking Rates and Digital Filter

As explained earlier, out-of-band noise density is an important parameter in wireless transmitters. Just as increasing the amplitude resolution reduces the quantization noise floor, so too does oversampling. Since the baseband I and Q signals are 10MHz each, higher clock rate can be employed to spread out the quantization noise and reduce the noise density. In general, higher clock rate needs to be employed to do the cartesian to polar conversion, since such a conversion is non-linear and results in bandwidth expansion. In general, about $4 - 5X$ the I/Q bandwidth (10MHz in our case) is needed to preserve most of the signal energy [65]. Hence, a 80-100MHz clock rate would have been sufficient. But, nanoscale CMOS process offers us the advantage of higher switching speeds, which has been utilized in this work. In this design, the cartesian-to-polar conversion has been done at a clock rate of 200MHz and then the amplitude and phase signals are further oversampled to a clock rate of 1GHz. Combined with 8 bits of resolution on the amplitude path and 10 bits on the phase path, such a high clock frequency can ensure noise density at 200MHz offset better than -140dBc/Hz. It should be noted that while oversampling has been employed to reduce the noise density, no noise-shaping (delta-sigma) modulation has been employed here. While noise-shaping can reduce the in-band noise floor, thereby increasing the effective resolution,
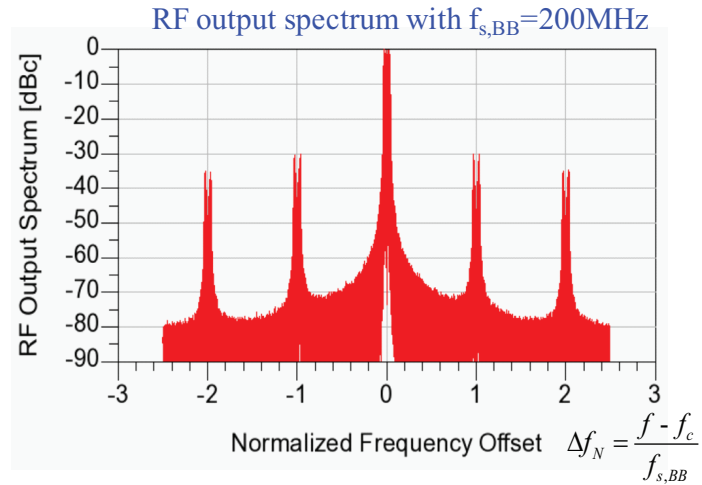
Figure 5.30: Simulated output spectrum with signal replicas at multiples of the baseband frequency (200MHz).

it has a detrimental impact on out-of-band noise and makes filtering requirements harder. Hence, only oversampling, but no noise-shaping is employed in this design.

The up-sampled 1GHz amplitude signal, however, still has spectral images at multiples of the baseband frequency of 200MHz. If the conversion to higher frequency is done by sample-and-hold, then the spectral images will be attenuated by sinc filtering. Fig. 5.30 shows the simulated output spectrum. Clearly, the images at multiples of 200MHz are visible. Depending upon the chosen clock frequencies and the resulting sinc filtering (due to zero-order hold), these spectral images may or may not violate the 802.11g spectral mask requirements, shown in Fig. 5.19. However, they can inject a lot of noise into another radio's receive band, exacerbating co-existence issues. Hence, a digital filter is needed which can attenuate these images to the noise floor. Such a filter needs to be low-power, since it is always on, independent of the transmitted power. The filter design will be described in the next section.

5.2.6 Implementation and Layout of the RF-DAC

A block diagram of the complete system developed so far is shown in Fig. 5.31. In this prototype, the baseband data generator, cordic algorithm and look-up tables for predistortion are implemented in an FPGA. The 8 amplitude bits at 200MHz are fed directly to the chip, while the phase bits are fed to an external DAC/modulator chip to create a 2.4GHz phase-modulated signal. On the phase path, all LO distribution and clock trees are integrated on-chip. On the amplitude path, digital filters, binary-to-thermometer decoder, NAND gates, driver inverters and PA array are all implemented on-chip.

Layout plays a critical role in the implementation of such an RF DAC. An array-based
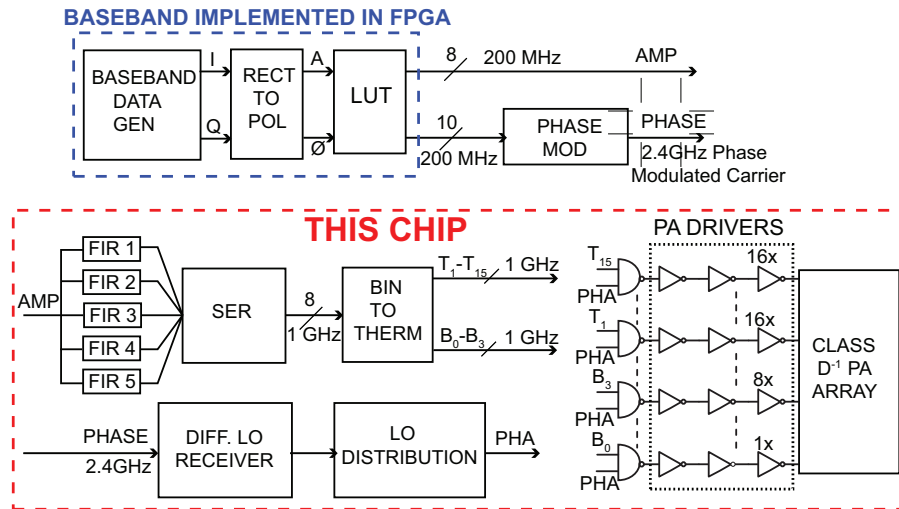
Figure 5.31: Block diagram of the mixed-signal polar transmitter.

technique has been adopted for the DAC layout. Its a 4×4 array, as shown in Fig. 5.32. As seen in the figure, the 15 thermometer cells, corresponding to the top 4 MSBs, occupy 15 of the cells, while the 4 binary-weighted elements are fitted in one square. The unit cell placement is disordered in order to reduce the effect of thermal and process gradients. Dummy cells were added to each side of the matrix to avoid boundary effects that could cause mismatch in the unit-cells [33]. Shielding was used intensively in the unit-cells to prevent the RF input signal from leaking to the RF output. The layout of the unit-cells was also optimized so that the 65nm density rules were automatically fulfilled, thus preventing the tiling procedure from causing mismatches across the matrix of unit-cells.

The transistors in the Class-D^{-1} PA are driven by rail-to-rail phase-modulated signals. In order to ensure that the signal is not degraded as it travels along the PA grid, the first driver is integrated into each unit cell. This ensures that signals with sharp slopes drive the PA cell for high efficiency.

The amplitude and phase-modulated RF signal are combined using digital NAND gates, as shown in Fig. 5.31. The 15 NAND gates corresponding to the thermometer elements and the 4 binary-weighted gates are also laid out in an array structure, just below the PA grid. It is important to ensure that the phase-modulated signal arrives at all the NAND gate inputs with as little skew as possible, since excess skew introduces EVM degradation. Hence, a clock tree has been implemented to distrbute the phase-modulated 2.4GHz signal generated off-chip to all the NAND gates. Tapered buffers have been used on the LO path, so that very low input power (less than $-10dBm$) is needed to drive the chip. The LO distribution is shown in Fig. 5.33. The leaves of the clock tress have been shorted to restrict the worst-case skew to the RC delay of the shorting wire.

In addition, cross-coupled inverters have been added in between the differential phase

Figure 5.32: Layout of the RF DAC.
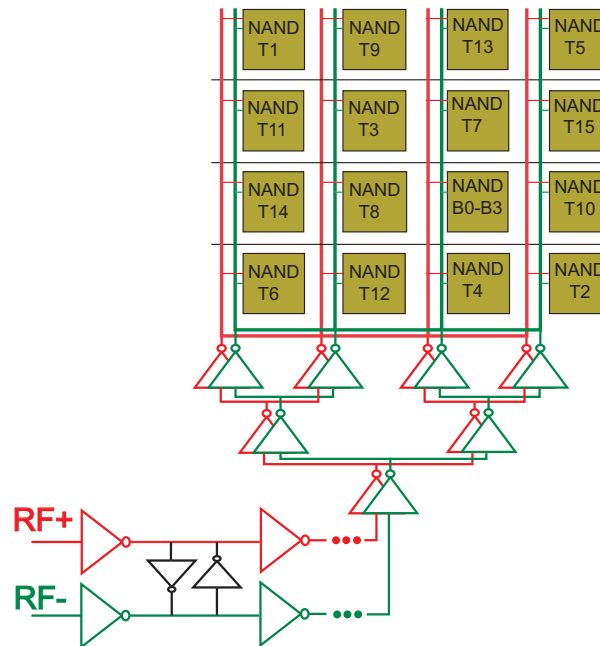
Figure 5.33: On-chip LO distribution network.
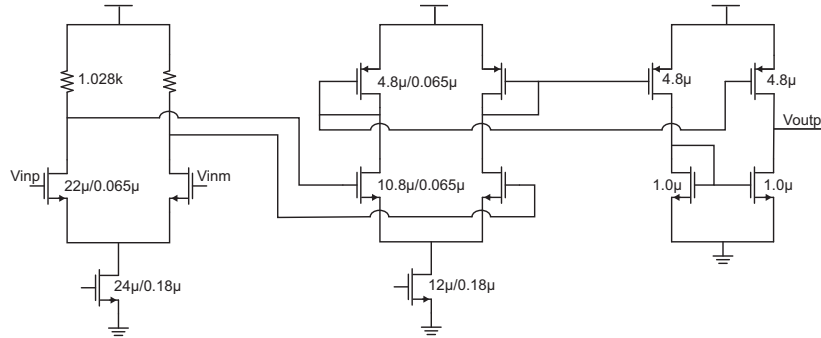
Figure 5.34: Differential receiver for phase-modulated RF signal.

signals. This ensures that signals remain complementary as they propagate down the chain. If the signals are not fully differential when they reach the PA, it causes crowbar current and efficiency reduction. Though not shown in Fig. 5.33, proper shielding between digital amplitude bits and RF phase signal is very important to minimize phase distortion.

