CMOS Circuits and Devices beyond 100 GHz

Babak Heydari

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2008-121
http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-121.html

September 18, 2008

CMOS Circuits and Devices Beyond 100 GHz

By

Babak Heydari

B. Science, Karshenasi. (Sharif University of Technology) 2002
M. S. (University of California, Berkeley) 2006

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 Chenming Hu
Professor Paul K. Wright

Spring 2008

Abstract

CMOS Circuits and Devices beyond 100 GHz

by

Babak Heydari

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Ali M. Niknejad, Chair

Mm-wave CMOS circuits are expected to enter the consumer market in the next few years and become a part of most mobile devices offering a drastic increase in the data transfer speed compared to the available systems today. mm-wave car radars is also expected to become ubiquitous and increase the safety of the roads. High performance silicon technology will also find several applications in medicine once it can efficiently operate in the THz frequencies.

This dissertation follows two basic goals. On one hand, the goal is optimizing the performance of CMOS for the mm-wave technology. To reach this goal, mm-wave systems have been considered from device and circuit angles. From the device point of view, a systematic way of making high performance active devices have been proposed and various device parameters such as power gain, noise performance, power efficiency and stability have been analyzed and optimized. An extensive investigation of accurate modeling methods for transistors up and beyond 100 GHz has also been performed. On the circuit side, several high performance, low power 60 GHz low noise amplifiers and a power amplifier were

designed and implemented. To increase the performance of such circuits, some novel techniques such as unilateralization have been proposed and implemented.

The other goal of this research was to investigate the true limits of CMOS technology. More specially, it has tried to answer this question that at any given technology node, what would be the ultimate frequency that one can design active circuits with acceptable performance and a reasonable power dissipation. To this end, some techniques to maximize the f_{max} of a CMOS transistor were proposed and few circuits, an amplifier and an oscillator were designed and implemented using these devices to operate beyond the f_t of the technology.

To My Parents, Shahin and Reza

Acknowledgments

I would like to thank many people who made my life at Berkeley memorable and enjoyable. First and foremost, I wish to express my deep gratitude to my advisor, Professor Ali M. Niknejad. Without his continuous supports and guidance, this work would not have been possible. During my PhD years, I had weekly phone conversations with many of my undergraduate classmates who were also doing their PhDs in different schools in the country. In a lot of occasions, after several minutes of listening to their complaints about their advisors as being detached from the research, not available and problematic, I would always feel really lucky to have such an advisor, with a deep knowledge and a pleasant personality . I also would like to thank my other thesis committee member, Professor. Chenming Hu. I learned a lot from Prof. Hu while I was working in the BSIM group during my first years of graduate studies and continued to benefit from his comments and ideas. I would like to also thank my other thesis committee and qual committee members,Professor. Paul Wright. Professor Tsu Jai King and Professor Micheal O'Hare. I would also like to thank Professor Robert G. Meyer for teaching EE142 and EE242 on high frequency circuits. I found the intuition that I learned in those classes very useful in a lot of occasions during my research. My thanks also goes to Professor Vivek Subramanian for teaching EE231, Solid-State Device physics, in the Spring 2003 that exposed me to an effective intuitive way of thinking about solid state devices. I think I should also thank Prof. Mehrdad Sharif Bakhtiar and Prof. Ali Fotowat, my undergraduate professors at Sharif University of Technology from whom I learned about electronic circuits for the first time. This work would not be possible without the immense help of my other team members, Ehsan Adabi and Mouir Bohsali who

would always kindly share their experiences and offer their support. I would like to thank other members of the mm-wave CMOS research group, specially Bagher Afshar, Mohammad Amin Arbabian, Cristian Marcu from whom I received support and with whom I had some great research discussions. Most importantly, I would like to thank my parents for their support and encouragement, and my brothers, Behnam and Pedram without them this work would not be done. I should also thank my dear uncle, Hossein Lotfi, who were always as supportive as a father to me and together with my parents had a deep influence on setting my academic and career goals and standards.

Contents

CONTENTS

List of Figures

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

List of Tables

1

Introduction

It's hard to imagine a world without silicon-based RF systems anymore! The revolution in the world of communication is mainly the result of advances in such devices and microwave technologies. Today, with a small device in her pocket, one can read news, check Emails, chat with friends and watch her favorite video clips, all of which are available wirelessly to her. In such a world, a natural question for technology developers and investors is "'What will be the next big wave?" Can we imagine a technology that facilitates data communication even more, making it even faster? Can I use my laptop wireless ethernet with multi-Gbps speed in 10 years from now? And with so many devices in hand, how can I make them communicate to each other in an ultra-fast fashion?

Most of technologists think the answer should be looked for in the mm-wave silicon technology. Although other technologies such as 3-10 GHz UWB are also potential candidates, mm-wave looks more promising due to several reasons, most importantly the large available bandwidth at 60 GHz and virtually unlimited allowable transmission power. UWB, the nearest competitor has some shortcomings including problems with interference at least in part of its available band and a

limited data rate. Due to these reasons, the last few years have witnessed an increasing interest in millimeter-wave communication systems and as this dissertation is being written, several main communication IC companies are about to start mm-wave CMOS divisions. Still, as it is the case with any new technology, the question is whether mm-wave CMOS is just a temporary hype or it is a reality of the future. Is Multi-Giga bit per second data transfer really needed, at least with a commercial volume? How large its market can be? and can such systems be developed with reasonable price and power efficiency? To have a valid estimate of the future of the technology, a good practice is looking into the ecosystem surrounded that technology and analyze various trends that are related to it. This is done in the next section.

1.1 The Ecosystem Of Multi-Gbps Wireless Networks

Moore introduced the concept of business ecosystem for the first time as an economic community supported by a foundation of interacting organizations and individuals the organisms of the business world (1). These organisms and communities come together in a partially intentional, highly self-organizing, and even somewhat accidental manner." In a business ecosystem, various elements are inter-related and change based on a so-called *co-evolutionary paradigm* according to which organisms are not simply the result, but also the cause of their environment (2). For a new technology to become successful and sustainable, all the related pieces of the puzzle should come together at the right time and in an

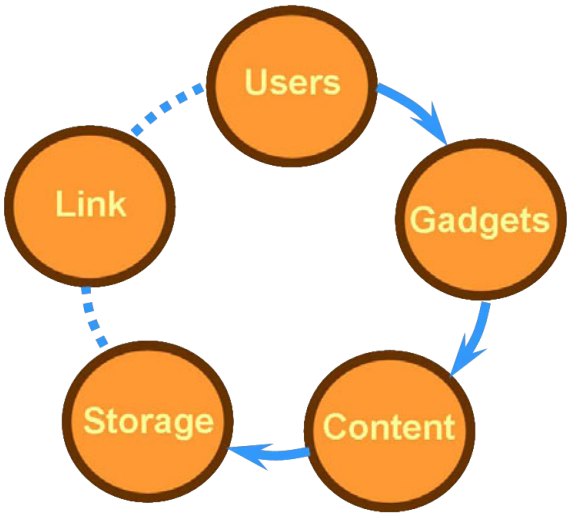


Figure 1.1: The ecosystem diagram of mm-wave wireless networks.

interactive fashion.

For mm-wave communication systems, important elements of the ecosystem are users, wireless devices and gadgets, high quality content, embedded memory and a cheap reliable technology. All the organisms of the ecosystem seem to be in their places in the next few years and in the system,the wireless technology is a missing link for which mm-wave technology can be used. This is summarized in the diagram shown in Fig.1.1

People and their behavior are always the most important portion of an ecosystem. For mm-wave technology, the important points to know are how many people have access to mobile devices, what different activities they do on a daily based with these devices and how these trends are going to change in time. Figure.1.2 shows the increase in the number of mobile users during the last decade and compares it to the number of internet users. As the graph shows, half of the

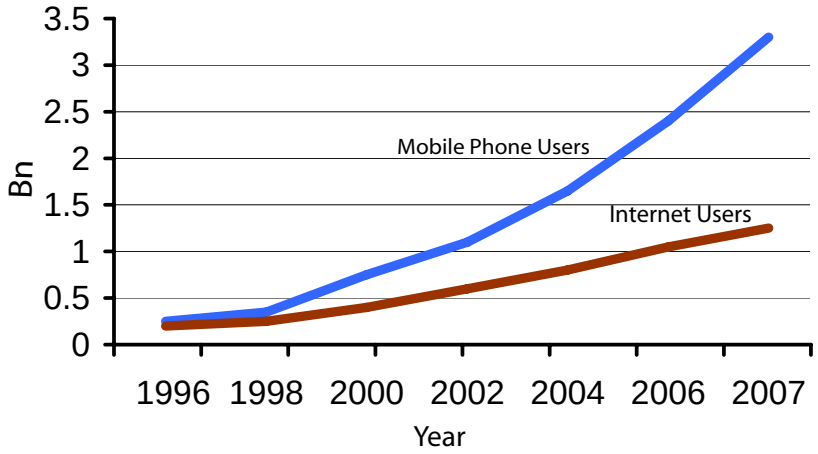


Figure 1.2: Mobile phone and internet usage trends in the last decade (3).

world population are currently using mobile phones. Although most of these people use their phones solely for voice communication, one would expect that the younger generation show a different usage behavior and do many more activities with their devices. Table. 1.1 shows how different age groups have used their mobile phones during the last year in the US. The *mobility* of people are not at the same level in different parts of the world. In fact, East Asian countries such as Japan and South Korea are far ahead of the rest of the world in adapting wireless technologies and using cellular devices for various applications. It's interesting to note that in the year 2007, five out of ten best seller novels in Japan were written on cellular phones (3). With a considerable percentage of young generation using their phones as a camera, video and gaming devices, one could reasonably expect that voice communication will be only one among many activities that will be done by cellular phones in the next few years.

Wireless devices are the next pieces of the puzzle. Although wireless enabled mobile devices such as cellphones, PDAs, cameras and game devices have been

Age	18-29	30-49	50-64	65+
Send or receive text messages	85	65	38	11
Take a picture	82	64	42	22
Play a game	47	29	13	6
Play music	38	16	5	2
Record a video	34	19	8	3
Access the internet	31	22	10	6
Send or receive e-mail	28	21	12	6
Send or receive instant messages	26	18	11	7
Watch a video	19	11	4	2
At least one of these activities	96	85	63	36

Table 1.1: Use of mobile phone or PDA to do various activities, by age group (4), %, 2007

around for several years, the introduction of *iphone* in 2007 should be considered as a revolutionary turning point. *Iphone* basically moved the concept of internet on-the-go from the realm of early adapters and technology fans to the mainstream users and the trend is expected to continue with the price reduction of the phone and introducing of similar devices by other manufacturing companies. It is worth mentioning that Google for example, received 50 times more search queries from iphone in 2007 compared to all other mobile devices combined (3).

Embedded memory is another important part of the game. It is expected that the average embedded flash memory in mobile devices reach 30 GB by 2010. Synchronization of devices with this amount of embedded memory with current wireless technologies take more than half an hour which is not acceptable for a lot of occasions and faster highways should be built between these devices. Moreover, high quality digital data in the form of video and game are also growing exponentially. The total revenue from video download market was 2 billion dollars in 2007, but is expted to exceed 10 billion dollars by 2010. It is also predicted

that the worldwide mobile gaming market will be worth 6.6 billion dollars by 2011 (6).

With so many mobile phone users, multimedia devices, available embedded memory and high quality content, the only missing link is a fast wireless link to transfer the data to and from these devices and synchronize them together. Another imaginable scenario is purchasing a movie from a digital kiosk in a local Cafe or an underground station in a matter of few seconds. Using Digital Right Management (DRM) technologies, one could even rent a movie or a game for a limited time period from these kiosks (7).

Figure 1.3: Various wireless devices and their communication and synchronization with each other.

Another important point to consider is the regulatory environment. In 2001, FCC announced a 7GHz bandwidth, between 57 GHz-64 GHz as an unlicensed band. This regulatory shift was the key motivation for using 60 GHz spectrum ever since. As other parts of the world such as Europe and Japan have also similar regulations for 60 GHz, this band became the center of attention for many ultra-high speed applications. In July 2003, the IEEE 802.15.3 working group for WPAN began investigating the use of this bandwidth as an alternative physical layer to enable very high data rate applications with a targeted data rate greater than 2 Gbps. It is worth mentioning that 60 GHz is one of the oxygen absorption bands that creates a very large path loss (about 10 dB per meter) and limits the range of 60 GHz communication. But this path loss could be beneficial for WPAN systems as it provides an extra spatial isolation and higher implicit security to the systems.

1.2 Why Silicon?Why CMOS

Silicon is not the natural candidate when it comes to mm-wave systems. This realm has been dominated by III-V technologies for many years and even today, silicon performance is still inferior to technologies such as GaAs, PHEMT,InP HEMT and InP HBT. Silicon carrier mobility is lower, limiting the performance of the technology.While the f_t of today's CMOS technology is well under 200 GHz, InP devices with an f_t of approximately 400 GHz and an f_{max} of over 1 THz have been reported.(9). Moreover, the silicon substrate resistance is relatively low, resulting in poorer isolation and higher losses in integrated passive devices.

Despite all these advantages of III-V technology, the interest in silicon for

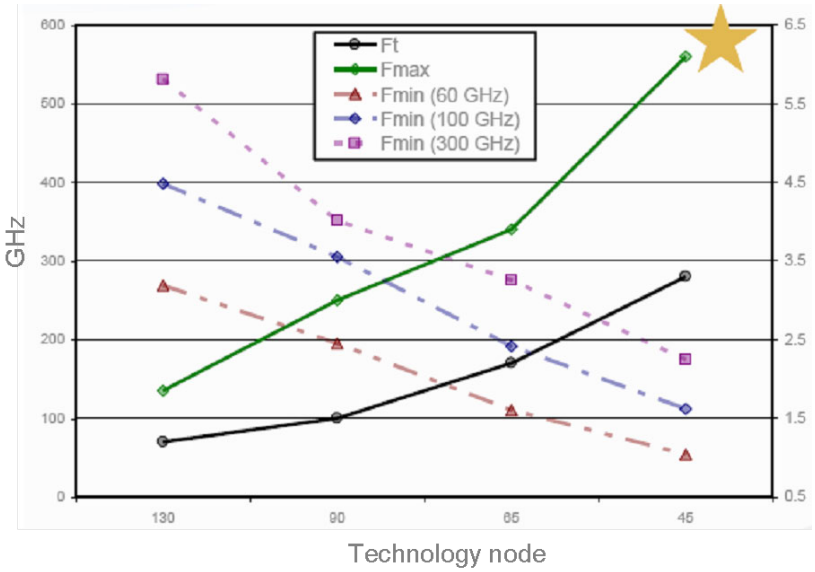


Figure 1.4: CMOS performance vs. technology node

commercial mm-wave systems is increasing. Commercial applications discussed in the previous section are cost sensitive and need to benefit from economy of scale. The worldwide manufacturing capacity of silicon technologies and its superior yield compared to III-V rivals can assure the low cost required for commercial applications such as cellular phone or gaming devices. In recent years, advances in silicon technology, mainly driven by digital applications, have given a chance to silicon devices to perform well in the mm-wave band. In particular, the performance of CMOS technology, quantified by f_t,f_{max} and NF_{min} has significantly improved with geometry scaling and other relevant technology enhancements. As the figure 1.4 suggests, at the $45nm$ technology node, CMOS will preform with a $f_t/f_{max)}$=$280/550GHz$ and an NF_{min} of about $1dB$ making it a better candidate for mm-wave systems than most of today's III-V devices.

Apart from cost benefits, silicon provides a unique integration advantage to

III-V. The silicon capability of higher level of integration at a high yield enables the implementation of entire array-based transceivers topologies on a single die. These topologies are keys to many commercial mm-wave systems such as car radar systems (10; 11; 12).

When it comes to silicon technologies, CMOS has a clear cost advantage over BiCMOS technology. As a rule of thumb, CMOS offers the same level of performance as BiCMOS with half the geometry length. The routing metal density for BiCMOS with equivalent performance as CMOS is however considerably lower. Even if the cost per area for the two technologies are equal, this difference in the metal density will directly increase the area of the digital portion of the system, resulting in a significant boost in the cost (73).

1.3 mm-wave Beyond 60GHz

The superb future performance of silicon makes one to consider the technology for frequencies, even higher than 60GHz. New potential applications beyond 60 GHz include automotive radar, mm-wave imaging and sub-THz chemical detectors with applications in medicine, security as well as scientific research.