Finally, since the phase-modulated RF signal is generated off-chip in this prototype, proper care needs to be taken to design the interface circuit to bring the RF signal on-chip. In particular, since the on-chip ground may bounce relative to the off-chip (board) ground, there is a possibility of phase distortion. Hence, a differential amplifier with high common-mode rejection is employed to receive the off-chip low-swing RF signal. The first stage of this clock receiver, shown in Fig. 5.34, is a resistively-loaded differential amplifier with low gain and very high common-mode rejection. The next stage provides more rejection, while also converting from CML-to-CMOS logic. Simulations have been performed to confirm that the phase-modulated signal can be received with high accuracy (less than $0.2°$) even in the presence of ground bounce.

5.3   Low-power Digital Filter Design

In the previous section, it has been explained that we need a digital filter on the up-sampled amplitude and phase paths to remove spectral images. A finite impulse response (FIR) filter is chosen due to the possibility of achieving linear phase response. It should be noted that the filter will attenuate spectral images at multiples of the baseband (BB) frequency ($200MHz$ in this prototype). But, the spectral images at multiples of the final clock rate ($1GHz$ in this prototype) will remain, only attenuated by sinc filtering. However, the resultant spurs in the PA output spectrum are far enough from the carrier frequency and can will be filtered out by the PA output network as well as any harmonic-reject filter.

In the presence of AM-AM and AM-PM distortion, the interaction between the filtering and the look-up table based pre-distortion should be considered carefully. Ideally, the lower-rate ($200MS/s$) BB data should be first interpolated to a higher-rate ($1GS/s$) data before
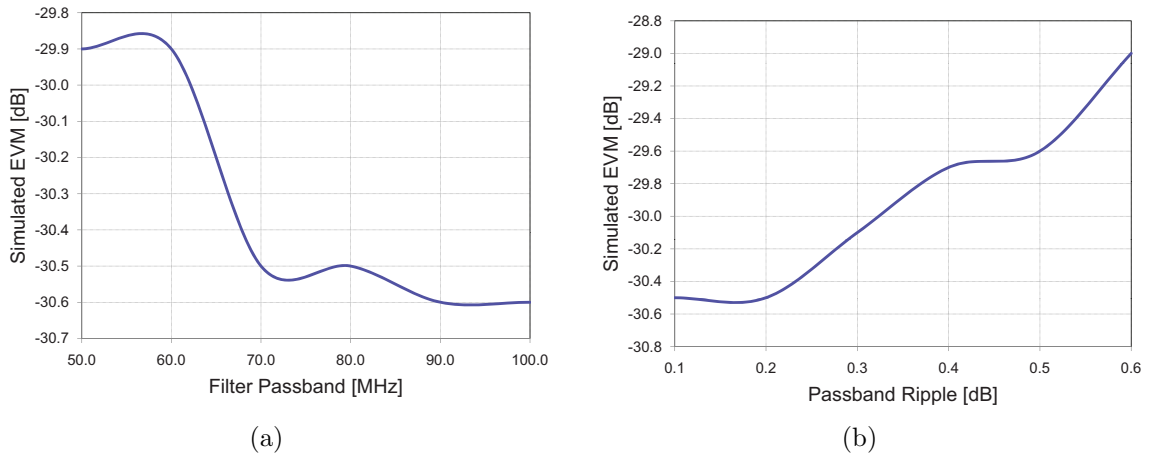
Figure 5.35: a) Effect of the filter passband on system EVM. Here, the filter passband ripple is assumed to be smaller than $0.3dB$. b) Effect of the filter passband ripple on system EVM. Here, the filter passband is assumed to be $80MHz$.

predistortion ("filter-first" arrangement). In this case, however, the LUT implementing the static pre-distortion has to run at a clock rate of $1GHz$ and will thus be power-hungry. For low power consumption, it is more desirable to pre-distort first and then do the filtering ("filter-last" arrangement), which allows LUT to run at a lower rate of 200MHz.

In order to allow this configuration, we have to make sure that the filtering after pre-distortion faithfully preserves the signal integrity. The condition which ensures that the "filter-last" arrangement will work is related to the filter passband. Predisortion, being a non-linear function, results in bandwidth expansion. If the filter bandwidth is chosen so that it will pass the pre-distorted signal without any distortion, then the "filter-last" scheme works well.

An alternative way to look at this is in the time domain. In such an over-sampled system, the signal changes slowly every 1GHz clock cycle. In the time domain, the difference between consecutive 200MS/s codewords (which themselves originate from 10MHz I/Q signals) is usually small. As a consequence, it makes little difference to predistort this 200MS/s data and then do interpolation filtering as compared to reversing the order between the two. We have verified through system simulations that the EVM degradation due to "filter-last" arrangement is negligible, even for an EVM level of $-35dB$.

In our design, the amplitude-path FIR filter is always on, regardless of the PA output power. To introduce negligible overhead to the whole TX, especially at backed-off power levels, the filter power consumption should be minimized. To ease the power-constrained filter design, zero-order hold (ZOH) is employed in the 5x up-sampling (from 200MS/s to 1GS/s) preceding the filter, which offers around 19dB attenuation at 180MHz offset. The
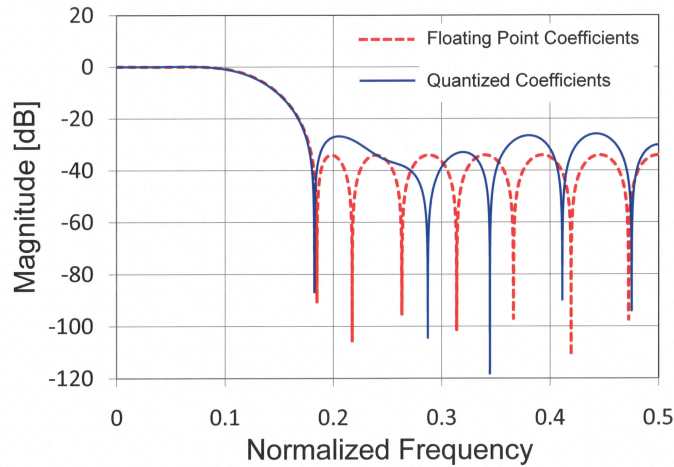
Figure 5.36: Simulated magnitude response of the filter with and without coefficient quantization.

effect of the in-band "sinc" droop resulting from zero-order hold is small (less than $0.3dB$ in EVM). In order for the filter to have negligible impact on system performance (mainly EVM), the filter specifications need to be chosen carefully. Simulations (Fig. 5.35) show that a passband $\geq 70MHz$ and passband ripple $\leq 0.25dB$ does not degrade the system EVM. Since the filter has to attenuate the spectral images at 200MHz and its multiples to the noise floor, it needs a stopband attenuation of $\geq 30dB$ starting from $180MHz$. From Matlab, it has been derived that a 19-tap 1Gs/s filter is needed to achieve the desired specifications.

A multiplier-less filter architecture has been adopted to conserve power. It is well known that if each filter coefficient value is a sum of signed power-of-two (SPT) terms, the filter can be conveniently implemented as a combination of adders with necessary shift operations. The complexity of the filter now depends only on the number of SPT terms required to represent a coefficient. Thus the primary goal in quantizing the coeficients is to have minimal total number of SPT terms. In doing this optimization, a cost-based algorithm, similar to that presented in [69], was adopted. This algorithm allows different number of SPT terms to be allocated to different coefficients. After the search, a minimal of 24 SPT terms have been used to represent all coefficients for our 19-tap filter. The filter magnitude response with and without coefficient quantization is shown in Fig. 5.36.

To facilitate IC synthesis and save power, the generic 1GS/s filter is designed as five parallel 200MS/s sub-FIR filters followed by a 5-to-1 1Gs/s serializer. Since the filter input signal is obtained from 5x up-sampling (sample and hold) of the 200MS/s amplitude data,
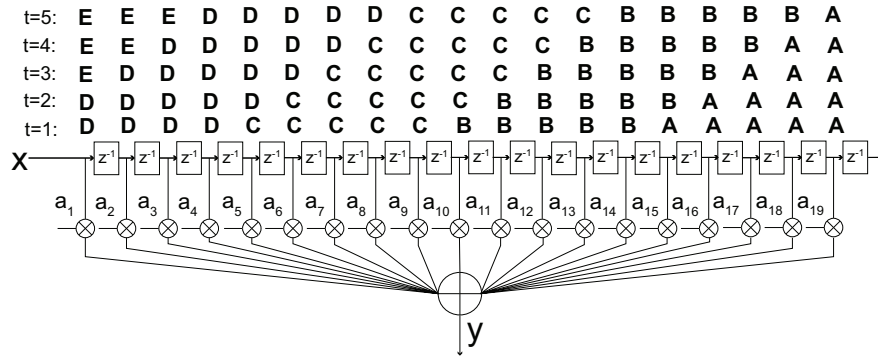
Figure 5.37: Concept of coefficient grouping in a parallelized FIR filter.

every five consecutive data bits at the filter input are identical. This fact can be exploited to greatly reduce the adder count and power.