1.3.1 Automotive Radar

The FCC has allocated two frequency bands, 22-29 GHz for short range radar and 76-77 GHz frequency band for long range automatic cruise control automotive radar applications. Currently, these radar systems are made using III-V semiconductor technology and are only available on luxury cars. Considering the high

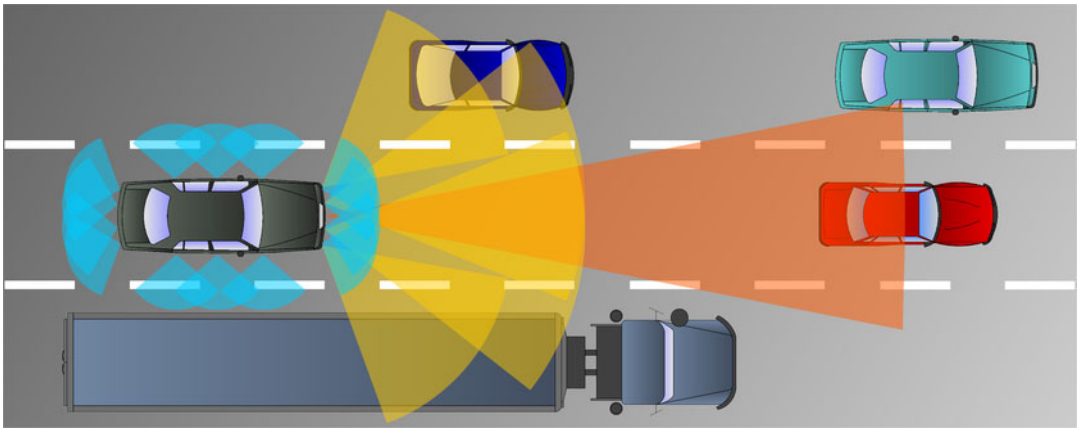


Figure 1.5: A typical automotive radar system.(77)

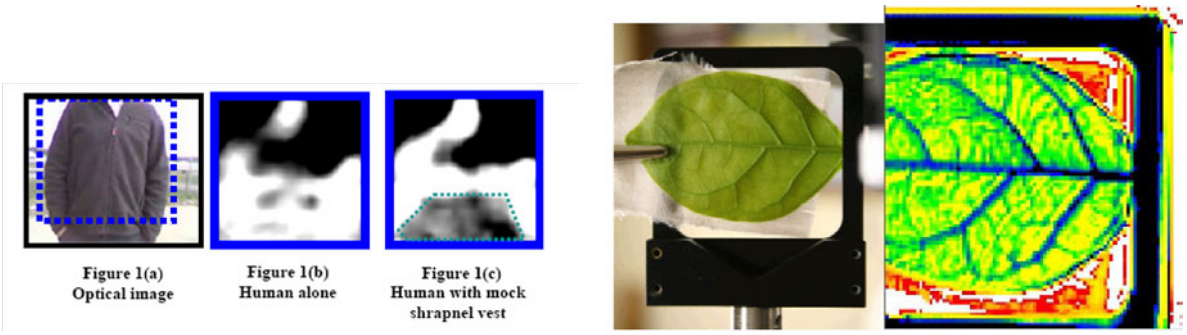
mortality and injury rates as a result of car accidents [1] using silicon technology to make these radar systems ubiquitous looks quite tempting. The technology, radar, is an all-weather system with adequate resolution of few centimeters. In a typical automotive radar, an EM impulse is transmitted toward potential targets. The reflected signal reveals information about the shape, distance and speed of the target, based on its shape and reflection time. A series of these radar sensors are embedded all around the vehicle as shown in figure 1.5 to detect all potential obstacles and hazards.

One potential difficulty with radar systems is the ability to discriminate between various signals, scattered from different objects back to the system. This requires sophisticated signal processing schemes and potentially introduces detection errors in the form of false alarm. The reliability of the automotive radar can dramatically increase by spatially scanning the visible field through antenna steering. The usual mechanical ways of steering the antenna is impractical for low cost systems. An alternative cheap solution is using electronic steering in

[1] In U.S. 42'000 people die annually in car accidents with a 1.5 million injured

the form of phase array systems. Phase array transmitters can form a narrow beam steered toward intended object in a narrow field of view. The receiver is then only sensitive to the reflected signals from that specific angel, substantially reducing the undesired interferences. Silicon technology, CMOS in particular, enables the integration of an entire phase array radar, resulting in a cheap reliable radar system that could play a major role in the road safety in the future.

1.3.2 mm-wave Imaging



(a) (b)

Figure 1.6: (a) mm-wave imaging for concealed weapon detection. (b) Photo of a leaf in an optical domain and a 615 GHz system. (13)

Ultra-fast silicon circuits have also promises in medicine by providing safer and cheaper imaging technology. As silicon technology is capable of integrating large arrays of transceivers, the mm-wave imaging technology could be a low cost competitor to existing technologies such as MRI and CAT scan. A millimeter wave imager can be passive or active. The passive mm-wave imaging system is based on detecting the difference between natural thermal radiation intensity of different materials (14). Currently these systems are being employed for concealed weapon detectors as shown in figure.1.6(a), but same idea could be applied to medical

imagers where a resolution on the order of millimeter is desired. Figure.1.6(b) shows the idea by comparing an optical photo of a leaf to an unprocessed THz image at 615 GHz. At mm-wave frequencies(broadly defined as 30-300GHz), a human body is almost a black body; i.e., it emits 90% of electromagnetic energy across the spectrum. By equating the brightness of the emission source with temperature TA human body naturally emits mm-wave energy and reflects little whereas metal or plastic weapons reflect much mm wave energy and emit less than a human body. While a passive imager detects the contrast of thermal energy between a human body and a weapon, an active imager transmits mm-wave energy to the object of interest and detects reflected energy. By illuminating the target in this way, signal to noise ratio (SNR) is improved, compared to a passive system.

1.4 Research Goals And The Organization Of The Dissertation

The research based on which this dissertation is written follows two basic goals. On one hand, the goal was optimizing the performance of CMOS for the mm-wave technology. To reach this goal, mm-wave systems have been considered from device and circuit angles. From the device point of view, a systematic way of making high performance active devices have been proposed and various device parameters such as power gain, noise performance, power efficiency and stability have been analyzed and/or optimized. An extensive investigation of accurate modeling methods for transistors beyond 100 GHz have also been performed. On

the circuit side, several high performance, low power 60 GHz low noise amplifiers and a power amplifier were designed and implemented. To increase the performance of such circuits, some novel techniques such as unilateralization have been proposed and implemented.

The other goal of this research was to investigate the true limits of CMOS technology. More specifically, it was tried to answer this question: At any given technology node, what is the ultimate frequency that one can design active circuits with acceptable performance and a reasonable power dissipation ? To this end, some techniques to maximize the f_{max} of a CMOS transistor were proposed and few circuits, an amplifier and an oscillator were designed and implemented using these devices to operate beyond the f_t of the technology.

The organization of this dissertation is as follows: Chapter 2 talks about modeling procedures of CMOS devices up to 100 GHz. Chapter 3 discusses the optimization of physical structure of the device in order to maximize its performance. Chapter 4 is about thermal noise in mm-wave devices and circuits, its modeling and optimization methods. Chapter 5 shows the design and implementation of several 60 GHz amplifiers including an LNA and a power amplifier. This is followed by a discussion about unilateralization in mm-wave devices to increase the gain in chapter 6. The designs of two 104 GHz circuits are then demonstrated in chapter 7. Finally chapter 8 concludes the dissertation and proposes some suggestions for future related research.

2

mm-wave Device Modeling

Circuit designers are mostly used to assuming device models as given, instantiate their desired devices in their schematic windows, set up the simulation and run! They might perform their simulations in a number of different process corners and this is as much as they should worry about the whole notion of device modeling. mm-wave circuit design, at least for now, is an exception and both active and passive devices need extensive modeling. In this section, first reasons for this importance are discussed, then the device modeling procedure up to 100GHz is presented and modeling results for single-transistor devices are shown. This follows by a discussion about measurement and de-embedding at these frequencies. Finally modeling of cascode devices are is included as an example of a multi-transistor structure.

2.1 The Importance of Modeling in mm-wave

Available models that circuit designers use in their daily simulations are the so-called "compact" models. Compact models are the interface between the technology and the design. A circuit designer learns about a process by experimenting with the compact model, rather than running expensive and time-consuming experiments

Several good compact models have been developed for digital, analog, and RF applications(23; 24; 25; 26). These models use a combination of physical and empirical methods to develop general equations, usually a large number of them, to describe the behavior of the device. Several parameters are embedded in each equation in order to capture the details of a given technology. These parameters are necessarily determined through complicated curve fitting procedures (parameter extraction) and shape the familiar model card for circuit designers. Most compact models have the advantage of describing the behavior of the device in all regions of operation at the same time. Furthermore, they provide small and large signal analysis as well as noise analysis. They also operate over a fair range of geometry, width and length of the device, over which the extracted parameters are valid. This generality however comes with an accuracy penalty if the model is used over a bias or geometry range outside of the extraction process. Moreover, the core equations in most compact models have been derived under quasi-static assumptions. This, together with the fact that most of available extracted parameters are also for low frequency applications, make these compact models less desirable and inaccurate for millimeter wave applications. Figure 2.1 shows the foundry modeled S parameters of a common-source device and compares it with

the actual measurement of the device.

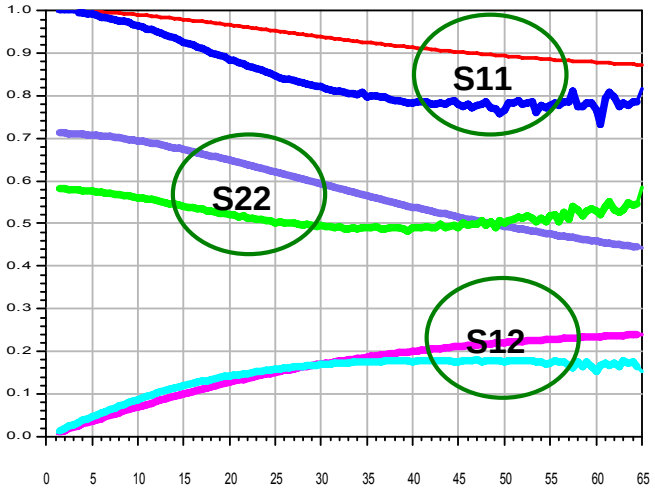


Figure 2.1: S parameters of an $80\mu m$ common-source device, measurement versus model using available BSIM3v3 foundry model

There are two main reasons for this inaccuracy: First of all, as mentioned before, the fact that the parameter extraction has been done in lower frequencies makes the extrapolation to mm-wave frequencies problematic(27). Some of device mechanisms that are not well captured at low frequencies, and naturally not modeled properly, have considerable effect on the performance of the device in higher frequencies, resulting in some inaccuracy. The substrate network including capacitances and resistances is an example of such an effect(28). The inaccuracy due to this effect could be addressed by increasing the frequency range of parameter extraction process.

The second reason for the error in modeling – which is more difficult to address – is due to the layout effect.(16; 22) The device interconnections to the outside world introduce small inductors, resistors and capacitors to the model. These

small components are generally negligible at lower frequencies making the device model more or less independent of layout. These components however change and in fact dominate the performance of the device as the frequency increases and therefore should be included in the model. An accurate prediction of these parasitic requires a detailed full-wave electromagnetic simulation, which is difficult and lengthy. Therefore existing compact models are used as the core for a hybrid customized mm-wave model. In essence, each small finger of the transistor is modeled with the "intrinsic" transistor model and interconnects are captured by a combination of selective electromagnetic simulation and experimental techniques. Due to the importance of device modeling in this project, two round of test structures were fabricated and modeled. The micrograph of one of these chips is shown in figure 2.2. The test chip contains common-source, common-gate and cascode transistors with various sizes as well as different transmission lines and capacitors. The characterized devices were used in all circuits designed and fabricated in $90nm$ process in BWRC.

Given the difficulties in modeling the device, one may be tempted to work directly with measured data. In traditional microwave design the common approach is to use measured S-parameter data for a specific device and treat the transistor as a black box(72; 80). This approach is very accurate in nature and accounts for all parasitics and distributed effects associated with the device and the layout. While this method is sufficient for small-signal circuit design applications, the accuracy of the S-parameter data hinges on reliable measurements of the device and de-embedding structures. As a result, the accuracy of the method may deteriorate for very high frequencies, both due to limited accuracy of test equipment and due to de-embedding errors. Besides, since S-parameters

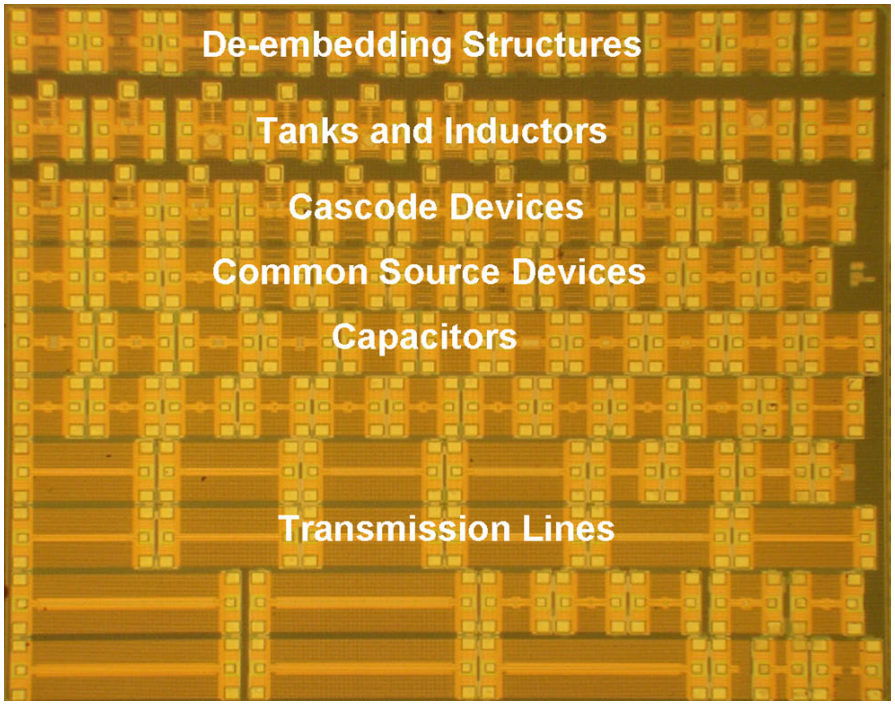


Figure 2.2: A sample $90nm$ test chip fabricated for modeling and characterization.

are small-signal in nature, this method is not suitable for simulation of any non-linear circuit such as mixers or oscillators or the assessment of the dynamic range of amplifiers. Moreover, because the transistor is treated as a black-box, there is no physical insight for improving the device performance or layout. Due to these issues, for mm-wave application, a combination of "RF" and traditional microwave methodology is preferred even for small-signal applications.

2.2 High Frequency Modeling Procedure

A typical transistor layout designed for high frequency is shown in figure 2.3a. The device usually is long and narrow as it is designed with a large number of short fingers to minimize the gate resistance. A connection at the gate and the

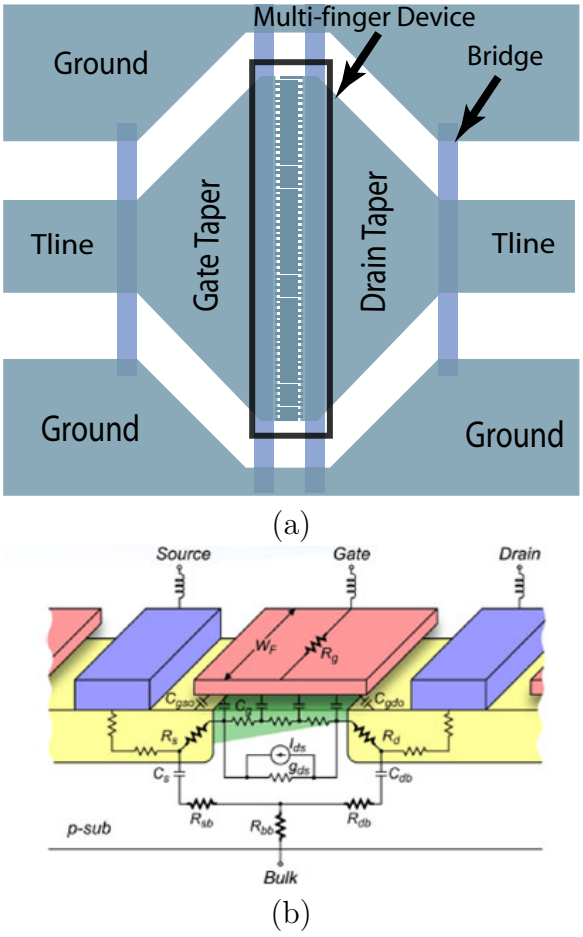


Figure 2.3: (a) Layout of a typical high frequency MOSFET. (b) A cross-section of a MOS transistor showing various parasitics.(Courtesy of Chinh Doan.)

drain, usually in the form of a transmission line connects the device to the outside world. These transmission lines are connected through a 45 degree taper to the transistor for a proper current distribution to and from the device. The cross section of a device is also shown in figure 2.3b to show several parasitics that should be considered in the high frequency modeling of the transistor.