In order to understand how this can be achieved, assume that the 200 Ms/s time-domain consecutive data samples are A, B, C, D, E, where A arrives first. After 5x upsampling, the five consecutive samples are identical at filter input, as shown in Fig. 5.37. Therefore, for each of the individual sub-FIR filters in the interleaved implementation, all coefficients with the same input data can be lumped together into a single term. For example, assume the "first" sub-filter corresponds to time instant t=1 in Fig. 5.37. The difference equation representing that sub-filter at that time instant would be

$$y = \sum_{k=0}^{3}(a_k \cdot D) + \sum_{k=4}^{8}(a_k \cdot C) + \sum_{k=9}^{13}(a_k \cdot B) + \sum_{k=14}^{18}(a_k \cdot A) \tag{5.12}$$

However, the above equation can also be written as

$$y = 0 \cdot E + \left(\sum_{k=0}^{3} a_k\right) \cdot D + \left(\sum_{k=4}^{8} a_k\right) \cdot C + \left(\sum_{k=9}^{13} a_k\right) \cdot B + \left(\sum_{k=14}^{18} a_k\right) \cdot A \tag{5.13}$$

Thus, we see that because of repeating samples, it is possible to group five consecutive coefficients. The above difference equation can be readily implemented by a 4th order FIR filter

$$y[n] = \sum_{k=0}^{4} b_k x[n-k] \tag{5.14}$$

where the coefficients are given by

$$b_0 = 0 \tag{5.15}$$
$$b_1 = a_0 + a_1 + a_2 + a_3 \tag{5.16}$$
$$b_2 = a_4 + a_5 + a_6 + a_7 + a_8 \tag{5.17}$$
$$b_3 = a_9 + a_{10} + a_{11} + a_{12} + a_{13} \tag{5.18}$$
$$b_4 = a_{14} + a_{15} + a_{16} + a_{17} + a_{18} \tag{5.19}$$
$$\tag{5.20}$$

Similarly, the filter outputs at time instants t = 2, 3, 4 and 5 can be expressed as the sum of each 200Ms/s data sample (A, B, C, D, E) multiplied with a different set of coefficients, with these coefficients repeating every 5 up-sampled cycles. This observation allows us to implement the generic interpolating 1GS/s FIR filter as five parallel 200MS/s sub-FIR filters followed by a 5-to-1 serializer. Unlike conventional parallelism, our architecture dose not requires 5X increase in hardware cost since the total number of SPT terms representing the sub-filter coefficients (b_i) is now greatly reduced by the coefficient grouping.

To verify the FIR response in measurements, a 1MHz sine wave is fed on the amplitude path, sampled at 200MHz. The phase-modulated signal is a single-tone carrier at RF frequency. The highest image measured around 200MHz offset from carrier is about -55dBc and matches simulation. The measured power consumption of the FIR and the serializer together is only 1.4mW.

5.4   Measurement Results

5.4.1   Board Layout and Measurement Setup

The chip has been fabricated in 65nm digital CMOS process without any RF options. The chip micrograph is shown in Fig. 5.38. It measures $1.7mm \times 1.8mm$ and is mostly pad-limited. All the I/Os have been bonded to a PCB (Fig. 5.39). The board features four conductor layers, with the two middle planes devoted to ground and supplies, and is built of FR4. FR4 has a loss of about 0.03 dB/cm/GHz. Hence, the traces of the RF signals, particularly the PA outputs, have been kept small. The 8 amplitude lines are laid out such that their lengths, and therefore their delays, are almost identical. The signals are also retimed on-chip. Since all of the matching circuits are integrated on-chip, there is no RF matching network on the board. The only passives on the board are large capacitors for supply and bias line bypassing at lower frequencies, since such capacitors are too big to be integrated on-chip.

The measurement setup is shown in Fig. 5.40. The baseband amplitude (A) and phase signals (ϕ) are generated using a Virtex-V FPGA, embedded within the ROACH platform [70]. Eight LVDS signals at 200MHz, corresponding to the amplitude bits, are fed directly
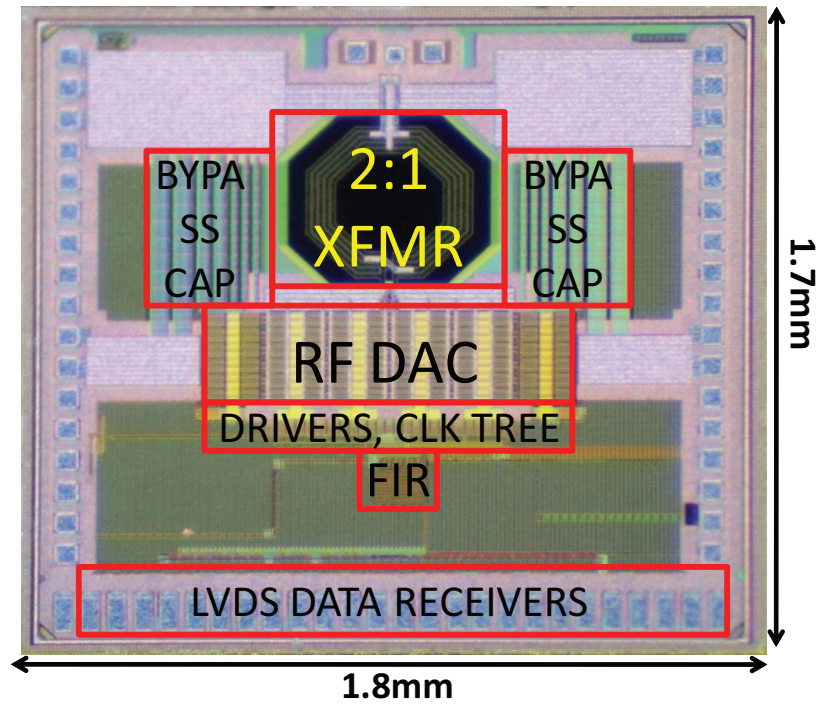
Figure 5.38: Chip micrograph of the 65nm mixed-signal polar PA.
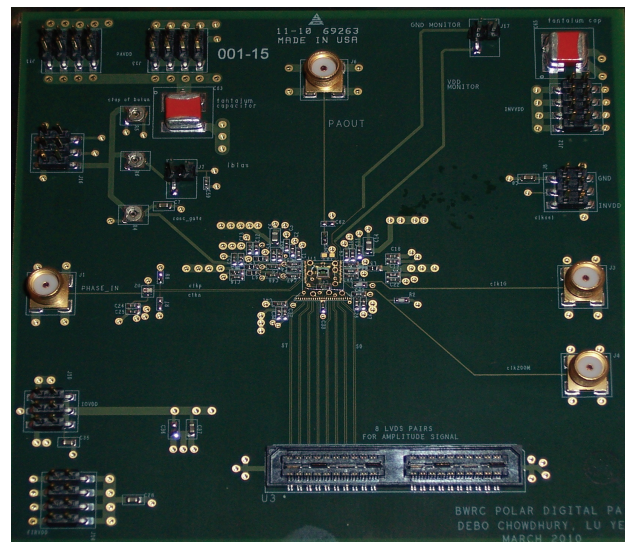
Figure 5.39: CMOS die mounted on an FR4 PCB.
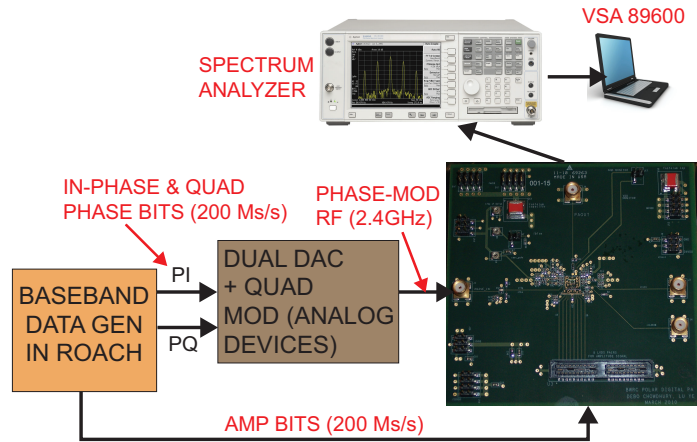
Figure 5.40: Measurement setup of the polar transmitter.

from the FPGA to the chip and interfaced via high-speed LVDS receivers designed on-chip. Two back-to-back flip flops have been used to retime the data on-chip. The 10 phase bits are converted to in-phase (ϕ_I) and quadrature (ϕ_Q) components and fed to an external 16-bit dual-channel DAC (AD9779A). The analog outputs from the dual-channel DAC are fed to a quadrature modulator (ADL5372), which produces the phase-modulated RF signal. The dual-channel DAC and quadrature modulator are housed in an evaluation board AD9779A-EB. The software for the AD9779A-EB board allows choice of rise or fall and also offers a delay element, thus enabling adjustment of the clocking delay [31]. Finally, the phase-modulated 2.4GHz signal from the modulator is converted to differential form using a balun on the board and interfaced to the chip via the differential clock receiver, described in an earlier section.