At mm-wave frequencies, series resistive and inductive parasitics become more significant. While the resistive parasitics are always a part of the device, the in-

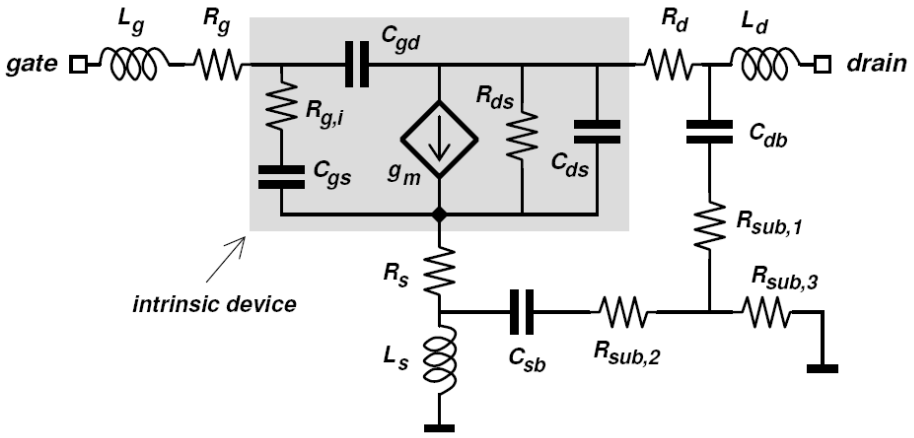


Figure 2.4: Small Signal high frequency equivalent circuit of a MOS transistor.

ductive portion is ususally more significant in high-frequency transistors because of the special layout considerations as mentioned earlier. Consequently, it is critical to properly model these parasitics, in addition to the capacitive effects that are traditionally captured by digital CMOS models. Moreover, neglecting or oversimplifying the substrate network of the device can introduce a considerable error at these frequencies. Figure 2.5 shows the error in the S_{11} and S_{22} of the device caused by ignoring the gate resistance and the substrate network in the small signal model of a typical NMOS transistor.

Equivalent circuits have been an effective approach to analyze the electrical behavior of a device by representing the important components(33; 34). As shown in Figure 2.4 a mosfet device can be divided into two portions: intrinsic part

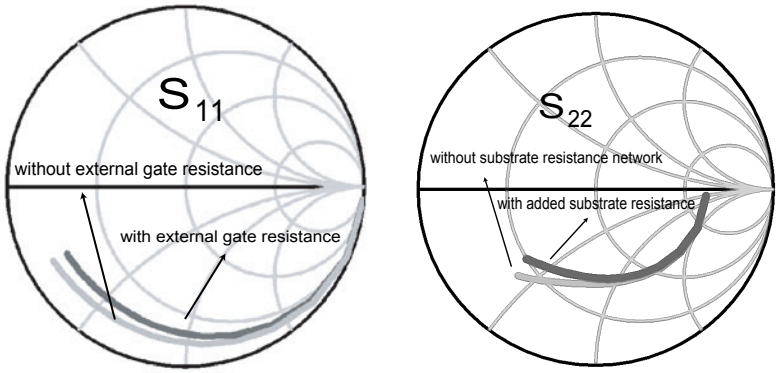


Figure 2.5: Model of a common-source NMOS (a)S_{11} with and without the gate resistance. (b) S_{22} with and without the substrate resistance network.

and extrinsic part. The intrinsic part (the shaded area in figure 2.4) is the familiar hybrid-π model of the device, used for low frequency circuit analysis. The extrinsic part consists of parasitic resistances and inductances at the gate, drain and source as well as a proper substrate network. It is shown that a three resistor substrate network is sufficient to model the device behavior in the mm-wave frequencies(30). Note that extrinsic parasitic capacitances between various terminals could be embedded in the internal device capacitances and be modeled as a part of the intrinsic part.

For each model, the extrinsic component values and device parameters were extracted from measured data using a hybrid optimization algorithm in Agilent IC-CAP(35). Values of the components that could minimize the measurement to model error are not unique and one could come up with several equivalent circuits of the device for the same set of measurement data. This is acceptable as long as the model is used within the measured range of frequency. However, if the component values in the model are made close to their physical values,there is an

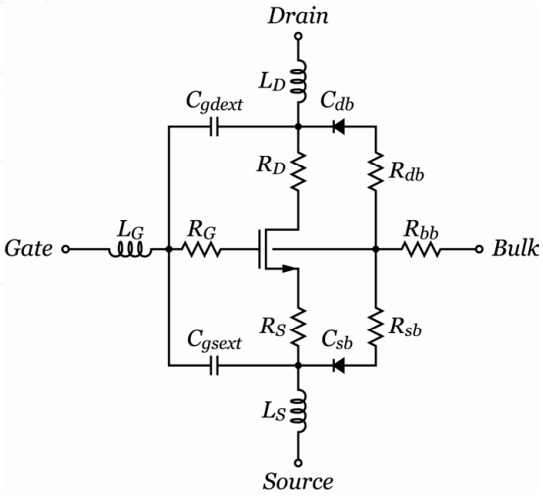


Figure 2.6: Large Signal high frequency equivalent circuit of a MOS transistor.(Courtesy of Sohrab Emami.)

additional benefit and they can be used in frequencies well beyond the maximum measured frequency. Moreover, having a physical equivalent model can help with an accurate assessment of the value of parasitics and the sensitivity of the device performance to them. This information is very useful in optimization of the device physical structure as will be discussed in chapter 3. Because of these reasons, the initial values of the components are calculated using proper equations and based on the measured Y parameters of the device up to 20 GHz(33). The initial value of external resistances and inductances could also be estimated by simulating the connection leads and contacts on the terminals using EM simulators. These initial values then are fed to the optimizer with reasonable tunning ranges to get an accurate physical model.

2.2.1 Large Signal Modeling

Although small signal models are usually sufficient for the design of linear circuits, the design of high performance non-linear blocks such as mixers, oscillators and power amplifiers depends on capturing the nonlinear characteristics of the active devices over a wide range of voltage and current.

Developing a large-signal equivalent model from the scratch is a very complicated process and many physical effects that affect the DC behavior of the device need to be considered. Fortunately, available compact models, such as BSIM3v3 or BSIM4 are specifically created to capture most of these effects. By adding proper parasitics to these foundry given compact models as shown in Fig.2.6, both DC nonlinearities and high frequency effects could be captured simultaneously. Since external terminal resistances and the substrate network are added manually, the BSIM model should be adjusted to turn-off the internal options for these parasitics. Moreover, due to the inherent accuracy compromise in these models to enable them to cover all geometries, the DC behavior of each individual device could be made more accurate and should be also fitted to the measurement by adjusting proper BSIM parameters.(36).

2.3 Measurement and De-embedding

In the high frequency measurement of active and passive devices, the effect of probing pads and extra leads are typically subtracted from the measurement through a de-embedding method (38). In direct de-embedding, the measured results from test structures (such as open and short circuits) are used directly and subtracted from the measurements. In a model based approach, a suitable

physical equivalent circuit topology is selected and rough values for these equivalent circuit parameters then are estimated using a combination of equation-based calculations based on low frequency data. The final fine tuning and fitting is done using an optimizer such as Agilent IC-CAP (35). In this section we review the major de-embedding procedures and discuss the problems and advantages of the various techniques.

2.3.0.1 RF Measurement Pads [1]

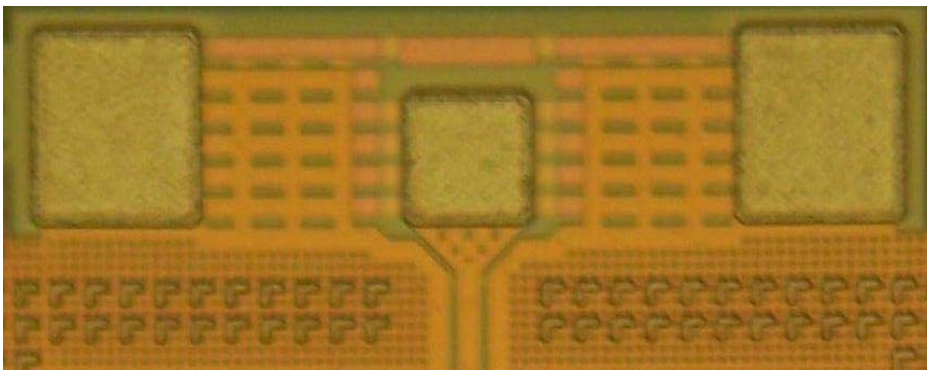


Figure 2.7: Layout of a common RF GSG pad.

Since most connections to the external world go through measurement pads, a good model for the pad is critical. In the model based de-embedding approach, this model also serves as a foundation for de-embedding the effects of the pads whereas in the design of building blocks, the effects of the pads must be included in order to predict the real world performance of the device or circuit.

[1] The pad and transmission line models are contributed by Mounir Bohsali

A common RF pad arrangement is the ground-signal-ground (GSG) structure shown in Fig. 2.7. These pads mate naturally with CPW probes and have good performance in the mm-wave band. Often the signal pad is shielded from the substrate, forming a grounded CPW (G-CPW) structure at the pad. If the transmission line leads to the rest of the circuit are microstrip or G-CPW, then this is the best option to use. Otherwise, if CPW is used, the decision to ground the pad is not clear cut. A shielded pad will form a high-Q structure, since the fields are isolated from the lossy substrate, but the shield adds extra capacitance and a discontinuity in the fields from the probe to the device. In order to reduce the pad capacitance, the signal pad is reduced to the minimum allowable probing dimensions, or about $90\mu m \times 90\mu m$, for $150\mu m$ pitch pads. For smaller pitch pads, smaller pads can be used. The RF pad is considerably smaller than the ground pads. A short 45° taper is used at the output of the pad in order to reduce the reflections due to discontinuities. In the example shown, a $40\mu m$, 50Ω transmission line connects the pad to the rest of the circuits.

2.3.1 Open-Short De-Embedding

A popular de-embedding approach is the so-called open de-embedding, which simply removes the effects of the pads from the measurement structure shown in Fig. 2.8-a by subtracting the measured Y parameters of the pad from the measured device, as shown in Fig. 2.8-b. The key assumption is that the pads are connected in parallel to the DUT, which neglects the physical nature of the pads and treats the signal entry/exit points as lumped circuit nodes. The equivalent circuit for parasitics that could be captured in the open measurement is shown

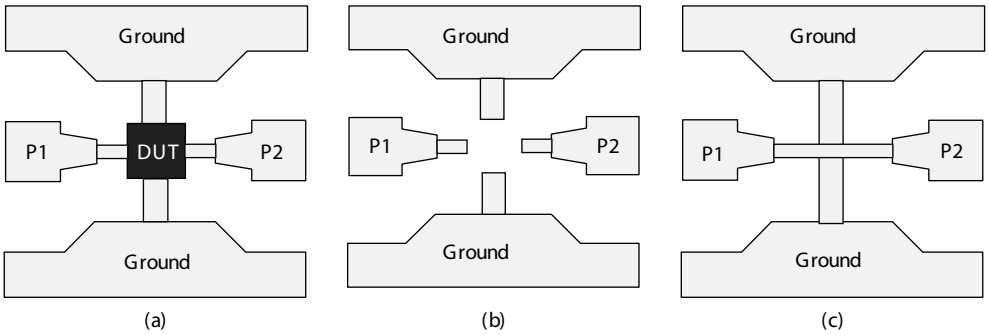


Figure 2.8: (a) The device under test and the measurement pads. (b) The open test structure.(c) The short test structure.

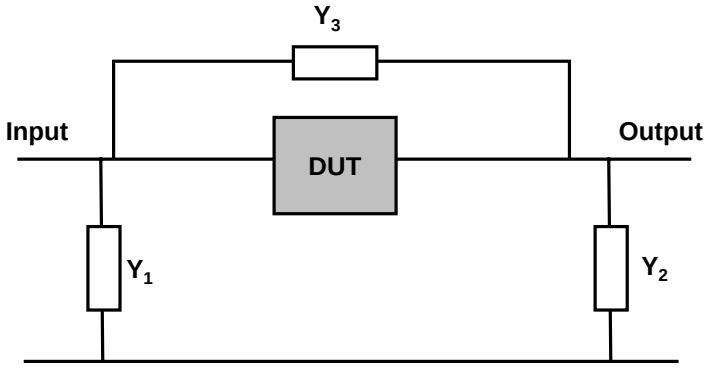


Figure 2.9: The equivalent model of parasicits for the open de-embedding.

in Fig. 2.9. For the DUT we can write:

$$Y_{dut} = Y_m - Y_o \tag{2.1}$$

And we can write these equations for the parasitics:

$$Y_3 = -Y_{12,o} = -Y_{21,o} \tag{2.2}$$

$$Y_1 = Y_{11,o} + Y_{12,o} \tag{2.3}$$

$$Y_2 = Y_{22,o} + Y_{21,o} \tag{2.4}$$

The other assumption for in the open de-embedding is that we can indeed measure a true pad *open* structure by simply open circuiting the pad test structure. In reality, the open circuits have finite fringe capacitance and radiation, which invalidates the above assumptions. In practice this procedure is quite accurate up to 10 GHz for small on-chip structures. In summary, open de-embedding removes the *shunt* parasitics from the measured device.

As the frequency increases, open de-embedding is not sufficient to de-embed all the parasitics and a more common approach is the so called open-short de-embedding. In this approach, in addition to measuring the embedded structure and open pads, a short structure as shown in figure 2.8c is also measured. A typical DUT with parasitics can be represented by an equivalent circuit shown in Figure 2.10.

If we device the matrix Z'_s as

$$Z'_s = \begin{pmatrix} Z_1 + Z_3 & Z_3 \\ Z_3 & Z_2 + Z3 \end{pmatrix} \tag{2.5}$$

Then we can calculate the Z'_s matrix from this equation:

$$Z'_s = (Y_s - Y_o)^{-1} \tag{2.6}$$

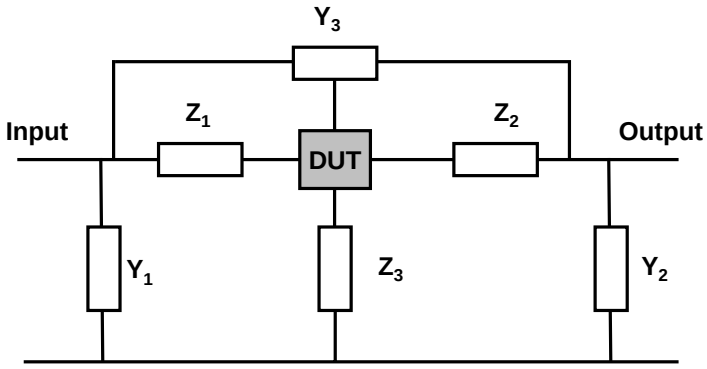


Figure 2.10: The equivalent model of parasicits for the open-short de-embedding.

the same correction is applied to the measured data of interest

$$Y'_m = Y_m - Y_o \tag{2.7}$$

and then the modified short measurement is subtracted from the measurements

$$Z''_m = Y'^{-1}_m - Z'_s = (Y_m - Y_o)^{-1} - \left(Z_s^{-1} - Y_o\right)^{-1} \tag{2.8}$$

In practice this technique is reliable up to about 40 GHz or more, depending on the size of the test structures. By neglecting the distributed nature of the pads, we are limited to frequencies where all dimensions are negligibly small compared to the wavelength.

2.3.2 Recursive Modeling Process

Evidently, the de-embedding step is a major source of inaccuracy at mm-wave frequencies. It introduces error in the data due to imperfect assumption about

the de-embedding structures. For example, for open-short de-embedding, the error arises from imperfect open and short especially at higher frequencies and the distributed nature of the structure. These inaccuracies make the de-embedded result noisy, directly affecting the accuracy of the extracted model. In order to resolve this problem several high frequency de-embedding methods have been proposed (39; 40**?**). Here as an alternative a model based de-embedding approach, dubbed the *recursive modeling* has been employed.