5.4.2   Measured Results

The first step in measuring the performance of the prototype amplifier was to confirm the RF performance with a single-tone sine wave. In the first measurement, the phase bits coming from the FPGA were held constant and thus the output of the external quadrature modulator was a fixed-phase sine wave at RF frequency. The frequency of the sine-wave was varied and the output power plotted (Fig. 5.41). As seen from the plot, the output power variance is less than 1dB over a bandwidth of 800MHz. Such a high bandwidth is possible since the digital PA has only one output matching network and no inter-stage inductors or transformers. Such a wide bandwidth also opens up the possibility of the use of this transmitter as a multi-mode radio. The peak-power frequency is shifted slightly from a desired value of 2.4GHz to 2.25GHz, which may be due to the bond-wires at the PA output. It should be noted that no tuning was performed on the board to adjust the output matching
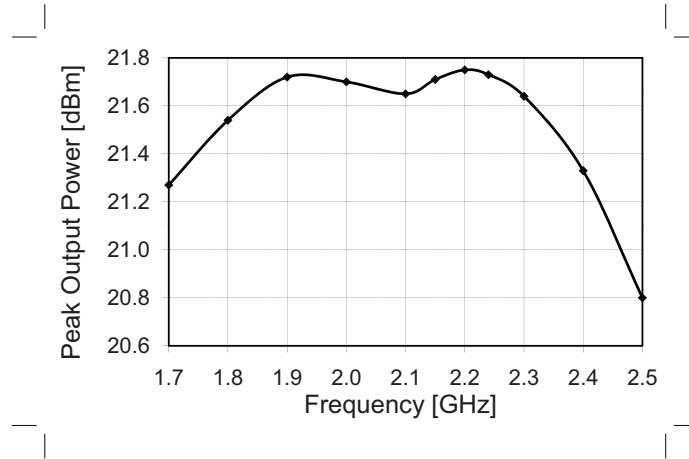
Figure 5.41: Measured PA output power over frequency.
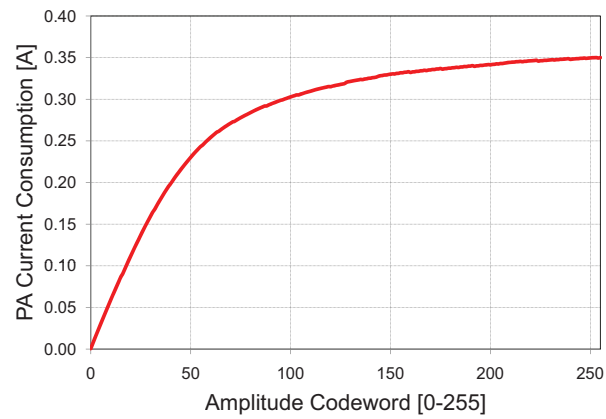
Figure 5.42: Measured DC current consumption of the PA array.
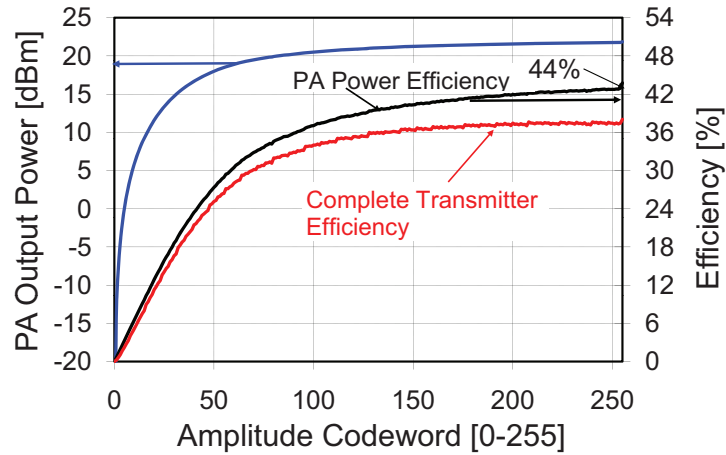
Figure 5.43: Measured PA output power, drain efficiency and transmitter efficiency at 2.25GHz for different amplitude codewords.

network.

Next, the RF signal is still held at a fixed phase, but the amplitude codeword is swept from 0-255. The DC current drawn by the inverse-D PA is plotted in Fig 5.42 and matches simulation well. The measured output power and efficiency is plotted in Fig. 5.43. Operating from a 1-V supply, the PA achieves a peak output power of 21.8dBm, which is about 0.5dB lower than simulation. The peak efficiency (including matching network loss) is 44%, while the PAE (including power of all drivers, decoders, LO distribution and clock receivers) is 38%.

The peak power consumption of all the digital circuitry (PA drivers, NAND gates, clock distribution, FIR) is only 50mW, showing the benefit of such a mixed-signal architecture. The amplitude dependent current consumption of the digital driver inverters is shown in Fig. 5.44. In the plot, however, it is seen that the current is not monotonically increasing with amplitude codeword. This is not a measurement artifact, but is by design and can be understood as follows.

The arrangement of the PA array and drivers is shown in Fig. 5.45. As described in a previous section, the first driver inverters are integrated locally with the core PA cells so that the PA can be driven by sharp digital pulses. In this segmented architecture, the 4 LSB cores (B_0-B_3) are binary weighted and the top 15 cells (T_1-T_{15}), corresponding to the 4 MSBs, are thermometer coded. Of course, the first drivers are also similarly weighted, as shown in Fig. 5.45. The second driver array is outside the PA-driver1 grid and thus long wires, traversing the height of the PA-driver1 grid, are needed to connect the output of driver-2 to the input of driver-1. Assume that the cells in driver-2 array are also weighted appropriately, i.e. the bottom 4 LSB devices are binary weighted and the top ones thermometer weighted. However, the parasitics of the wires connecting the binary cores versus those connecting the
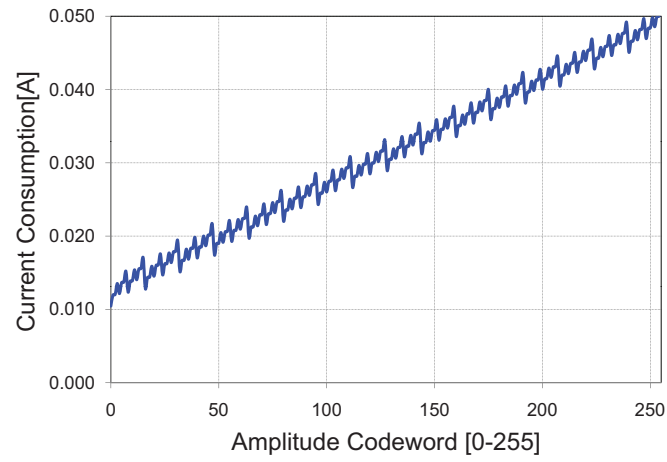
Figure 5.44: Measured DC current consumption of the digital inverters.
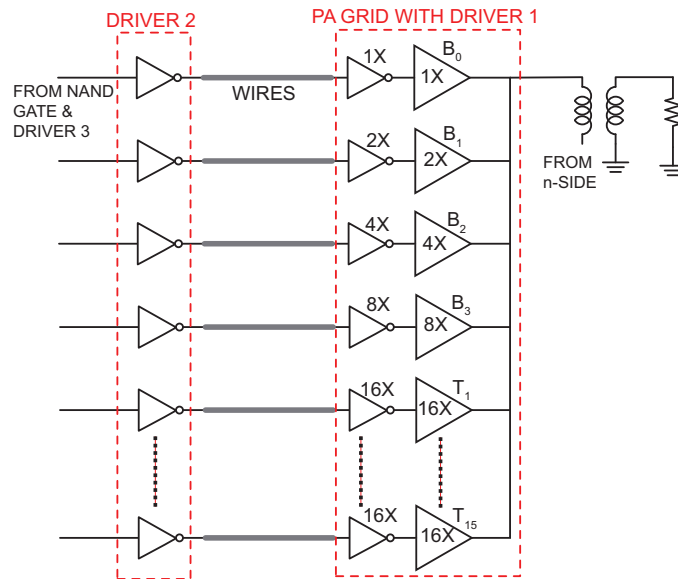
Figure 5.45: Layout of the PA grid with drivers.
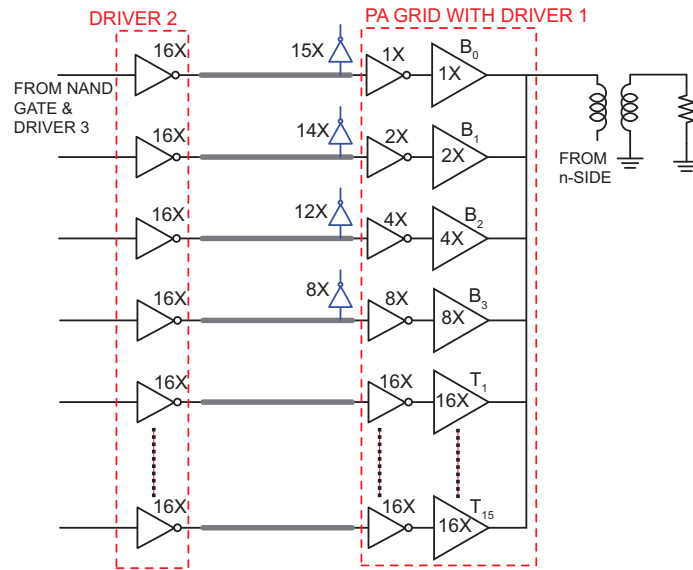
Figure 5.46: Low-skew layout technique for the PA array with drivers.

thermometer-coded cores are almost the same, as shown in Fig. 5.45. In other words, it is very hard to make the wire parasitics track the weight of the amplifier. This means that the signals in the binary cells might have a skew relative to each other and also relative to the thermometer cells.

To avoid this, we have used dummy loads, as shown in Fig. 5.46. Here, all the cells in driver 2 array are of the same size ($16X$). In order to keep the delay same, dummy inverters are added at the gate of driver-1 cells. The dummies are so sized that all the cells in driver-2 array see the same loading (including wire parasitics). At the expense of little extra power consumption ($< 0.5mW$), such a technique helps to keep the skew between the different signals very low. The addition of dummy loads and the non-binary weighting of cells in driver-2 array is the reason why the current profile in Fig. 5.44 is not monotonic. However, this is not a problem either for PA operation or for predistortion, since the outputs of the dummy inverters are not connected to the PA input.

Next, the distortion of the PA was measured. Due to the code dependent output impedance, the PA exhibits AM-AM and AM-PM distortion, as shown in Fig. 5.47. Measurements have verified that the PA has negligible PM-PM and PM-AM distortion. The top eight MSB bits of the phase codeword were varied and Fig. 5.48 shows the measured linear PM-PM response. The phase difference between successive codewords is shown in Fig. 5.49 and has an error of only 0.1°, which could be the result of measurement inaccuracy. In such a case (no PM-PM distortion), the AM-AM and AM-PM errors have been corrected using independent look-up tables (LUT).

To illustrate the effect of predistortion, a 2.25GHz carrier was modulated by a low-
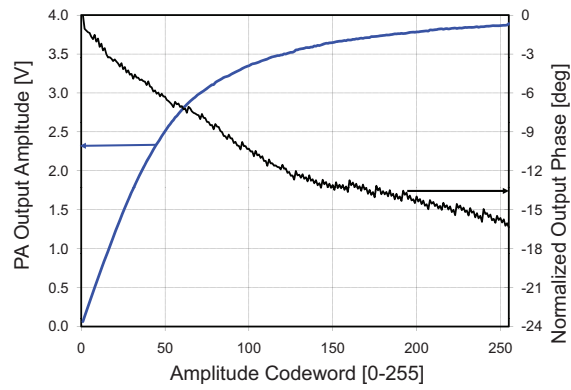
Figure 5.47: Measured AM-AM and AM-PM characteristics.
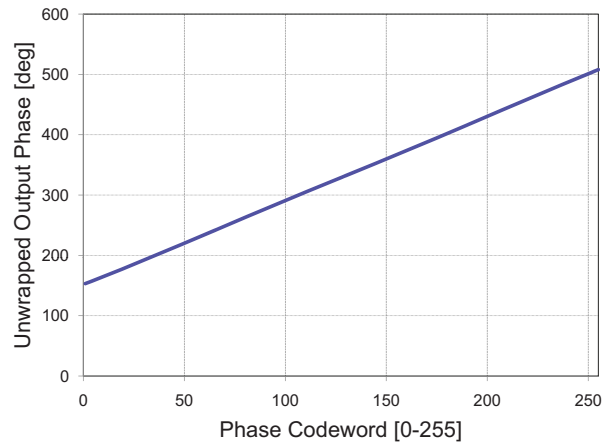
Figure 5.48: Measured PM-PM characteristics.
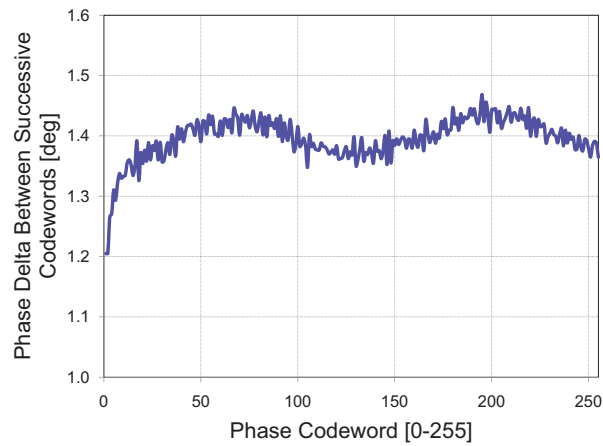
Figure 5.49: Measured phase difference as the phase codeword is incremented.
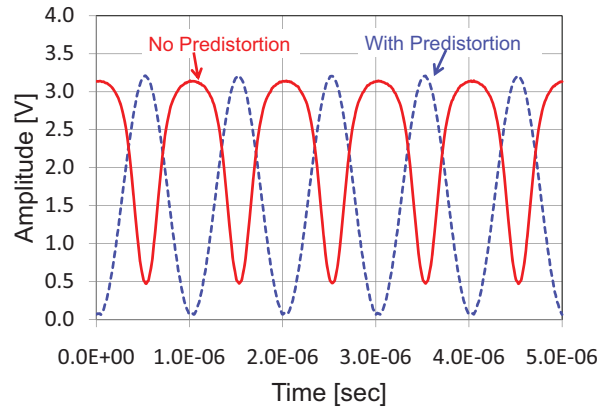
Figure 5.50: Effect of amplitude predistortion.
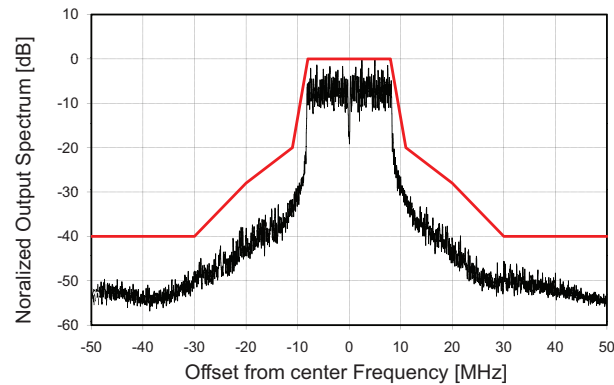
Figure 5.51: Output spectrum when transmitting 802.11g 54Mbps 64-QAM OFDM data.

frequency (1MHz) sine wave to create an AM-modulated signal. This was passed through the PA and the output is a distorted sine wave due to compressive non-linearity of the PA. Next, static pre-distortion is applied on the amplitude path through the LUT and the test again repeated. The results are shown in Fig. 5.50, which clearly shows how predistortion linearizes the PA.

Finally, the transmitter was tested with 802.11g 54Mbps 64-QAM WLAN OFDM signal. The alignment between the amplitude and phase signals was achieved using clock delays in the digital domain. The transmitter output spectrum (Fig. 5.51) meets WLAN requirements with high margin. The full spectrum (Fig. 5.52) shows that the FIR has reduced the spectral images to nearly the noise floor. At 200MHz offset from the center frequency, the measured noise is -120dBm/Hz without any external filtering. The measured EVM is -28dB, while transmitting an average power of 14dBm with 18% average PA efficiency. The improvement
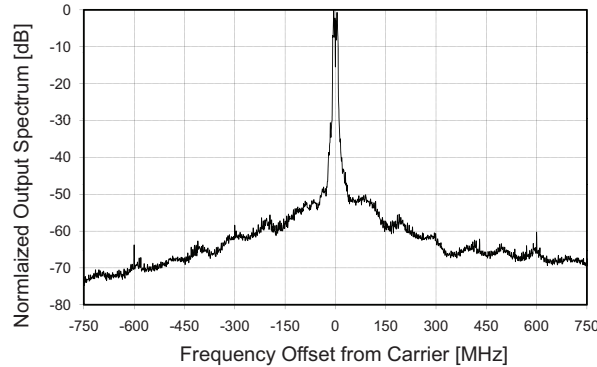
Figure 5.52: Far out spectrum when transmitting WLAN data.

Table 5.2: Performance of the mixed-signal PA in comparison with a commercial WLAN power amplifier

Metrics	This PA	RFMD 5622 PA
Supply Voltage (V)	1	3.0-3.6
Technology	Digital CMOS (no UTM)	InGaP GaAs HBT
Average output power (dBm)	14-15	13 (V_{DD}=2.8V)
		18 ($V_{DD} = 3.3V$)
Average Efficiency (%)	18	20 (V_{DD}=3.3V)
		10 ($V_{DD} = 2.8V$)
EVM (%)	3.6	3.0 (min) - 4.0 (max)
Quiscent power (mW)	20	165
Test Setup	802.11g 54Mbps	802.11g 54Mbps

obtained from phase predistortion (to compensate for AM-PM non-linearity) is lower than what simulations predicted. This could be due to the bandwidth limitations in the external quadrature modulator used for testing purposes. With the phase predistortion fully in effect, simulations predict an output power of about 15dBm with greater than 20% efficiency.

The average power consumption of the entire transmitter is only 150mW, which is much lower than traditional WLAN transmitter power consumption. Simulations show that with the use of ultra-thick metal (UTM) for transformer layout and power supply distribution, the average efficiency can be boosted to around 25%.

Table 5.2 shows a comparison of the performance of the designed prototype with that of a commercial power amplifier implemented in an exotic compound semiconductor process [71][1] The output power of the CMOS prototype is lower than the maximum available from the commercial PA. However, no power combining has been employed in this design. The use

---

[1]The performance metrics of the commercial PA has been inferred solely based on the data sheet available online. No other information was available.

of a power combiner can help achive 3dB higher output power with negligible degradation in performance. The efficiency and linearity (as measured by EVM) of the CMOS PA is on par with the compound semiconductor PA. Thanks to CMOS, the static power consumption is also much lower.

5.5   Summary

This chapter has presented the design of a fully-integrated digitally-modulated polar transmitter. Such an architecture has the potential of breaking the efficiency-linearity trade-off inherent in linear Class A/AB power amplifiers. In addition, such an architecture offers a more scalable solution and is well-poised to take advantage of technology scaling. In order to leverage the real benefit of this system, an efficient, switching power amplifier is desired. This work has proposed an inverse Class-D PA as a compact yet efficient solution. The PA has been designed as an RF DAC and integrated with the rest of the digital circuitry. Low-power pre-filtering is performed to enable co-existence. Operating from a 1-V supply, this fully-integarted 65nm CMOS PA has a high efficiency of 44% while delivering 22dBm of output power. The use of digital predistortion enables the PA to meet the stringent linearity requirements of modern wireless standards like 802.11g with 18% efficiency.

Chapter 6

Conclusion

6.1 Thesis Summary

Wireless has been the buzzword of the last decade. Explosive growth in wireless communication market has led to consumer demand for low-cost, small form factor and low-power terminals. High level of integration is the most effective way to provide such a solution. The power amplifier, which is often implemented in niche processes, is one of the critical blocks that still needs to be integrated with the rest of the radio in commercial solutions. This thesis studies key issues in the design and implementation of fully-integrated efficient CMOS power amplifiers at RF and 60GHz.

An important challenge in any PA is the design of the output matching network. The area and power efficiency of such matching networks are critical to the PA design. Typically, distributed circuits like transmission lines have been used as matching networks at mm-wave frequencies. Though such transmission lines are easier to model, they are quite bulky even at 60GHz, increasing silicon area and cost. This thesis has shown the feasibility of compact on-chip transformers at 60GHz and their potential to perform impedance transformation at mm-wave frequencies with low insertion loss.

In order to facilitate a transformer-PA co-design, we have proposed a distributed modeling approach in this thesis. Compared to a traditional lumped element model, such a distributed model of a transformer requires far fewer parameters to predict its performance and is also size-scalable. The accuracy of the proposed model has been verified through measurement results.