A typical test structure comprises of probing pads, lead transmission lines and the device under test (DUT). In this method, probing pads are modeled in the first step. Pad models are then used to model the transmission line leads and finally the two models are employed to model the complete DUT. Typically all different test structures use identical probing pads and lead transmission lines making it sufficient to model them only once for all the structures. This modeling technique in principal is applicable for any structure, passive or active, as long as an equivalent circuit can capture the behavior of the structures.

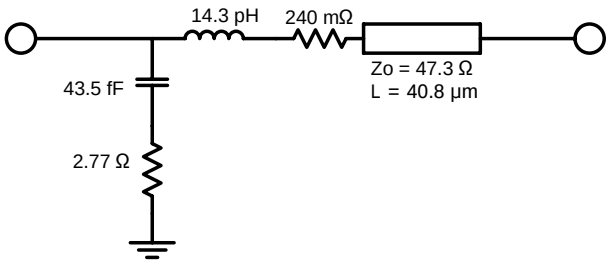


Figure 2.11: Equivalent circuit of pad includes a section of transmission line.

The circuit shown in Fig. 2.11 is used to model the pad over a broad frequency

range from 40 MHz to 65 GHz. The parallel branch represents the equivalent circuit for the pad itself and the series branch models the extra lead. The 1-port S parameters measurements are performed for the pad in two configurations, the first with the output port connected to ground, and the second with the output port left open. To increase the accuracy of the modeling, the Z and Y parameters of the model are simultaneously matched with the measured parameters in both configurations. The results in Fig. 2.12 show that the model accurately captures the RF pad behavior over the frequency range of interest.

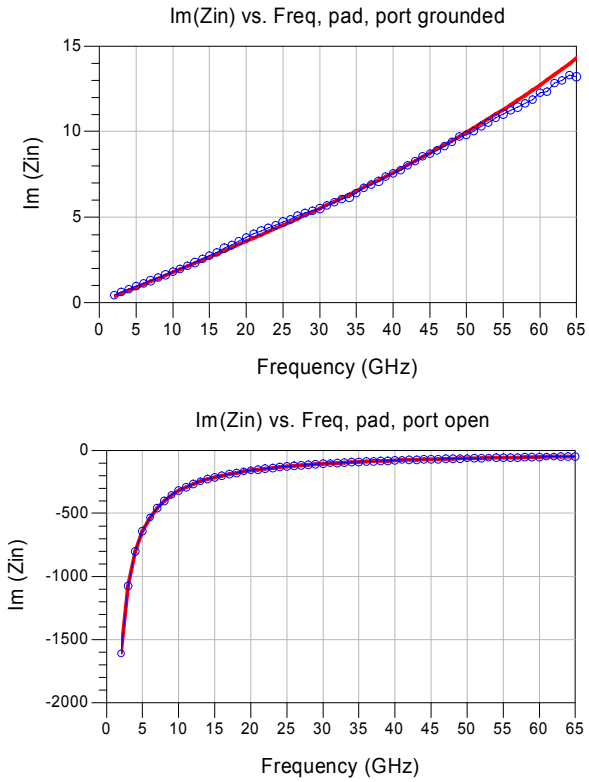


Figure 2.12: Measured vs. simulated $\Im(Z)$ parameters for the RF pad with the port (a) grounded and (b) open.

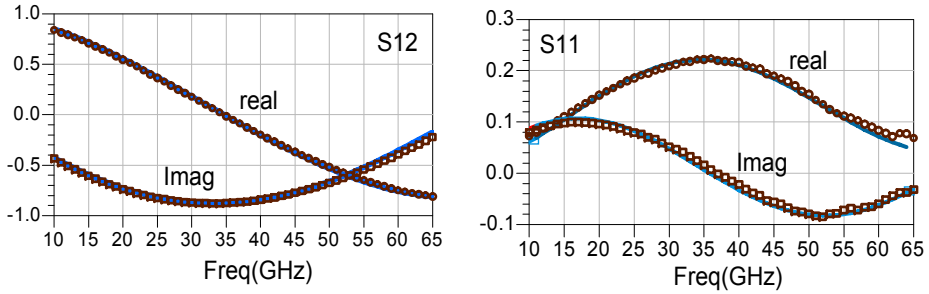


Figure 2.13: Measured versus model of a $500\mu m$ long CPW transmission line with gap spacing $S = 4\mu m$.

The transmission line that is used as a lead from the probe to the device has to be modeled in the next step. Coplanar waveguide (CPW) transmission lines are used in all test structures. A length scalable electrical transmission-line model has been developed to capture the complex propagation constant and frequency dependent characteristic impedance. Fig. 2.13 shows the modeling result for a $500\mu m$ CPW transmission line with a $4\mu m$ signal-to-ground gap.

In the next step, the DUT is modeled. The complete model of the measured structure is made by connecting the previously modeled pad and transmission line whose models should be kept unchanged during this step and the equivalent model of the DUT that could be both small-signal or large-signal as was discussed in the previous section. The initial guesses of the equivalent circuit components are calculated and the whole model is then fitted to the raw device measurement through optimization of transistor core and external parameters.

An experimental verification of this approach has been performed. All the measurements have been done using on-chip probing up to 110 GHz. Agilent IC-CAP software and the hybrid optimization method has been employed to

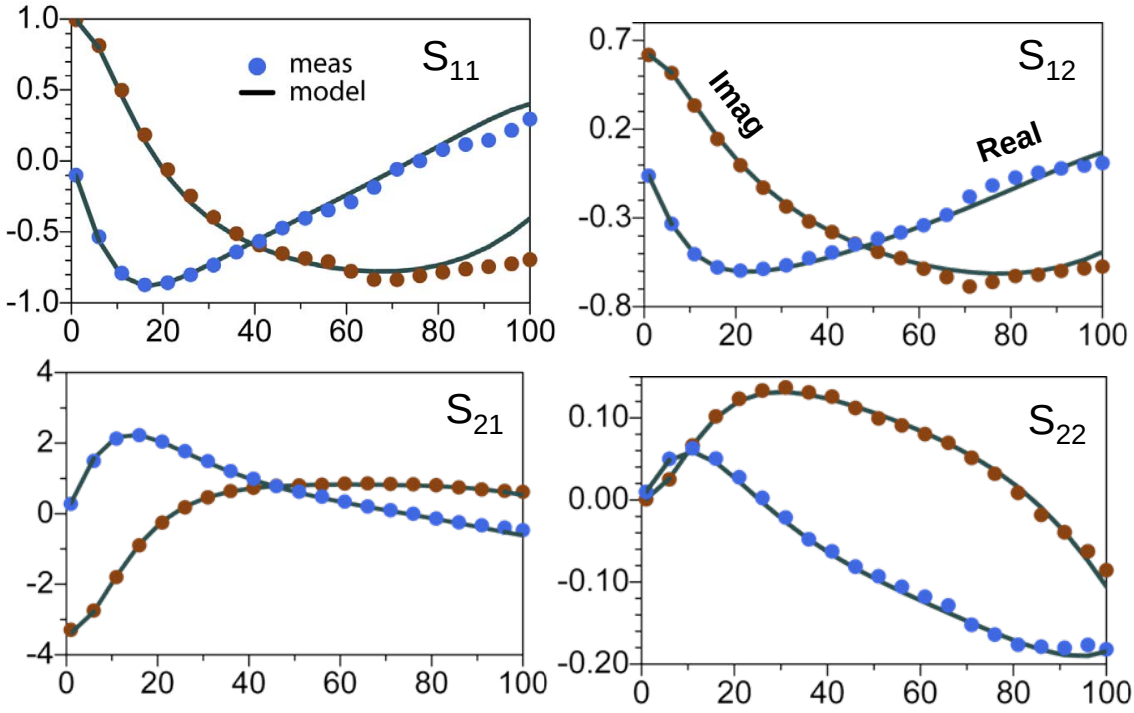


Figure 2.14: Measured (marker) versus simulated (lines) S-parameters of a $40\mu m/90nm$ transistor modeled using the recursive approach.

perform model optimizations. Fig. 2.14 shows the modeling result, the measured and modeled real and imaginary parts of S-parameters for an $80\mu m/0.09\mu m$ transistor up to 100 GHz. The cleanness of the measured data is an advantage of this method which helps the accuracy and speed of the modeling process. The good agreement between the model and measurement suggests that the extended lumped hybrid-pi model is valid to frequencies as high as 100 GHz.

Fig. 2.15 compares the result of the proposed modeling technique to the open-short de-embedding method. The differnece can be best noticed by comparing Y-parameters. The de-embedded data is clearly noisy especially for frequencies in the millimeter-wave bands. For frequencies in the K and Ka bands, the two models give similar results. The discrepancy however gets significant for frequen-

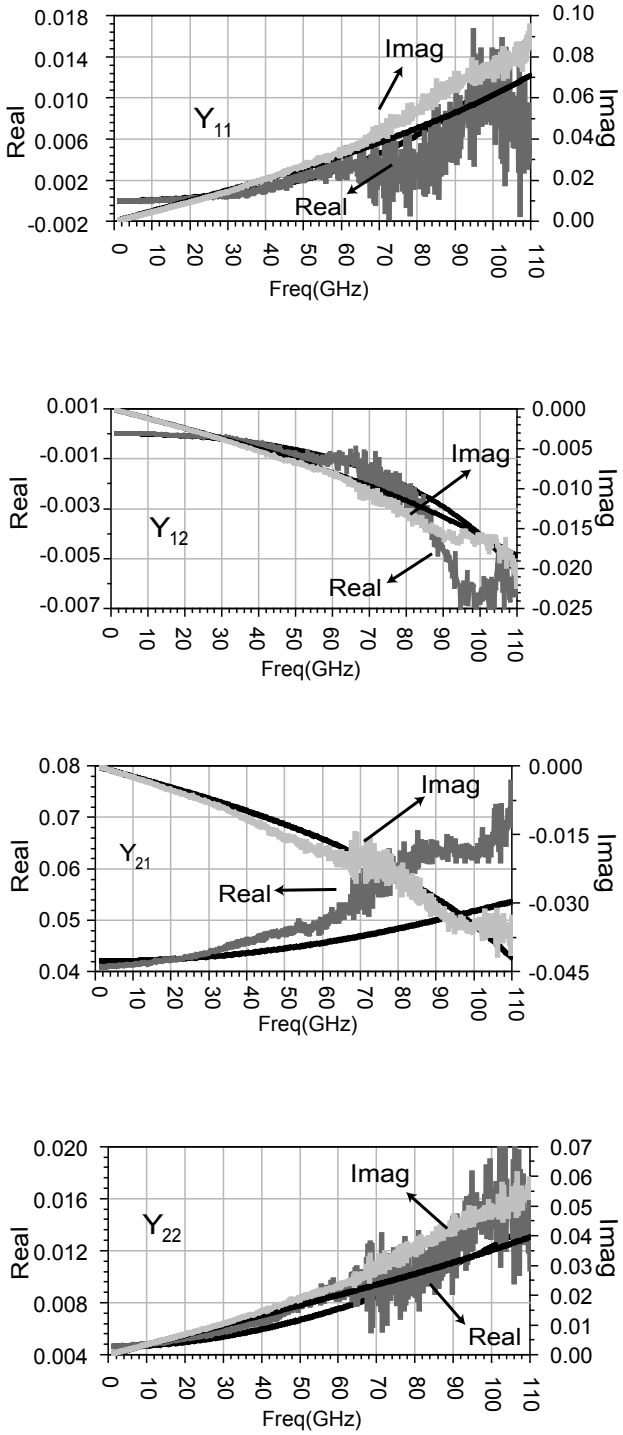


Figure 2.15: Comparison of of the open/short versus recursive de-embedding/modeling approach for the Y parameters of a $40\mu m/90nm$.

cies higher than 40 GHz showing the inaccuracy of open-short de-embedding for millimeter-wave applications. The error becomes specially significant in the imaginary parts of Y_{11} and Y_{22} and the real part of Y_{21}. These would result in major circuit performance degradations as we approach 100 GHz. The method was also tested with measured data up to 60GHz and the predicted data at 100GHz were compared to the actual measurment at this frequency and found to be in a good agreement. This indicates another important advantage of this modeling method that is its ability to extend beyond the measurement frequency without introducing significant error.

2.4 Cascode Modeling

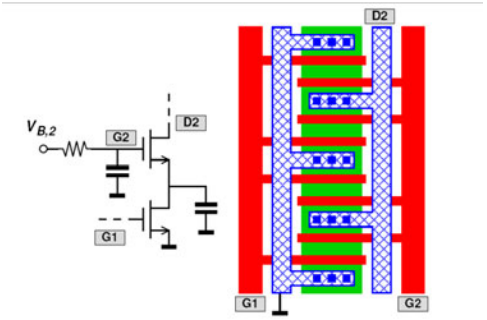
Cascode devices are used extensively in mm-wave design. These devices could potentially provide higher gain compared to common-source devices and are usually unconditionally stable at these frequencies due to the isolation between input and output(41). To minimize the capacitance at the junction of the input and cascode device, a shared junction structure as shown in figure 2.16a is usually used[1].Because of using this structure, cascode devices need special treatment in modeling and a simple connection of the two single transistor models does not accurately predict the device high frequency behavior and specifically can introduce substantial error in the Y_{22} of the device(42).

An equivalent circuit of a cascode transistor is shown in figure 2.16b.The model is essentially similar to the common-source model that was discussed earlier. One important differnece is the way the substrate network is modeled. Because of the

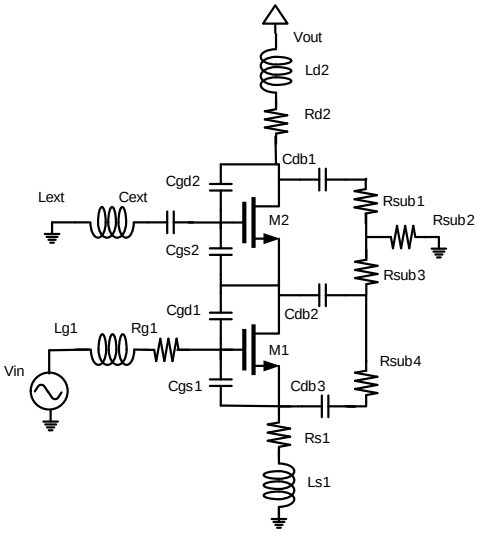
[1]This is explained in more detail in chapter 5.

shared junction structure, the substrates of the two devices are shared and this needs to be considered in the equivalent circuit. This substrate can be a source of feedback between the output and input (43). The second gate of the device is also connected to a bypass capacitor to ensure a high frequency ground to avoid oscillation. Because of the sensitivity of the cascode gate to parasitics , the proper equivalent model of the capacitor should be included on the cascode gate.

To test the accuracy of the model, a sample 40μm/90nm is measured and



(a)



(b)

Figure 2.16: Equivalent circuit of a cascode device. The transistors can be replaced with a hybrid-π model for small signal modeling.

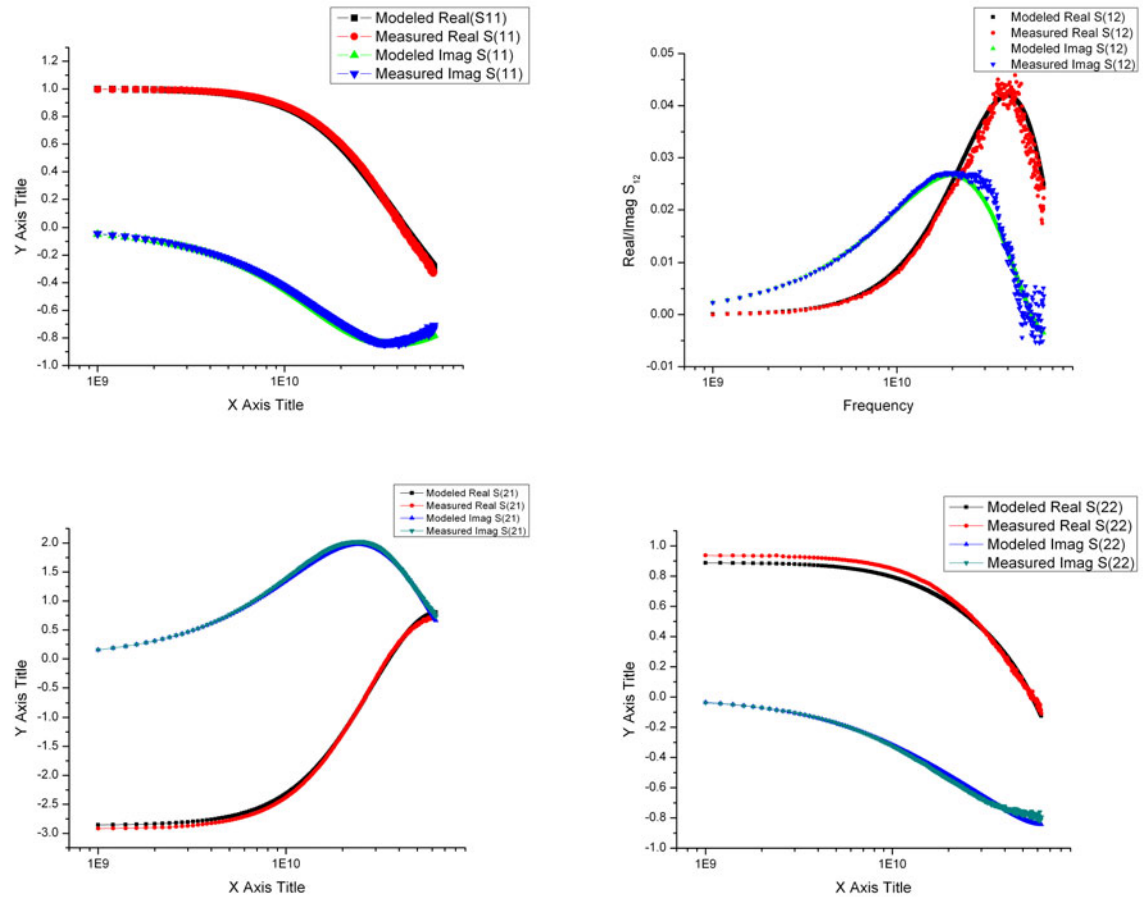


Figure 2.17: Comparison of measured and modeled device S-parameters.

compared to the model using the proposed method. The close match between measured and modeled S-parameters up to 65 GHz as shown in Fig. 2.17 confirms the validity of the model as well as the modeling procedure.

2.5 Summary

Modeling of active and passive devices is an essential step in designing high performance mm-wave circuits. A CMOS transistor can successfully be modeled

using a proper lumped equivalent circuit even beyond 100 GHz and available compact models can be used as the core of large-signal models once proper bias independent equivalent circuit of parasitics and substrate is added to it. On the measurement side, *recursive modeling* proves to be capable of accurately de-embedding the effect of measurement pads and extra connection wirings from the device. Cascode devices need special modeling due to their shared junction nature and their sensitivity to the parasitics on the second gate.

3

mm-wave Device Optimization

As was mentioned in the previous chapter, device performance in mm-Wave frequencies is deeply under the influence of layout parasitics. Apart from the urge for layout dependent models, as was pointed out before, this has another important consequence: Unlike low frequency circuit design in which the device design is absolutely in the realm of process engineers, here the circuit designer could-and should- alter the device performance by changing the device layout(22; 46). This enables the designer to layout the device based on the performance metric which is more important in any specific application. It might be astounding in the first look how much the device layout could vary certain device parameters. f_{max}, for instance, which is an indicator of the speed of the transistor, have been reported for a similar process, CMOS 90nm, from 80GHz to up to 300GHz mainly due to differences in layout(49; 50).

Millimeter-wave Device design is essentially customizing the device layout in order to maximize certain performance metrics. Performance metrics for mm-wave devices are several, f_t, f_{max}, Maximum Stable Gain at a given frequency,

Maximum unilateral gain at a given frequency and Minimum noise figure are the most important ones. That which metric is to be considered as an optimization target depends on the specific application of the device. In this chapter, we first look more closely at some of the most important device pefromance parameters. Then few examples of device optimization including a novel "round-table" device are presented. In the end, there is a breif discussion about the optimization process for large devices used as power delievering transistors.

3.1 Device Performance Metrics

There are several performance parameters for a transistor, and each defines the performance in a different way. Unity current gain frequency (f_t), unity power gain frequency (f_{max}), maximum stable gain (MSG), and maximum unilateral gain (U) are the most popular metrics as shown on figure 3.1. Noise performance and linearity of the device are also crucial when it comes to specific circuit blocks such as low noise or power amplifiers. In order to maximize the performance of a device through layout optimization, it is crucial to consider the correct figure of merit and the criteria for optimization.

The unity current gain frequency , f_t, is the most popular high freqeucy number of a process and is basically the frequency in which the current gain of the device becomes one. It can be calculated from the h_{21} of the device and is equal to (37):

$$f_t = f_{h_{21}=1} \approx 2\pi \frac{G_m}{C_{gs} + C_{gd}} \tag{3.1}$$

In this equation G_m is the effective transconductance of the device and C_{gs} and

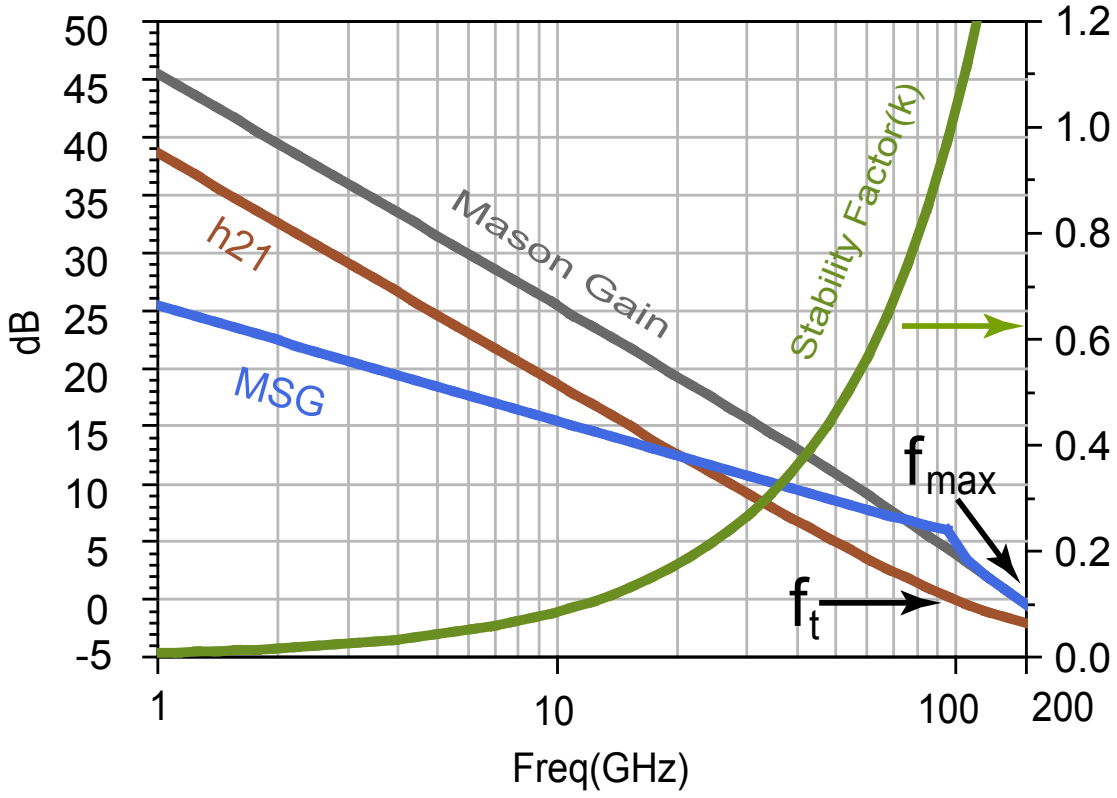


Figure 3.1: Maximum Stable Gain, Maximum Unilateral Gain and h_{21} of a sample Common-Source device. The f_t and f_{max} of the device is defined by the unity values of h_{21} and U curve respectively.

C_{gd} are total gate to drain and gate to source capacitances of the device, including parasitic capacitances. When the layout is reasonable in a way that does not add considerable extra capacitances, f_t is mostly determined by the intrinsic device characteristics and is improved by scaling and/or process optimization. Because resistive losses in the input and output do not affect this parameter, the device layout has negligible effect on f_t.

Although f_t is the most common performance metric for a given technology

node, it does not reflect the performance level of a specific device at that technology node. What actually matters more is the power gain of a device rather than its current gain (51). A device can remain active (have power gain larger than one) in frequencies well above f_t. For this reason f_{max}- the maximum frequency in which the device has power gain- is a more valid metric to show the limit of the activity of the transistor. The unity power gain frequency, f_{max}, strongly depends on parasitic losses of the device and can be improved (or degraded) by optimizing the layout of the transistor. Depending on the layout, the f_{max} could vary from below f_t to values considerably higher than f_t. The ratio, f_{max}/f_t is a figure of merit that shows the optimality of the layout(22). Depending on the assumptions, several equations have been proposed to evaluate the f_{max} of a MOS transistor. Assuming the reasonable assumption that R_s is smaller than the total gate resistance the f_{max} can be written as (45):

$$f_{max} \approx \frac{f_t}{2\sqrt{R_g^t \left(g_{ds} + 2\pi f_t c_{gd}\right)}} \tag{3.2}$$

In which R_g^t is the total gate resistance. This equation shows the effect of R_g and C_{gd} on the value of f_{max} and indicates these are the main factors that need to be minimized to increase the f_{max}. We will see the detail of this optimization in the next section.

While f_t and f_{max} indicate unity gain frequencies and indicate limits of the performance of the device, MSG and Mason's Gain (U) represent the performance of the device at the frequency of interest.

U, is the gain of the device under the condition that the device is unilateral-

ized through some feedback mechanisms or some circuit techniques (47)[1]. More importantly, it is also a figure of merit independent of the topology in which the device is employed and is important in applications where device or circuit unilateralization techniques are used. The value of U can be calculated from this equation (47):

$$U = \frac{|S_{21}/S_{12} - 1|^2}{2\left(k|S_{21}/S_{12}| - \Re S_{21}/S_{12}\right)} \tag{3.3}$$

Where k is the stability factor and can be calculated from this equation[2].:

$$k = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + \Delta^2}{2|S_{12}S_{21}|} \tag{3.4}$$

In an open loop structure, the achievable gain could be considerably lower than the Mason Gain specially in frequencies below $f_t/2$; thus we need a different metric for open loop applications where reaching U is not an option. Maximum Stable gain is a good candidate to serve this goal. To better grasp a sense out of MSG, the case for the stability of the device in question should be discussed first. Under frequencies of high gain condition, devices are generally conditionally unstable. This means that under certain source/load impedance conditions, there is a chance that the circuit starts oscillating. Considerable amount of literature is available to come up with parameters that represent an instability potential of a device. The most common measure is k parameter that is directly calculated through S parameters and ensures and unconditional stability when is above 1.As the figure 3.1 suggests, the MSG curve consists of two separate regions

[1]A general condition for N-port unilateralization is presented in chapter 6
[2]Stability conditions and stability factor are discussed in chapter 5.

distinguished by a kink in between. The kink happens at k=1 and at a frequency after which the device becomes unconditionally stable. The equation for MSG is also piece-wised based on k:

$$\Delta = S_{11}S_{22} - S_{12}S_{21} \tag{3.5}$$

$$MSG = \frac{|S_{21}|}{|S_{12}|} \qquad if \quad k < 1 \tag{3.6}$$

$$MAG = (k - \sqrt{k^2 - 1})\frac{|S_{21}|}{|S_{12}|} \qquad if \quad k > 1 \tag{3.7}$$

It's worth noting that at frequencies before the kink, the unconditional stability is ensured by adding as much loss to the input/output ports to drive k exactly equal to 1 and MSG is calculated based on this assumption. As a result, the internal gate and drain resistances of the device do not affect the MSG as long as they are less than the required add-on resistance to make the device stable. This is an important fact that helps designing especially high performance power devices as will be discussed in the next section. In the $k < 1$ region, the maximum stable gain of a MOS transistor is proportional to g_m/c_{gd}, the ratio of feedforward and feedback factors of the device. As a result c_{gd} is very crucial factor in MSG optimization. The source resistance can also change the value of MSG through changing the effective g_m of the device. After the frequency in which device becomes unconditionally stable, the parasitic losses at the input and output terminals increase the value of K and make the MSG drop faster.

Although these metrics are in correlation with each other, changing one does not always guarantee a change in the other and ultimately the application and the topology in which the device is used determine the goal parameter. Generally

Mason's gain and it's unity cross over, i.e. f_{max}, are more sensitive to the layout than the maximum stable gain, especially where the device is conditionally stable, due to the loss compensation in MSG calculation{[1]

3.2 Layout Effect on Device Performance

To determine the effect of layout on device parameters, a physical small signal model is used in order to ascertain the effect of each parasitic on the desired performance metrics of the device. The small-signal model of the device, discussed in the previous section, is not necessarily unique and different combination of lumped element values could satisfy the required matching between the measurement and the simulation result. As a result, in order to make the model physical and extendable to higher frequencies, the values of these parasitics were partly determined through 3D EM simulation (HFSS) and were set as the initial value for optimization.

The developed physical small-signal model helps determine the effect of each parasitic element on the performance of the transistor. Ultimately this insight can be used to determine the optimal transistor layout. For example, the layout of a common source device has been optimized for f_{max}. The starting point for this procedure was a $80 \times 1\mu m/90nm$ sized device with an MSG of $7.5dB$ at 60 GHz and the extrapolated f_{max} of 143 GHz. A sensitivity analysis was performed for the developed model and the variation of maximum unilateral gain (Mason's Gain) and maximum stable gain together with maximum frequency of operation were determined.

[1]Noise analysis and optimization is discussed in chapter 4.

As expected, the gate to drain capacitance and the gate resistance have the largest impact on f_{max}, and thus layout methods should give the first priority to their minimizing. As we noted earlier, MSG does not change with a reduction in gate and drain series resistances when the transistor is conditionally stable ($k \leq 1$). The source resistance, however, changes the MSG since it changes the effective transconductance through its local feedback effect. A more detail analysis of parasitic resistances and their minimization methods are discussed in the next section.

3.2.1 Parasitic Resistance Optimizations

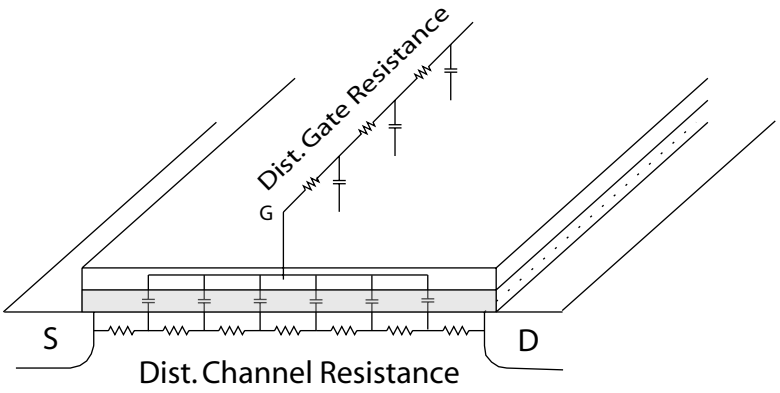


Figure 3.2: Distributed gate and channel resistance.

Gate resistance is the most important resistive parasitic that needs to be minimized. A large gate resistance significantly reduces the f_{max} and available gain. It also affects the noise performance of the device as will be shown in chapter 4. The gate network can be viewed as a distributed RC transmission line as is shown in figure 3.2. The total gate resistance can be divided into three

parts(32):

$$R_{gate} = R_g^{poly} + R_g^{NQS} + R_g^{wire} \tag{3.8}$$

The NQS resistance is an intrinsic device parameter and is a function of the geometry and bias conditions. To optimize the gate resistance one should focus on the other two portions. The R_g^{wire} can be minimized by increasing the number of connection vias from the gate to the top metal layer. R_g^{poly} can be written as (38):

$$R_g^{poly} = \frac{R_{sh}}{N_f L_f}\left(W_{ext} + \frac{W_f}{\alpha}\right) \tag{3.9}$$

In this equation, R_{sh} is the gate sheet resistance, W_f is the channel width per finger, L_f is the channel length, N_f is the number of fingers, and W_{ext} is the extension of the polysilicon gate over the active region that is imposed by the design rules. The factor α is related to the distributed nature of the gate resistance and is equal to 1/3 or 1/12 for single and double gate contacts respectively (51). Equation 3.9 suggests that using more short fingers can reduce the poly gate resistance. The effect of finger width reduction becomes minor as soon as the poly gate resistance becomes a small portion of the total gate resistance in equation 3.8 and is not beneficial any more. Moreover, as will be seen later in this chapter, short finger width can negatively affect the performance of large power devices.