Using such a model, a two-stage as well as a high-gain three-stage transformer-coupled power amplifier has been demonstrated at 60GHz in 90nm CMOS technology. We have outlined a systematic design methodology which can enable simultaneous optimization of output power, efficiency and gain, while ensuring stability. The designed prototype has one of the highest reported output power for a 60GHz CMOS PA without any power combining. The power amplifier has been integrated into a complete 60GHz transceiver meant for short-

range high-data rate wireless communication.

The high bandwidth available at 60GHz makes this mm-wave band exciting for short-range Gbps wireless communication. For longer range communication, the lower frequency bands like 2.4GHz and 5.8GHz are still the de-facto standards. In order to increase the data rates, newer standards are packing more bits/sec/Hz using higher order modulation. Such modulation schemes place stringent linearity constraints on the power amplifier. In addition, output power requirements are also higher at these frequencies. This thesis has investigated the design of a highly-linear, high output power 2.4GHz power amplifier in 90nm CMOS technology. Transformer-based power combiners have been utilized to generate high output power in silicon technology. We have shown that the choice of an optimal bias point, together with capacitive phase compensation, allows the design of a linear yet efficient power amplifier. The designed prototype has one of the highest powers ($> 30dBm$) reported in literature for a linear PA.

Since the battery life in a portable radio is often dominated by the power amplifier efficiency, it is important to increase it as much as possible. Conventional transmitter architectures do not allow the use of non-linear power amplifiers which are more efficient. There is thus an efficieny-linearity trade-off in conventional implementations. In addition, conventional transmitter architectures are not benefiting much from technology scaling. The area of transistors shrink every process node but the passives do not scale well. Thus it is very attractive to come up with an architecture that requires fewer passives.

In order to break this linearity-efficiency trade-off as well as to come up with a more scalable solution, this thesis has investigated the design of digitally-modulated polar transmitters. The advantages and limitations of such an architecture have been studied in depth. We have proposed the use of an efficient non-linear inverse Class-D PA as an unit cell. Such an inverse class-D PA can achieve efficiency very close to a Class-E PA but requires a much simpler matching network. The amplitude information is transmitted in a digital manner by turning on or off unit cells. We have shown that in such a digital architecture the choice of the amplitude and phase resolution is impacted not by EVM or linearity requirements but mainly by far out-of-band noise considerations for co-existence issues. The need for a low-power digital filter has been justified and a modified polyphase filter has been proposed as a low-power, efficient solution.

A prototype of the polar system has been demonstrated in 65nm CMOS technology. It consists of a fully-integrated PA as well as all the associated digital circuitry. The PA delivers 22dBm output power with 44% peak efficiency, while operating from a 1-V supply. The transmitter has also been tested with 802.11g WLAN signal and meets all EVM and spectral mask requirements at an output power of 14-15dBm with more than 18% efficiency. This performance is even comparable to some commercial PAs that employ exotic compound semiconductor processes.

6.2 Future Directions

Almost every research contributes something to the understanding of the specific subject. Meanwhile, it also opens the door to many other new topics. This work is no exception. The mixed-signal polar transmitter architecture holds a lot of potential. The amplitude path, particularly the design of an efficient power amplifier core, has been investigated in this thesis. The phase modulation has been performed externally. Integrating the phase modulator with the rest of the transmitter would be a natural extension of this work. A low-power design of a phase modulator which can handle large bandwidth is an interesting research problem.

The power amplifier used for the digital PA uses an on-chip transformer as a matching network. It would be worthwhile to investigate the integration of the PA with a transformer-based power combiner. This can enable higher output power generation. More importantly, it will be interesting to study whether individual amplifiers can be turned on and off in a dynamic manner to improve the average efficiency. The fact that it can be turned off on a packet-to-packet basis to achieve power back-off efficiently has been demonstrated in this work by designing a dual-mode WiMAX/WiFi PA. However, it might be possible to use the top two bits of the amplitude codeword to turn unit amplifiers on and off in a dynamic manner (tracking the envelope). The impact of such a scheme on linearity (like DNL etc.) needs to be investigated.

Bibliography

[1] G. Fettweis, "Challenges in architecting and designing next generation cellular systems in silicon." BWRC Retreat Talk.

[2] "Nokia q3 2010 quarterly report."

[3] M. Y. Bohsali, *Millimeter-Wave CMOS Power Amplifiers Design*. PhD thesis, Univeristy of California at Berkeley, 2009.

[4] N. Guo *et al.*, "60-ghz millimeter-wave radio: principle, technology, and new results," *EURASIP Journal on Wireless Communications and Networking*, Jan 2007.

[5] I. Aoki, *Distributed Active Transformer for Integrated Power Amplification*. PhD thesis, California Institute of Technology, 2002.

[6] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press, 2003.

[7] I. Aoki *et al.*, "Distributed active transformer - a new power-combining and impedance-transformation technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, pp. 316–331, Jan 2002.

[8] A. Zolfaghari *et al.*, "Stacked inductors and transformers in cmos technology," *IEEE Journal of Solid State Circuits*, vol. 36, April 2001.

[9] W. Simuburger *et al.*, "A monolithic transformer coupled 5-w silicon power amplifier with 59% pae at 0.9 ghz," *IEEE Journal of Solid State Circuits*, vol. 34, Dec 1999.

[10] J. G. McRory *et al.*, "Transformer coupled stacked fet power amplifiers," *IEEE Journal of Solid State Circuits*, vol. 34, pp. 157–162, Feb 1999.

[11] H. Gan *et al.*, "Integrated transformer baluns for rf low noise and power amplifiers," in *IEEE Radio-Frequency Integrated Circuits*, Jul 2006.

[12] B. Heydari *et al.*, "A 60 ghz power amplifier in 90nm cmos technology," in *IEEE Custom Integrated Circuits Conference*, Sep 2007.

[13] T. Suzuki *et al.*, "60 and 77ghz power amplifiers in standard 90nm cmos," in *IEEE International Solid State Circuits Conference*, pp. 562–563, Feb 2008.

[14] Y. Jin *et al.*, "A millimeter-wave power amplifier with 25db power gain and +8dbm saturated output power," in *IEEE European Solid-State Circuits Conference*, pp. 276–279, Sep 2007.

[15] D. Dawn *et al.*, "17-db-gain cmos power amplifier at 60 ghz," in *International Microwave Symposium*, pp. 859–862, June 2008.

[16] Y. Cao *et al.*, "Frequency-independent equivalent-circuit model for on-chip spiral inductors," *IEEE Journal of Solid State Circuits*, vol. 38, pp. 419–426, Mar 2003.

[17] D. Chowdhury *et al.*, "Modeling and characterization of on-chip transformers for 60ghz power amplifier applications," in *Techcon*, Nov 2008.

[18] A. Shirvani *et al.*, "A cmos rf power amplifier with parallel amplification for efficient power control," *IEEE Journal of Solid State Circuits*, vol. 37, pp. 684–693, Jun 2002.

[19] P. Reynaert *et al.*, "A 2.45-ghz 0.13um cmos pa with parallel amplification," *IEEE Journal of Solid State Circuits*, vol. 42, pp. 551–562, Mar 2007.

[20] G. Liu *et al.*, "A 1.2v, 2.4ghz fully integrated linear cmos power amplifier with efficiency enhancement," in *IEEE Custom Integrated Circuits Conference*, pp. 141–144, Sep 2006.

[21] K. C. Tsai *et al.*, "A 1.9 ghz, 1-w cmos class-e power amplifier for wireless communications," *IEEE Journal of Solid State Circuits*, vol. 34, pp. 962–969, Jul 1999.

[22] C. Yoo *et al.*, "A common-gate switched, 0.9w class-e power amplifier with 41% pae in 0.25um cmos," in *Symposium on VLSI Circuits*, pp. 56–57, 2000.

[23] I. Aoki *et al.*, "Fully integrated cmos power amplifier design using the distributive active-transformer architecture," *IEEE Journal of Solid State Circuits*, vol. 37, pp. 371–383, Mar 2002.

[24] J. Kang *et al.*, "Highly linear 0.18-um cmos power amplifier with deep n-well structure," *IEEE Journal of Solid State Circuits*, vol. 41, pp. 1073–1080, May 2006.

[25] Y. Ding *et al.*, "A high-efficiency cmos +22dbm linear power amplifier," *IEEE Journal of Solid State Circuits*, vol. 40, pp. 1895–1900, Sep 2005.

[26] J. Kang *et al.*, "A single-chip linear cmos power amplifier for 2.4 ghz wlan," in *IEEE International Solid State Circuits Conference*, pp. 761–762, Feb 2006.

[27] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*. Artech House, 1999.

[28] L. Kahn, "Single-sided transmission by envelope elimination and restoration," *Proc. Inst. Radio Eng*, p. 803806, Jul 1952.

[29] D. Su *et al.*, "An ic for liniearizing rf power amplifiers using envelope elimination and restoration," *IEEE Journal of Solid State Circuits*, vol. 33, pp. 2252–2258, Dec 1998.

[30] T. Sowlati *et al.*, "Quad band gsm/gprs/edge polar loop transmitter," in *IEEE International Solid State Circuits Conference*, pp. 186–187, Feb 2004.

[31] A. Kavousian *et al.*, "A digitally modulated polar cmos power amplifier with a 20-mhz channel bandwidth," *IEEE Journal of Solid State Circuits*, vol. 43, pp. 2251–2258, Oct 2008.

[32] R. B. Staszewski *et al.*, "All-digital pll gsm and edge transmitter in 90nm cmos," in *IEEE International Solid State Circuits Conference*, pp. 316–317, Feb 2005.

[33] P. T. M. van Zeijl *et al.*, "A digital envelope modulator for a wlan ofdm polar transmitter in 90nm cmos," *IEEE Journal of Solid State Circuits*, vol. 42, pp. 2204–2211, Oct 2007.