The source and drain resistances have several components including the via resistance, the salicide resistance and the resistance of the LDD region as is shown in figure 3.3. However, the contact and the LDD sheet resistances usually dominate the total resistance. Accordingly, their values can be written as:

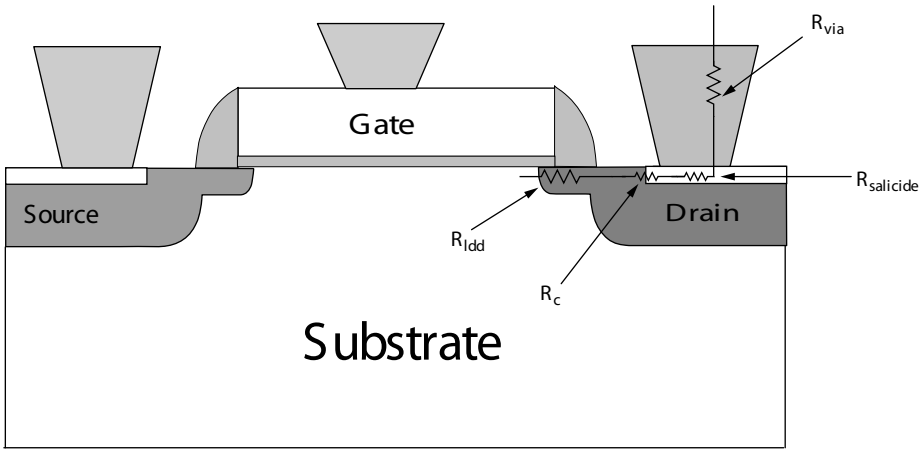


Figure 3.3: Various portions of a MOSFET drain resistance

$$R_d = R_{d0} + \frac{r_{dw}}{N_f W_f} \tag{3.10}$$

$$R_d = R_{s0} + \frac{r_{sw}}{N_f W_f} \tag{3.11}$$

where r_{dw} and r_{sw} are the parasitic drain and source resistances with unit width and R_{d0} and R_{s0} represent the width independent part. Increasing the number of fingers and connection vias help to reduce R_d and R_s too although the number of fingers has to be decided based on gate resistance considerations and other issues related to large power devices as will be described later in this chapter. The value of R_s needs to be minimized as it affects the effective g_m and the noise performance of the device. On the other hand, circuit performance metrics, f_{max} and NF_{min} are not very sensitive to the value of R_d and the circuit layout could be optimized in favor of other parameters if there is any trade-off.

3.2.2 Multi-Finger Layout Optimization

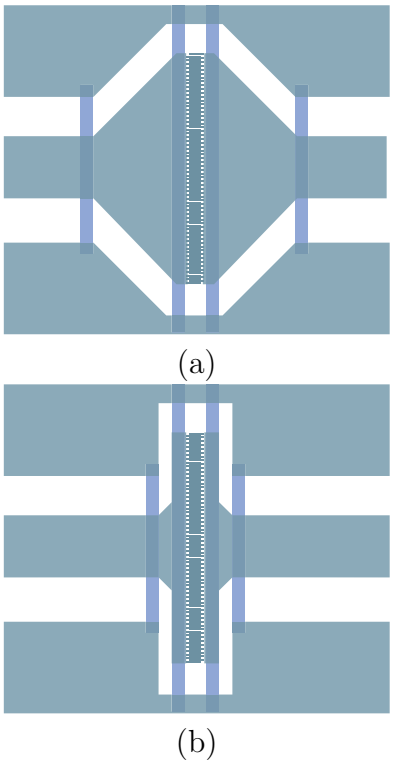


Figure 3.4: Initial (a) and improved (b) layout for an $80\mu m/90nm$ Device. The improved layout includes more substrate contacts, higher density of gate and drain vias and smaller taper

The NMOS structure was modified based on these findings. Mainly the shape of gate and drain tapers, number of gate vias, and width of connections and gate/drain overlap regions were changed. Fig.3.4 shows a layout comparison of the structure before and after modification. The measured performance of the initial device and that of the modified device is shown in Fig.3.5. The f_{max} for the improved structure is up to 178GHz. The maximum stable gain of the device is intact however since the device is in the conditionally stable region as was expected.

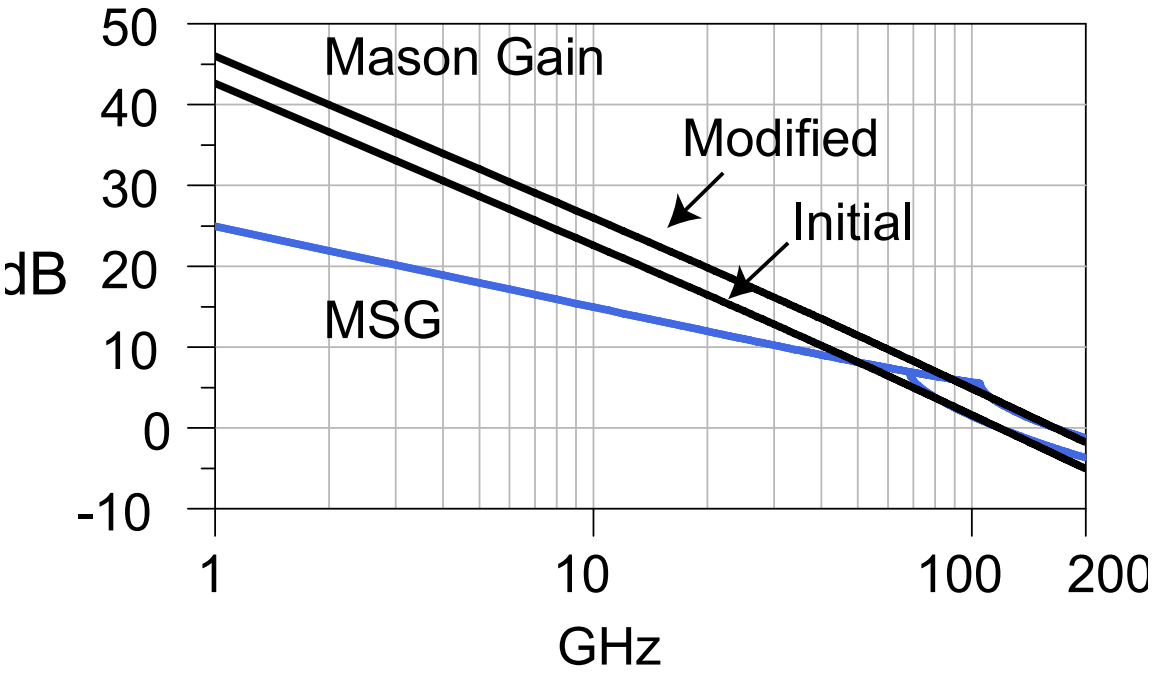


Figure 3.5: The effect of layout improvement on Mason Unilateral Gain and MSG.

3.3 Round-Table Structure

As was shown in the last section the optimal multi-finger layout of an NMOS device could increase the f_{max} up to 20%, but increasing the performance further required further innovation. This is particularly true of the available gain in the conditionally stable frequencies. In order to improve the performance of the device even further, a new structure for the device is proposed. The idea is to reduce the parasitic losses by using a modular approach in device design and using multi-path connections between various modules.

The building block is a standard $10\mu m$ cell with double-gate contacts in order to decrease the finger resistance of the device. Since each finger of the device

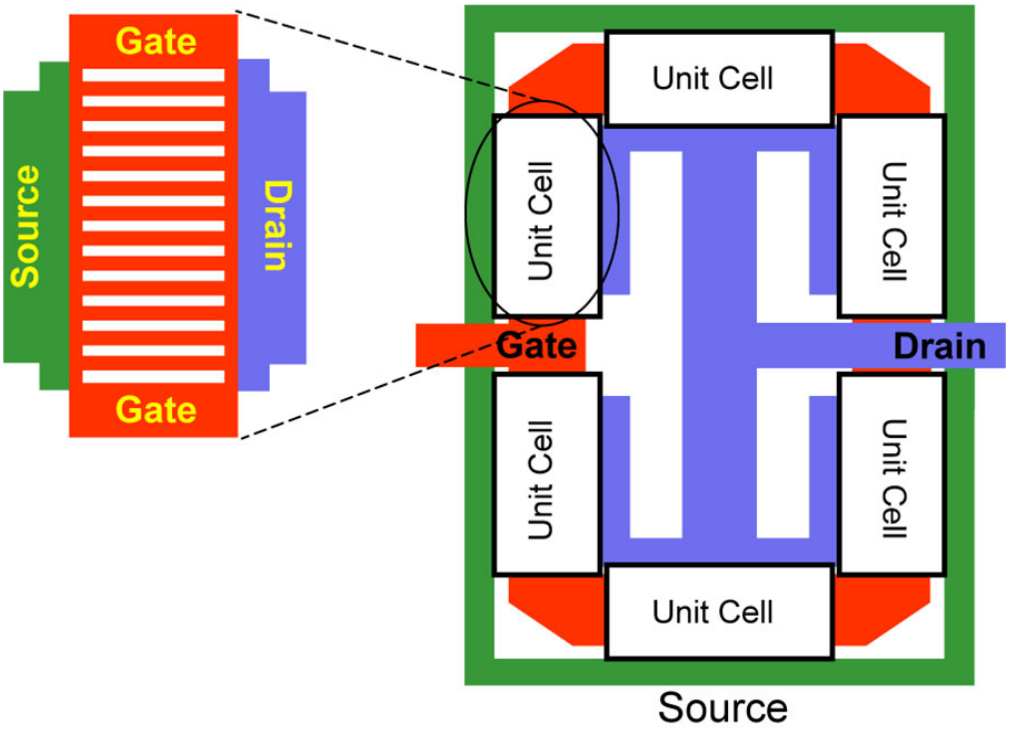


Figure 3.6: Layout of a round-table device.

forms a distributed RC network, double contact reduces the resistance of each finger by a factor of four (37). These cells are then connected in a matrix or circular fashion depending on the desired size of the final transistor. Fig. 3.6 shows a W=60μm NMOS using a circular connection, hence the name "Round-Table". This structure uses external double-contacts (between cells) and multi-path connections between sources and drain of the sub cells. The layout trade-offs were addressed based on the results of the predictive model discussed in the previous section.

Several dimensions of these devices were fabricated in 90nm CMOS process. Measurements were carried out up to 65 GHz and probing pads were de-embedded from the devices. Fig. 3.7 shows MSG, Mason's gain (U) and h_{21} of a W=40 μm

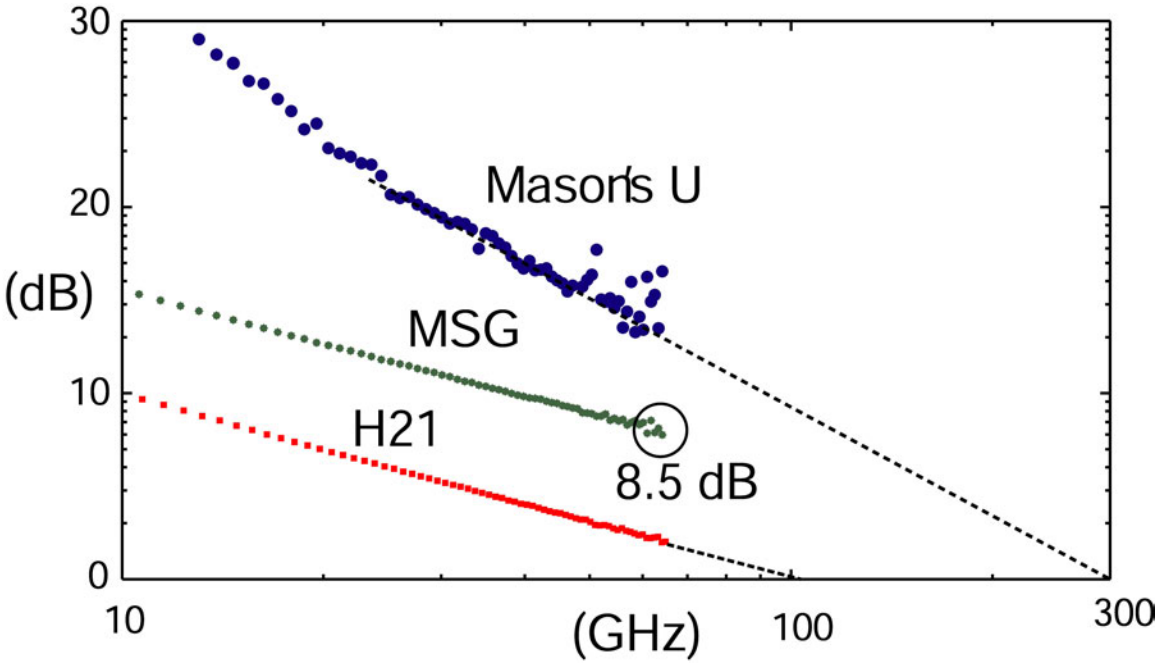


Figure 3.7: Measured h_{21}, Mason's Unilateral Gain (U) and Maximum Stable Gain (MSG) for a $40\mu m$ /90 nm round-table device

round-table NMOS. The f_{max} is calculated by extrapolation of the Mason's gain U for frequencies between 20 GHz to 50 GHz, a frequency range where the most reliable data occurs. As evident, measurements suggest significant improvement in both the speed and the desired gain of these devices as compared to regular RF transistors with the same number of fingers. Even though f_t remains almost constant (100 GHz), f_{max} improved by almost two fold, or to about 300 GHz. This is of course the extrapolated f_{max} since the device introduces new high frequency poles after 100GHz, rendering the linear approximation of the Mason curve inaccurate beyond 100 GHz. Unlike the improvement of the regular multi-finger device presented in the previous section, the MSG of the round-table device increases even at frequencies in which the device is conditionally stable. The MSG

value	Regular	Round Table
$R_g(\Omega)$	4.46	2.23
$R_d(\Omega)$	3.54	2.42
$R_s(\mathrm{m}\Omega)$	627	438
$C_{gs}(fF)$	35.7	57.1
$C_{gd}(fF)$	21.3	17.2

Table 3.1: Parasitic comparison between a regular layout and a round-table layout

at 60 GHz is 8.5dB up from the value of 7.5 dB (regular NMOS) for I=28μA/μm. The ratio f_{max} to f_t, a measure of the optimality of the physical structure of the device, is close to 3, the highest reported for CMOS. The improvement of the maximum stable gain is the result of decreased source resistance and parasitic drain to gate capacitance that both act to decrease the internal series and shunt feedback gains respectively. The improvement of f_{max} was mostly due to reduction in the gate and drain resistances. Table 1 compares the result of extracted small-signal parameters of a round-table W=40m device to a regular optimized multi-finger 40μm transistor. All the resistive losses have been reduced considerably as shown in the accompanying table. The parasitic gate-source capacitance of the device is increased. This is mainly due to the increased overlap capacitance between source and gate in order to reduce the gate and source resistances. This is a good trade-off since the c_{gs} can be tuned out by the matching network.

3.4 mm-Wave Power Device Optimization

The design of a power amplifier hinges around the selection of the appropriate power device. Large devices are quite popular to deliver a large amount of current to the output load in the ouput stage of power amplifiers. A large power device

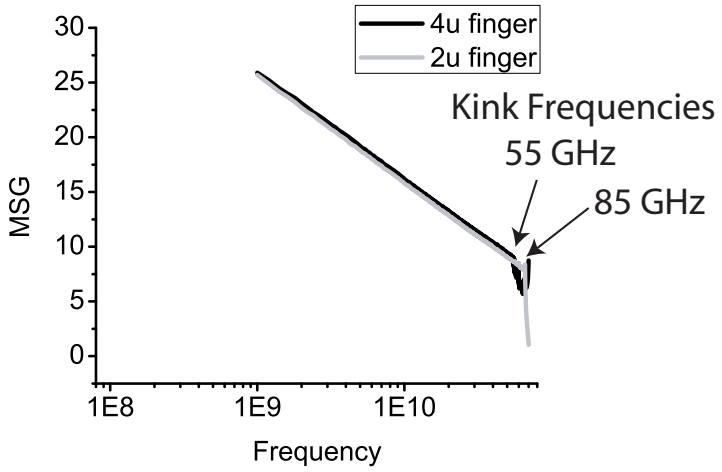


Figure 3.8: A comparison of device power gain as a function of device layout (2μ versus 4μ finger width).

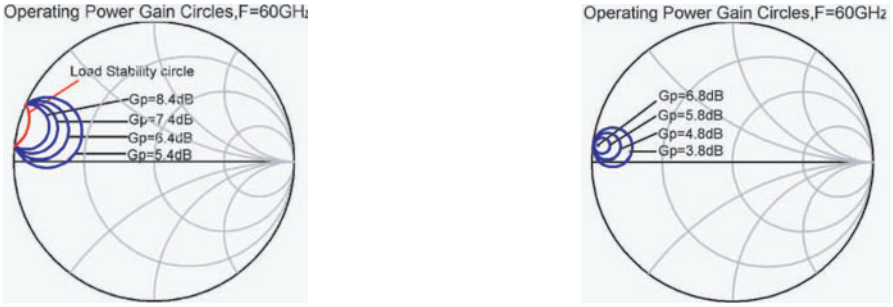


Figure 3.9: Load impedance contours of constant device power gain and contours of constant output power.

can be realized in different ways, with a standard multi-finger layout, an array "round table" layout as described in the previous section, or as a delay equalized structure. The primary considerations for the design of the amplifier include sufficient power gain G_p, stability, output power P_o, and the drain efficiency η. The maximum stable power gain of a 2μ versus 4μ finger width transistor are shown in Fig. 3.9. Both devices have 100 fingers and so the 400μ device

should in theory be able to deliver twice the power of the 200μ device. The 200μ device is biased with 47 mA whereas the 400μ device is biased at 94 mA. It is interesting to note that the 400μ device is unconditionally stable as the gate resistance stabilizes the devices. The 200μ device is only conditionally stable, but the unstable region occurs for only a small inductive range when the load is terminated in a small resistance. The larger device, though, has smaller maximal stable gain ($MSG \sim 6.8$ dB versus $MSG \sim 8.4$ dB). More importantly, the variation in gain is much more rapid as we move away from the optimal point, which means that process variations would lead to more variation of power gain. Utilizing a large device with small finger width, though, is problematic due to the difficulty in making the gate/drain transmission lines. This difficulty is apparent in the layout of such a device (Fig. 3.10), where the transition region introduces extra series resistance and shunt capacitance into the signal path. The measured MSG of this device is less than 5 dB, less than half of the optimal device width.

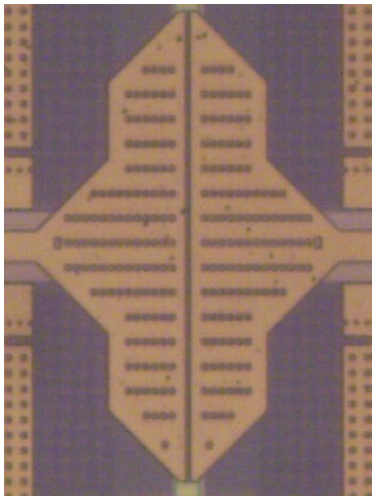


Figure 3.10: A $W = 400\mu$ device realized with 400 fingers.

3.5 Summary

The performance of mm-wave devices can be assessed based on several different parameters such as f_t, f_{max}, Minimum Noise Figure and the amount of available gain. Most of these parameters, especially the f_{max} of the device are strong functions of the physical structure. The performance of these devices can be optimized by a study of the sensitivity of these parameters to the layout parasitics and improving the layout based on these findings. The *round-table* device is a special structure that shows a superior performance with an extrapolated f_{max} of 300 GHz. These optimized devices show a considerable power gain even beyond the f_t of the device. Large devices such as those used in the output stage of power amplifiers need different optimization procedure and the decision about the size and number of fingers and wiring methods should be made accordingly.

4

Mm-wave CMOS Noise Analysis

The range of many wireless communication systems is limited by the sensitivity of their receivers, meaning the minimum amount of signal to noise ratio that the reciever can successfully detect(52). This sensitivity heavily depends on the noise figure of the entire receiver. However since the noise of each stage is normalized by the total gain of previous stages, as Friis equation predicts, the noise figure of the low-noise amplifier essentially dominates the entire receiver sensitivity (53).

Although several noise mechanisms such as flicker, shot-noise and generation-recombination noise can be considered for a MOS device, at high frequencies, the main source of noise that is important to linear circuits is the thermal noise(54). Due to the direct effect of the noise figure of LNA on the performance of the entire receiver, this parameter should be accurately predicted during the design process. Moreover, in the low noise amplifier design, comming up with a simultanous optimization of noise performance with other important circuit parameters such as gain and input match is always a challange and needs an accurate noise model. In this chapter, first, the general noise representation methods for any

two port network is overviewed, then the thermal noise model of CMOS devices is discussed. This is followed by a section about CMOS mm-wave noise model and the behavior of noise parameters versus frequency and layout parasitics. In the end, the developed model is compared with some experimental results in the 50-75 GHz frequencies.

4.1 Two Port Noise Models

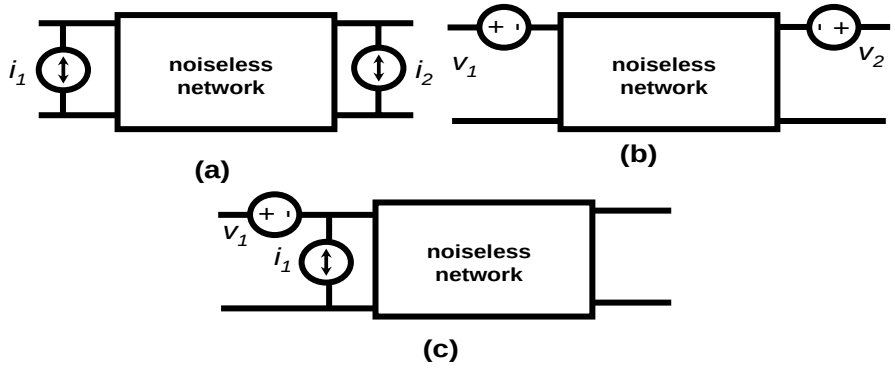


Figure 4.1: Equivalent two-port noise representation (a) Admittance(PRC) model (b) Impedance model (c) ABCD model

Two-port theory provides a mean to represent a noisy two-port in terms of a noiseless two port and its corresponding two noise sources. Modeling the noise of a two port network is essentially based on a generalized Thevenin's theorem. Just like deterministic two port parameters, the noise model can be represented in admittance, impedance or ABCD form as shown in Figure.4.1. These noise sources form a correlation matrix that is Hermitian and non-negative. The PRC

and $ABCD$ representations are the most common forms and their correlation matrix can be calculated using these equations(66; 78):

$$C_y^n = \begin{pmatrix} \overline{i_1 i_1^*} & \overline{i_1 i_2^*} \\ \\ \overline{i_2 i_1^*} & \overline{i_2 i_2^*} \end{pmatrix} \tag{4.1}$$

$$C_A^n = \begin{pmatrix} \overline{v_n v_n^*} & \overline{v_n i_n^*} \\ \\ \overline{i_n v_n^*} & \overline{i_n i_n^*} \end{pmatrix} \tag{4.2}$$

In both of these matrices, C_{11} and C_{22} are positive real numbers and $C_{12} = C_{21}^*$. Thus, C_{11}, C_{22} together with the real and imaginary parts of C_{12} form four noise parameters that are sufficient to fully characterize the noise behavior of a two port network. Most of the time, the $ABCD$ matrix can be calculated starting from the PRC representation using this equation:

$$C_A^n = \begin{pmatrix} 0 & B \\ 1 & D \end{pmatrix} C_Y^n \begin{pmatrix} 0 & 1 \\ B^* & D^* \end{pmatrix} \tag{4.3}$$

In which A, B, C and D are the $ABCD$ matrix elements. Circuit designers are more familiar with another four parameter noise representation

$$F = F_{min} + \frac{R_n}{G_s}|Y_s - Y_{opt}|^2 \tag{4.4}$$

in which F_{min} is the minimum achievable noise figure, R_n is the noise sensitivity resistance, and Y_{opt} is the optimal source noise admittance. Also $Y_s = G_s + jB_s$ is the source admittance of the network.

To bridge between this representation and the ones mentioned earlier, one can use the $ABCD$ representation. This form is a function of the circuit representation as follows:

$$C_A^n = \begin{pmatrix} R_n & \frac{F_{min}-1}{2} - R_n Y_{opt}^* \\ \\ \frac{F_{min}-1}{2} - R_n Y_{opt} & R_n |Y_{opt}|^2 \end{pmatrix} \tag{4.5}$$

The four circuit parameters can be calculated by solving the corresponding 4 equations (78).

4.2 CMOS Noise Model

The main source of transistor noise in mm-wave frequencies is thermal noise. Thermal noise as Johnson formulated for the first time is the result of the kinetic energy of particles (55). These thermally-excited particles in a conductor undergo a random walk Brownian motion via collisions with the lattice of the conductor. This random walk produces random electrical characteristics in the device terminals.

Among the various methods proposed for MOSFET noise modeling, the Van Der Ziel model is the most widely accepted one(56). Van Der Ziel modeled the FET noise as a voltage modulated resistor, capacitively coupled to the gate as depicted in figure 4.2. This way, two noise sources for the channel resistance and the induced gate noise are calculated. As the source of both these noises are the channel noise, there is a strong correlation between these two sources. However due to the distributed nature of the channel that translates to infinite number of

small noise sources along the channel, the represented sources are not completely correlated. Van Der ziel model is essentially a PRC model as described earlier. The value of the two sources and their correlation can be calculated using these equations:

$$i_d^2 = 4kT\Delta f\gamma g_{d0} \tag{4.6}$$

$$i_g^2 = 4kT\Delta f\delta g_g \tag{4.7}$$

$$c \triangleq \frac{i_g i_d}{\sqrt{\bar{i}_g^2 \bar{i}_d^2}} \tag{4.8}$$

$$g_g \triangleq \zeta\frac{\omega^2 C_{gs}^2}{g_{d0}} \tag{4.9}$$

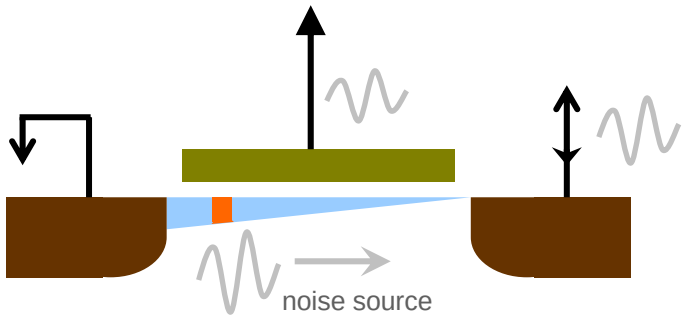


Figure 4.2: Generation of channel and induced gate noise in MOSFET.

where γ, δ, and ζ are bias-dependent factors; g_{d0} is the drain output conductance under zero drain bias; g_g is the real part of the gate-to-source admittance; and c is the cross correlation coefficient.

For a long channel MOSFET, γ is 2/3 when the channel is pinched off and 1

when the channel is symmetric. (56). Values of δ, ζ and c could be also calculated to 4/3, 0.2 and $j0.395$ respectively. Although the model is well matched with long channel transistors, substantial increase has been observed and reported in both γ and δ for short channel MOSFETs (61; 62; 63). There have been extensive discussions on the amount of noise increase and subsequently the value that γ an δ take as well as the physical source of the origin of this excess noise. However, recently there has been a consensus that the noise parameter, γ, is substantially smaller than what initially had been assumed and its value for a saturated MOSFET is close to twice as its long channel value.(58; 60; 64)

4.3 mm-wave Noise Model

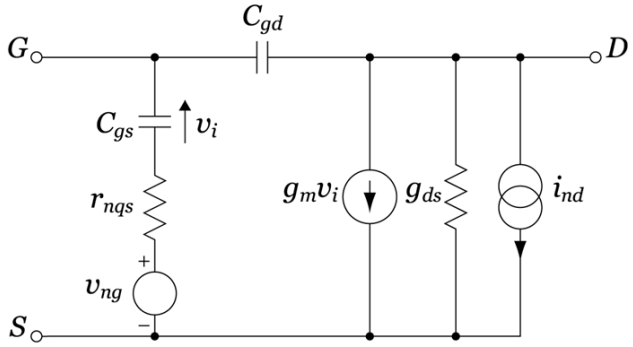


Figure 4.3: Pospieszalski model assumes two uncorrelated noise sources, r_{gs} at T_g and r_{ds} at T_d.

The main problem with PRC and ABCD models is that the correlation between the two noise sources, makes the simulation difficult as few circuit simulator tools offer correlated noise sources. To remedy this problem, Pospieszalsky proposed a new model, as shown in figure 4.3, based on two uncorrelated noise sources in the source and drain sides (57). This model, assumes two uncorrelated

noise sources, r_{gs} and r_{ds} to model the channel noise. The temperature of these two resistors are set to T_g and T_d, the only model parameters, respectively where T_g is close to the environment temperature while T_d a is much higher temperature and could go up to several thousands. In this model, the r_{gs} noise is responsible for the correlated part of the channel noise as if the T_d is set to zero, the model represents a noise process with $\rho_c = -j$.

Based on the procedure described in the previous section, the noise optimal source impedance can be calculated using the C_y and $ABCD$ matrices of the device. Ignoring the effect of $C - gd$ for now, the R_{opt} and X_{opt} can be calculated using these equations:

$$R_{opt} = \sqrt{\left(\frac{f_t}{f}\right)^2 \frac{r_{ds}T_g(R_g + r_{gs})}{T_d} + (r_{gs} + R_g)^2} \tag{4.10}$$

$$X_{opt} = \frac{1}{\omega C_{gs}} \tag{4.11}$$

At low frequency, when f is much smaller than f_t, the second term in the R_{opt} equation could be neglected and the equation simplifies to

$$R_{opt}^{lowf} = \frac{f_t}{f}\sqrt{\frac{(R_g + r_{gs})r_{ds}T_g}{T_d}} \tag{4.12}$$

This value could be much larger than $R_g + r_{gs}$, the optimal gain resistance. This considerable difference, makes it crucial to use simultaneous noise and gain optimization techniques.

As the frequency approaches the f_t of the device, this approximation becomes invalid. In fact, the two terms become comparable in value for such frequencies.

To get an approximate value for R_{opt} in this case, we need few assumptions. The value of r_{gs} is equal to $\frac{1}{ng_m}$ and n is a value between 3 to 5. We can also assume that R_g, by using some good layout techniques, is roughly equal to r_{gs}. We can further assume that T_d is 10 to 15 times larger than T_g and g_m is roughly 10 times larger than g_{ds}, the R_{opt} will become

$$R_{opt} = k_n(R_g + r_{gs}) \qquad k_n = \sqrt{2}....2 \tag{4.13}$$

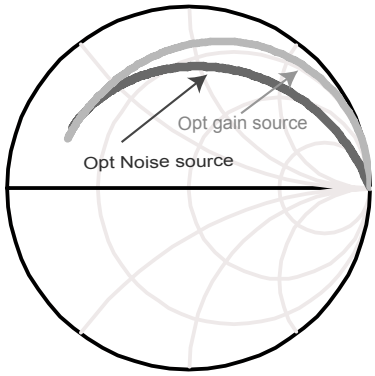
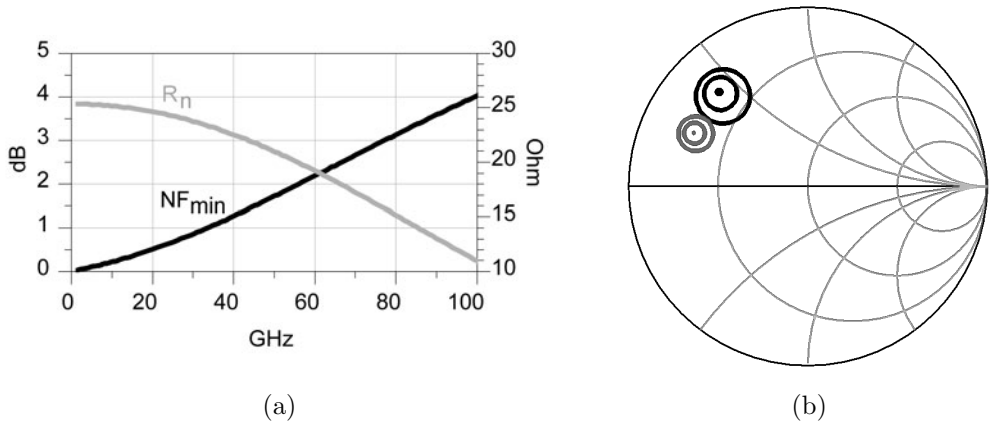


Figure 4.4: Optimal noise(gray) and gain(black) impedances for a 40μ device for f=1 to 100 GHz.