[34] P. T. M. van Zeijl *et al.*, "High-power digital envelope modulator for a polar transmitter in 65nm cmos," in *IEEE Custom Integrated Circuits Conference*, pp. 733–736, Sep 2008.

[35] G. Liu, *Fully Integrated CMOS Power Amplifier*. PhD thesis, Univeristy of California at Berkeley, 2006.

[36] A. M. Niknejad, *Electromagnetics for High-Speed Analog and Digital Communication Circuits*. Cambridge University Press, 2007.

[37] P. Haldi *et al.*, "A 5.8 ghz 1-volt linear power amplifier using a novel on-chip transformer power combiner in standard 90 nm cmos," *IEEE Journal of Solid State Circuits*, vol. 43, pp. 1054–1063, May 2008.

[38] D. Chowdhury *et al.*, "A 60ghz 1-volt +12.3dbm transformer-coupled wideband pa in 90nm cmos," in *IEEE International Solid State Circuits Conference*, pp. 590–591, Feb 2008.

[39] T. O. Dickson *et al.*, "30-100 ghz inductors and transformers for millimeter-wave (bi)cmos integrated circuits," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 123–133, Jan 2005.

[40] J. R. Long, "Monolithic transformers for silicon rf ic design," *IEEE Journal of Solid State Circuits*, vol. 35, pp. 1368–1382, Sep 2000.

[41] S. Mohan *et al.*, "Simple accurate expressions for planar spiral inductances," *IEEE Journal of Solid State Circuits*, vol. 34, pp. 1419–1424, Oct 1999.

[42] D. Chowdhury *et al.*, "Design considerations for 60ghz transformer-coupled cmos power amplifiers," *IEEE Journal of Solid State Circuits*, vol. 44, pp. 2733–2744, Oct 2009.

[43] T. S. D. Cheung *et al.*, "Design and modeling of mm-wave monolithic transformers," in *Digest of Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 1–4, Oct 2006.

[44] D. M. Pozar, *Microwave Engineering.* Wiley, 2004.

[45] S. B. Cohn *et al.*, "Characteristic impedances of broadside-coupled strip transmission lines," *IRE Transactions on Microwave Theory and Techniques*, vol. 8, pp. 633–637, Nov 1960.

[46] D. Chowdhury *et al.*, "A 2.4 ghz mixed-signal power amplifier with low-power integrated filtering in 65 nm cmos technology," in *IEEE Custom Integrated Circuits Conference*, Sep 2010.

[47] O. Degani *et al.*, "A 90nm cmos power amplifier for 802.16e (wimax) applications," in *IEEE Radio-Frequency Integrated Circuits Conference*, Jun 2009.

[48] D. Chowdhury *et al.*, "Transformer-coupled power amplifier stability and power back-off analysis," *IEEE Transactions on Circuits and Systems - Express Briefs*, vol. 55, pp. 507–511, Jun 2008.

[49] P. Haldi *et al.*, "A 5.8 ghz linear power amplifierin a standard 90nm cmos process using a 1v power supply," in *IEEE Radio-Frequency Integrated Circuits Conference*, pp. 431–434, Jun 2007.

[50] T. S. D. Cheung *et al.*, "A 21-26-ghz sige bipolar power amplifier mmic," *IEEE Journal of Solid State Circuits*, vol. 40, pp. 2583–2597, Dec 2005.

[51] A. M. Niknejad and H. Hashemi, *mm-Wave Silicon Technology: 60GHz and Beyond.* Springer, 2008.

[52] C. H. Doan *et al.*, "Design of cmos for 60ghz applications," in *IEEE International Solid State Circuits Conference*, Feb 2004.

[53] T. Yao *et al.*, "Algorithmic design of cmos lnas and pas for 60-ghz radio," *IEEE Journal of Solid State Circuits*, vol. 42, pp. 1044–1057, May 2007.

[54] G. Gonzalez, *Microwave Transistor Amplifiers: Analysis and Design.* Prentice Hall, 1996.

[55] T. Yao *et al.*, "A 24-ghz, +14.5dbm fully integrated power amplifier in 0.18um cmos," *IEEE Journal of Solid State Circuits*, vol. 40, pp. 1901–1908, Sep 2005.

[56] B. A. Floyd *et al.*, "Sige bipolar transceiver circuits operating at 60 ghz," *IEEE Journal of Solid State Circuits*, vol. 40, pp. 156–167, Jan 2005.

[57] U. R. Pfeiffer *et al.*, "A 77 ghz sige power amplifier for potential applications in automotive radar systems," in *IEEE Radio-Frequency Integrated Circuits Conference*, pp. 91–94, Jun 2004.

[58] B. Heydari *et al.*, "Internal unilateralization technique for cmos mm-wave amplifiers," in *IEEE Radio-Frequency Integrated Circuits Conference*, pp. 463–466, Jun 2007.

[59] C. Marcu *et al.*, "A 90nm cmos low-power 60ghz transceiver with integrated baseband circuitry," *IEEE Journal of Solid State Circuits*, vol. 44, pp. 3434–3447, Dec 2009.

[60] D. Chowdhury *et al.*, "A single-chip highly linear 2.4 ghz 30 dbm power amplifier in 90 nm cmos technology," in *IEEE International Solid State Circuits Conference*, pp. 378–379, Feb 2009.

[61] T. Sowlati *et al.*, "A 2.4ghz 0.18um cmos self-biased cascode power amplifier with 23dbm output power," in *IEEE International Solid State Circuits Conference*, pp. 294–295, Feb 2002.

[62] C. Wang *et al.*, "A capacitance-compensation technique for improved linearity in cmos class-ab power amplifiers," *IEEE Journal of Solid State Circuits*, vol. 39, pp. 1927–1937, Nov 2004.

[63] N. B. Carvalho *et al.*, "A comprehensive explanation of distortion sideband asymmetries," *IEEE Trans. Micro. Theory Tech.*, vol. 60, pp. 2090–2101, Sep 2002.

[64] D. Chowdhury *et al.*, "A fully integrated dual-mode highly linear 2.4 ghz cmos power amplifier for 4g wimax applications," *IEEE Journal of Solid State Circuits*, vol. 44, pp. 3993–3403, Dec 2009.

[65] M. S. P. Reynaert, *RF Power Amplifiers for Mobile Communications*. Springer, 2006.

[66] N. O. Sokal *et al.*, "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, Jun 1975.

[67] H. Kobayashi *et al.*, "Current-mode class-d power amplifiers for high-efficiency rf applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, Dec 2001.

[68] T. P. Hung *et al.*, "Design of high-efficiency current-mode class-d amplifiers for wireless handsets," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, Jan 2005.

[69] Y. C. Lim *et al.*, "Signed power-of-two term allocation scheme for the design of digital filters," *IEEE Transactions on Circuits and Systems -II*, vol. 46, May 1999.

[70] http://casper.berkeley.edu/wiki/ROACH.

[71] RFMD, *RF5622: 3.0V TO 3.6V, 2.4GHz TO 2.5GHz LINEAR POWER AMPLIFIER.*

[72] T. Biondi *et al.*, "Analysis and modeling of layout scaling in silicon integrated stacked transformers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, pp. 2203–2210, May 2005.

[73] A. Cote, "Matrix analysis of oscillators and transistor applications," *IRE Transcations on Circuit Theory*, vol. 5, pp. 181–188, Sep 1958.

[74] J. Smith, *Modern Communication Circuits.* McGraw Hill, 1986.

[75] C. D. Presti *et al.*, "A high resolution 24-dbm digitally-controlled cmos pa for multi-standard rf polar transmitters," in *IEEE European Solid State Circuits Conference*, pp. 482–485, Sep 2008.

Appendix A

PA Stability Analysis

In this appendix, a small-signal common-mode stability analysis of the output stage of the PA is presented. A simplified schematic of the final stage of the two-stage PA is shown in Fig. A.1. Transistors M_1 and M_2 constitute the differential pair, L_{pr} and L_s represent the inductance of the primary and secondary coils of the transformer, L_{DD}, L_{GND} are the inductance to the supply voltage and ground connections (including bond wires), C_b is the on-chip bypass capacitor (with series resistance R_1), and R_L is the 50 Ω load resistor. In any linear amplifier, the gate to drain feedback capacitor C_{gd} in conjunction with the drain inductor (L_{pr}) can cause potential instability. This is referred to as internal feedback in this work. In addition, the bond wire inductances (L_{DD}, L_{GND}) and on-chip bypass capacitors (C_b) form another feedback path, which is referred to as the external feedback.

Because of the presence of more than one loop in the circuit and possible interaction, a loop gain approach of determining stability is not suitable. We adopt a matrix based approach to analyze the oscillatory behavior of the amplifier. For ease of analysis of common-mode instability, a suitable common mode half-circuit is used, and the small signal representation is depicted in Fig. A.2. Here, R_g represents the gate resistance, while $L_g = 2L_{GND}$, $L_v{=}2L_{DD}$ and $L_p{=}\frac{1}{2}L_{pr}$. Writing the network equations using KCL and KVL for the circuit shown in Fig. A.2 results in a set of equations as a function of network currents and voltages. These equations can easily be arranged and represented in the following matrix form

$$[A][B] = [S_i] \tag{A.1}$$

where B represents the network currents and voltages shown in Fig. A.2 and S_i is the input matrix and are given by

$$B = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ V_{gs} \\ I_4 \end{bmatrix}, \quad S_i = \begin{bmatrix} V_i \\ 0 \\ V_i \\ V_i \\ 0 \end{bmatrix}.$$
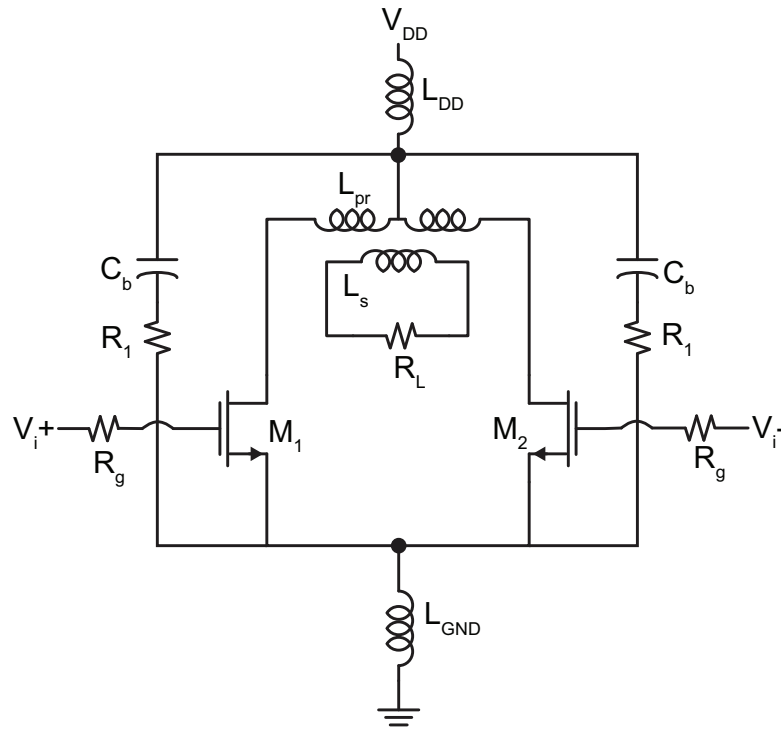
Figure A.1: Schematic of the output stage of the PA with bond-wire inductances and bypass capacitances.
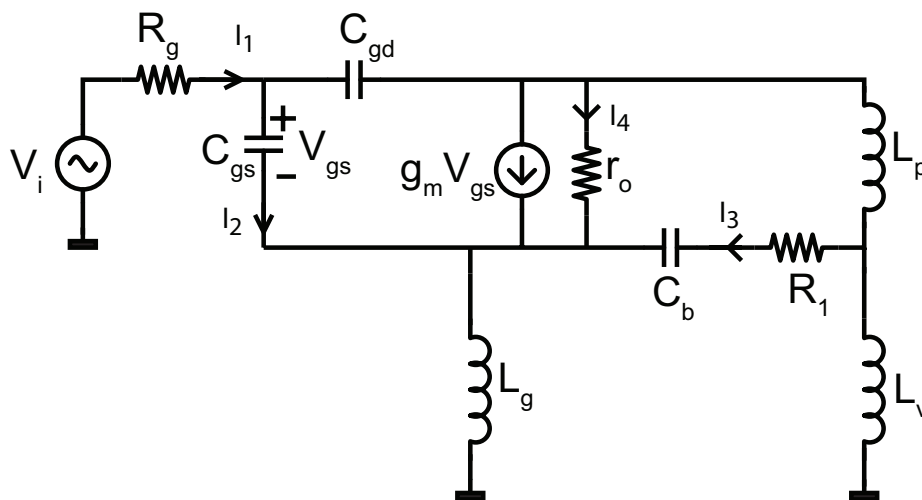

Figure A.2: Small-signal representation of common-mode half-circuit of the PA.

The matrix [A] is as shown below

$$
\begin{bmatrix}
R_g & j\omega L_g & j\omega L_g & (j\omega g_m L_g + 1) & j\omega L_g \\
0 & 1 & 0 & -j\omega C_{gs} & 0 \\
\begin{matrix}(R_g + Z_L + \\ 1/j\omega C_{gd} + j\omega L_v)\end{matrix} & \begin{matrix}(-1/j\omega C_{gd} \\ -Z_L - j\omega L_v)\end{matrix} & -j\omega L_v & (-j\omega g_m L_v - g_m Z_L) & (-j\omega L_v - Z_L) \\
R_g + j\omega L_v & -j\omega L_v & (-j\omega L_v - Z_C) & (-j\omega g_m L_v + 1) & -j\omega L_v \\
-Z_L & Z_L & -Z_C & g_m Z_L & (r_o + Z_L)
\end{bmatrix}
$$

where $Z_L = j\omega L_p + R_p$ (R_p is due to finite Q of on-chip transformer winding) and $Z_C = R_1 + 1/j\omega C_b$.

For the system to oscillate, there must be finite current flow in the circuit, meaning that matrix B must be non-zero, even in the absence of any applied signal [73] [74]. This can be true only if A is a singular matrix, i.e.

$$\det[A] = 0 \tag{A.2}$$

In general, the determinant has a real part as well as an imaginary part and thus we have two equations. The determinant for this circuit may be generally represented as

$$\det[A] = (k_{0r} + k_{2r}\omega^2 + k_{4r}\omega^4 + k_{6r}\omega^6) + j(k_{0i} + k_{2i}\omega^2 + k_{4i}\omega^4 + k_{6i}\omega^6) \tag{A.3}$$

where

$$k_{mr} = f(g_m, R's, C's, L's), \; for \; m = 0, 2, 4, 6$$
$$k_{mi} = f(g_m, R's, C's, L's), \; for \; m = 0, 2, 4, 6$$

Setting the real part equal to zero gives us the oscillation frequency, while setting the imaginary part equal to zero gives us the start-up condition. However, in many cases, as a starting approximation, we can make a high-Q assumption and neglect the losses (memoryless elements) in the system. Under such a condition, Fig. A.2 can be simplified as shown in Fig. A.3. The system matrix then simplifies to

$$
\begin{bmatrix}
0 & \frac{1}{j\omega C_{gs}} + j\omega L_g & j\omega L_g \\
\frac{1}{j\omega C_{gd}} + j\omega L_p + j\omega L_v & -\frac{1}{j\omega C_{gd}} - j\omega L_p - j\omega L_v & -j\omega L_v \\
j\omega L_v & \frac{1}{j\omega C_{gs}} - j\omega L_v & -\frac{1}{j\omega C_b} - j\omega L_v
\end{bmatrix}
\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix}
=
\begin{bmatrix} V_i \\ V_i \\ V_i \end{bmatrix}
\tag{A.4}
$$

Once again, for oscillation to occur, the matrix A must be singular, which in this high-Q case, simplifes to the following equation.

$$
\begin{aligned}
0 = {} & -L_p L_v L_g \omega^6 + \left[\frac{L_g L_p}{C_{gs}} + \frac{L_g L_v}{C_{gs}} + L_g\left(\frac{L_p}{C_b} + \frac{L_v}{C_b} + \frac{L_v}{C_{gd}}\right) + \frac{L_v L_p}{C_{gs}}\right]\omega^4 \\
& -\left[\frac{L_g}{C_{gd}C_{gs}} + \frac{1}{C_{gs}}\left(\frac{L_p}{C_b} + \frac{L_v}{C_b} + \frac{L_v}{C_{gd}}\right) + \frac{L_g}{C_b C_{gd}}\right]\omega^2 + \frac{1}{C_b C_{gs} C_{gd}}
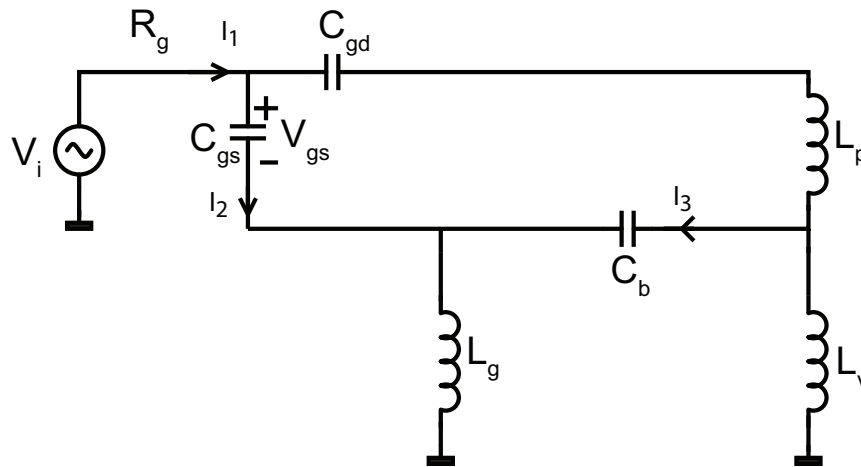\end{aligned}
\tag{A.5}
$$

Figure A.3: Small-signal representation of common-mode half-circuit of the PA under high Q assumption.

A solution of Eq. A.5 gives the potential oscillating frequencies for the designed power amplifier. For analysis of oscillating frequency, we had set the input matrix S_i [in Eq. A.1] to zero. However, if now a voltage V_i is applied at the oscillating frequency, the matrix equation can be re-solved to obtain the port input impedance ($Z_i = V_i/I_1$). For oscillation to occur, the real part of this port impedance must be negative. Having already determined the possible oscillation frequencies, it becomes very easy for us to calculate the input impedance of the amplifier and determine if start-up conditions are satisfied or not.

As an example, if L_p = 40 pH, L_v = 1 nH, L_g = 100 pH, C_b = 2 pF, and device capacitances (C_{gd} and C_{gs}) are estimated from the foundry model of the $80\mu m$ transistor, then the three solutions of Eq. A.2 are ω_1 = 21.13 GHz, ω_2 = 304.05 GHz and ω_3 = 1.01 THz. Of course, the frequencies ω_2 and ω_3 being very high (beyond f_{max}), the losses of the system are high enough to suppress these modes. On the other hand, the input resistance is negative at 21.13GHz and hence, it is a possible oscillation tone, which has been confirmed by simulation as well. Note that this tone is not harmonically or sub-harmonically related to the fundamental frequency of 60GHz. The analysis here thus accurately predicts the existence of common-mode instability in a linear power amplifier. It is interesting to observe that because of the presence of ground bond-wire, the bypass capacitor is not connected to a perfect ground and hence it can indeed cause instability.