In fact the optimal noise resistance is only a factor of k_n larger than the optimal gain resistance. In this simple model, the imaginary part of optimal noise and gain sources are also equal. As a result the optimal noise and gain source impedances approach each other as evident on Fig.4.4.These imply that, if the R_n is sufficiently small, minor compromise in the noise and gain performance, can result in the simultaneous noise/power match as we approach and pass the f_t of the device. Even small deviation from the optimal noise source, can have a considerable noise penalty if the value of R_n is large. Calculation of R_n is more

complicated and needs considering the effect of the gate to drain capacitance of the device to show the real trend.The value of R_n is equal to the value of A in the $ABCD$ noise matrix of the device, considering both R_g and C_{gd} and is equal to[1]:

$$R_n = r_{gs}\frac{T_g}{T_0}\frac{g_m^2}{(g_m + C_{gs}C_{gd}r_{gs}\omega^2)^2 + C_{gd}^2\omega^2} + \frac{T_d}{T_0}\frac{g_{ds}(1 + r_{gs}^2C_{gs}^2\omega^2)}{(g_m + C_{gs}C_{gd}r_{gs}\omega^2)^2 + C_{gd}^2\omega^2} \tag{4.14}$$



(a) (b)

Figure 4.5: (a) Simulated NF_{min} and R_n for a 40μ common-source round-table device. (b) Noise and gain circles for optimal, optimal-0.3dB and optimal-0.5dB at 60 GHz

Both these terms show a reduction with the frequency and the simulated R_n is shown in figure 4.5(a). The implication is that the penalty to be paid as a result of a deviation from the optimal noise impedance gets smaller as the frequency approaches f_t. In fact this together with the closeness of the optimal noise and gain impedances, suggest that with a compromise of about 0.5 dB in gain and noise figure, one can get a simultaneous noise and gain match as depicted in

[1]The detail of the calculation can be found in Appendix 1.

fig.4.5(b).

It is worth mentioning the importance of considering the C_{gd} of the device in the R_n calculation. If the C_{gd} is neglected, equation A.4 simplifies to the more familiar equation:

$$R_n = r_{gs}\frac{T_g}{T_0} + \frac{T_d}{T_0}\frac{g_{ds}}{g_m^2}(1 + r_{gs}^2 C_{gs}^2 \omega^2) \tag{4.15}$$

This would imply a direct relation between R_n and frequency that could be misleading. In fact, high frequency device noise measurements for non-silicon technologies, had shown the reduction of R_n with frequency and the effect is verified for CMOS as will be represented in the next section (59; 65).

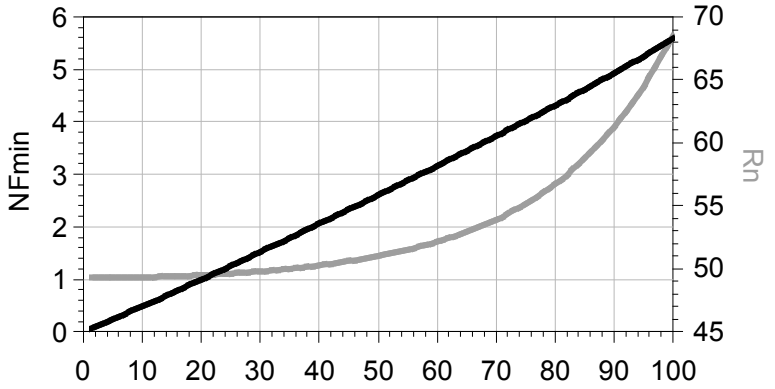
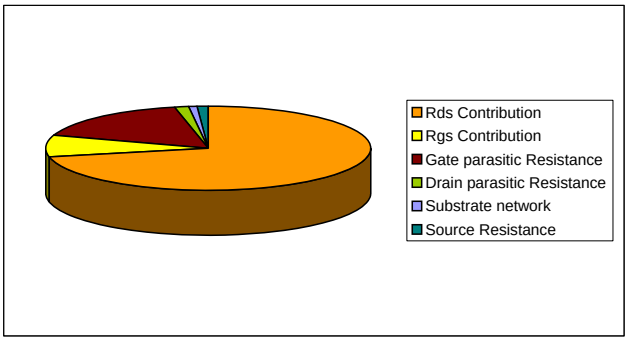


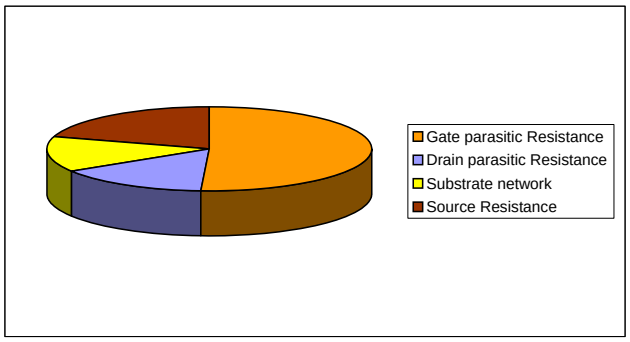
Figure 4.6: NF_{min} and R_n for a 40μ cascode device.

It is expected however, that if the output and input of the device are decoupled, as is the case for cascode devices, the R_n experiences a moderate increase with the frequency in the mm-wave region as predicted by equation 4.15.This prediction is verified both in simulation and measurements. The simulated R_n and F_{min} for a cascode device is shown in 4.6. This increase has also been observed

in measurements of $0.13\mu m$ cascode devices (73).



(a)



(b)

Figure 4.7: (a) Noise contribution of various noise sources for a round-table device (b) Noise sensitivity of the device to parasitic noise sources

4.4 Noise Sensitivity Analysis To Parasitics

Determining the noise contribution of various noise sources and the sensitivity of the noise performance to their values are essential specially when it comes to optimize a device for a low noise applications. Fig.4.7(a) shows the contribution share of different noise sources of a round-table device based on the proposed model.

The sensitivity of the noise figure to the increase in the parasitic resistances have been also shown in the Fig.4.7(b). It is clear that with a reasonable layout to keep the parasitic gate resistance low, most of the device noise comes from the drain side noise source. The parasitic gate resistance is the second large contributor and it is the main controllable noise source that shows a large sensitivity as well. This suggest that even further reduction of R_g could still significantly help in terms of noise performance of the device and should be considered for devices specifically tailored to low noise applications[1]. The source resistance, although showing rather a large sensitivity, does not contribute to the overall noise figure as it is sufficiently small for the round-table device.

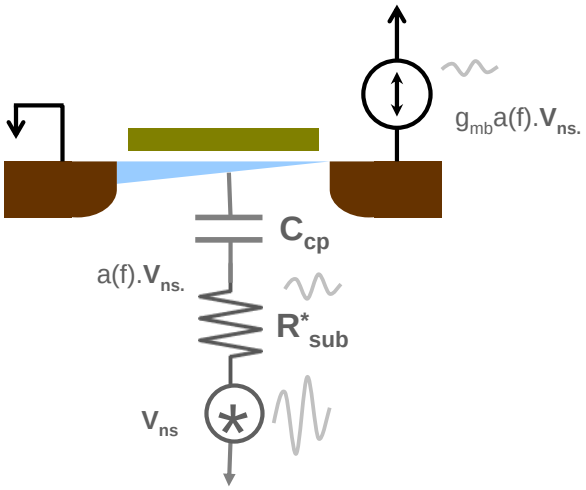


Figure 4.8: The substrate noise coupling to the channel.

The drain and substrate resistances demonstrate low noise contribution as well as low sensitivity. The noise of the drain resistance is scaled down by the

[1]The detail of the relationship between R_g and the device layout has been discussed in Chapter 3.

gain of the device. The substrate resistance noise capacitively couples to the channel and can add to the drain current noise based on the effective g_{mb} of the device. This is shown in Fig.4.8 and can be quantitavely shown as:

$$\bar{i_{sub}^2} = 4kT_0 \frac{R_{sub}^* g_{mb}^2}{1 + \omega^2 C_{cp}^2 R_{sub}^2} \tag{4.16}$$

R_{sub}^* is the effective substrate resistance assuming a simple one resistor substrate network and C_{cp} is the associated coupling body to channel capacitor as depicted in figure4.8. Due to the low pass nature of the filter formed by the substrate resistance and the coupling capacitor, this noise is not of a great importance in high frequency as verified by sensitivity and noise contribution analysis.

4.5 Experimental Results

For the complete noise simulation, the small signal model as shown in figure 4.9 has been used. The T_d and T_g of the device have been set to 4200K and 310K while all the parasitic resistances have the environment temperature to fit the measured data.

The noise measurement is performed for the frequency range of 50 to 75 GHz although the data in the lower frequency range is very scattered and most of the reliable data occur after 60 GHz. The noise data was de-embedded using the recursive de-embedding method described in chapter two. Figure 4.10(a) and 4.10(b) compare the simulated and measured minimum noise figure and noise resistance for the round-table $40\mu m$ device. The slight increase in the NF_{min} as well as the predicted reduction in R_n can be seen in these measurements while both values have good agreements with the model.

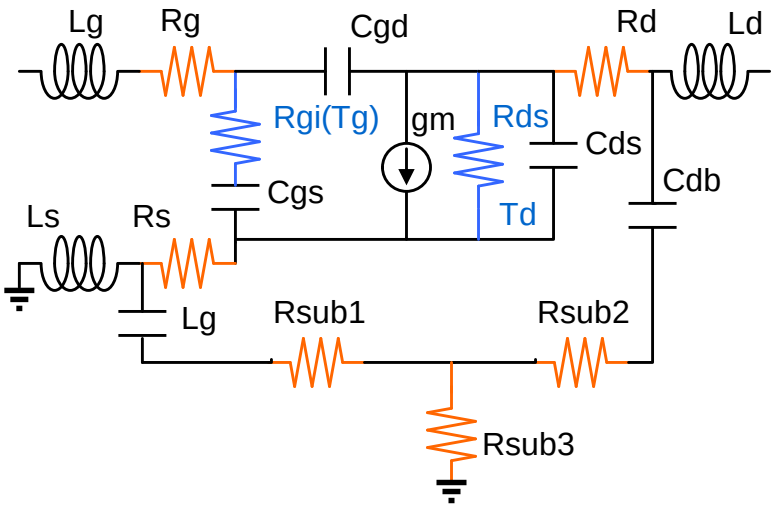
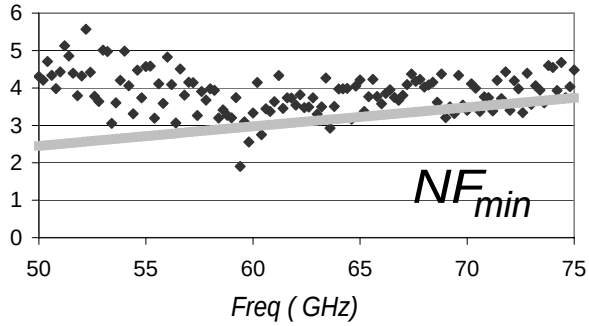


Figure 4.9: The employed model for noise analysis.

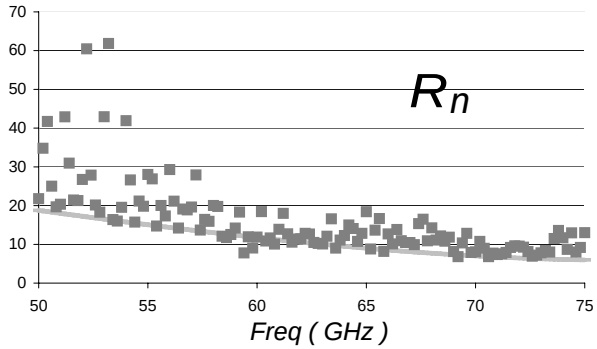
Figure.4.11(a) and 4.11(b) demonstrates the measured and modeled optimal noise impedance and compares it to the measured optimal gain impedance. While the model and measurement show a good agreement, the optimal gain impedance is also very close to the corresponding noise impedance as was explained and predicted in the previous section.

4.6 Summary

This chapter talked about noise behavior of CMOS devices in the mm-wave frequency. It was shown that the noise could be modeled using two independent noise sources at the source and drain sides of the transistor with proper choices of source and drain temperatures. The optimal noise impedance at the source of the device approaches the optimal gain impedance as the frequency approaches the f_t



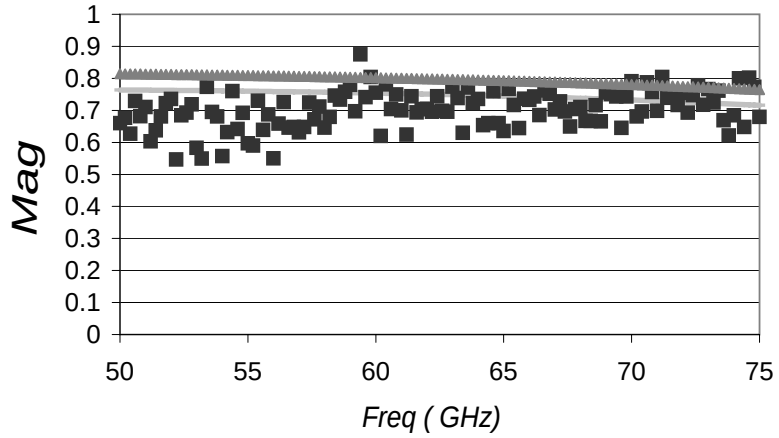
(a)



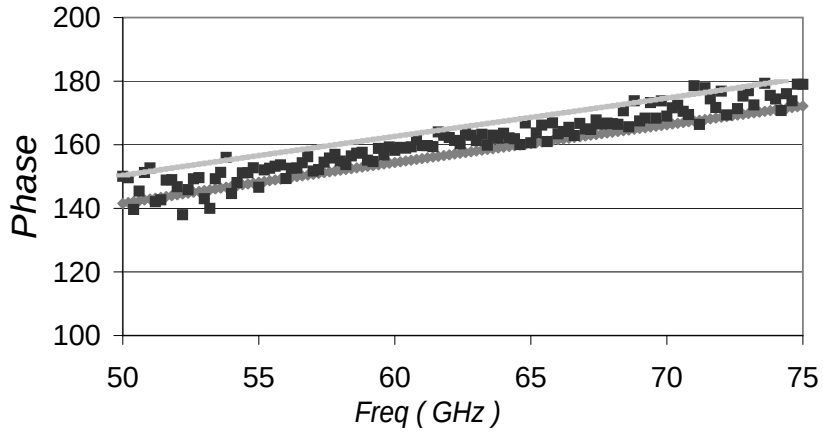
(b)

Figure 4.10: Simulated and measured (a) NF_{min} and (b) R_n for a 40μ common-source round-table device

of the device. On the other hand, the noise sensitivity of the device, R_n reduces with frequency, making the overall noise figure less sensitive to deviations from the optimal noise impedance. These imply that a noise figure close to NF_{min} is achievable once an amplifier is designed to maximize the gain. It was also shown that even in the optimal layouts, the gate resistance still contributes significantly to the overall noise figure of the device.



(a)



(b)

Figure 4.11: Simulated (solid) and measured (Black dotted) optimal noise impedance and measured (Gray dotted) optimal gain impedance.(a) Magnitude (b) Phase for a 40μ common-source round-table device.

5

mm-wave Amplifier Design

Amplifiers are essential building blocks in any communication system. For mm-wave circuits, low noise amplifiers are needed as the first block in the receiver chain to amplify the received signal and suppress the noise effect of all the subsequent stages in the chain and they essentially determine the receiving range of the system. In the transmitter, power amplifiers are the critical blocks that send out the modulated signal of the transmitter to the antenna and determine the transmit range of the system.

In this chapter, first, basic considerations and design procedure of mm-wave amplifiers are discussed. Then the design and implementation of a 60 GHz low noise amplifier is shown. Another two stage 60 GHz LNA with an interstage matching between two common-source and common-gate blocks is next discussed. These will follow by a design of a two stage 60 GHz class A power amplifier that delivers 6.7 dBm of output power with 20% of power added efficiency. All the circuits were implemented in a standard digital 90 nm CMOS technology.

5.1 mm-wave Amplifier Design

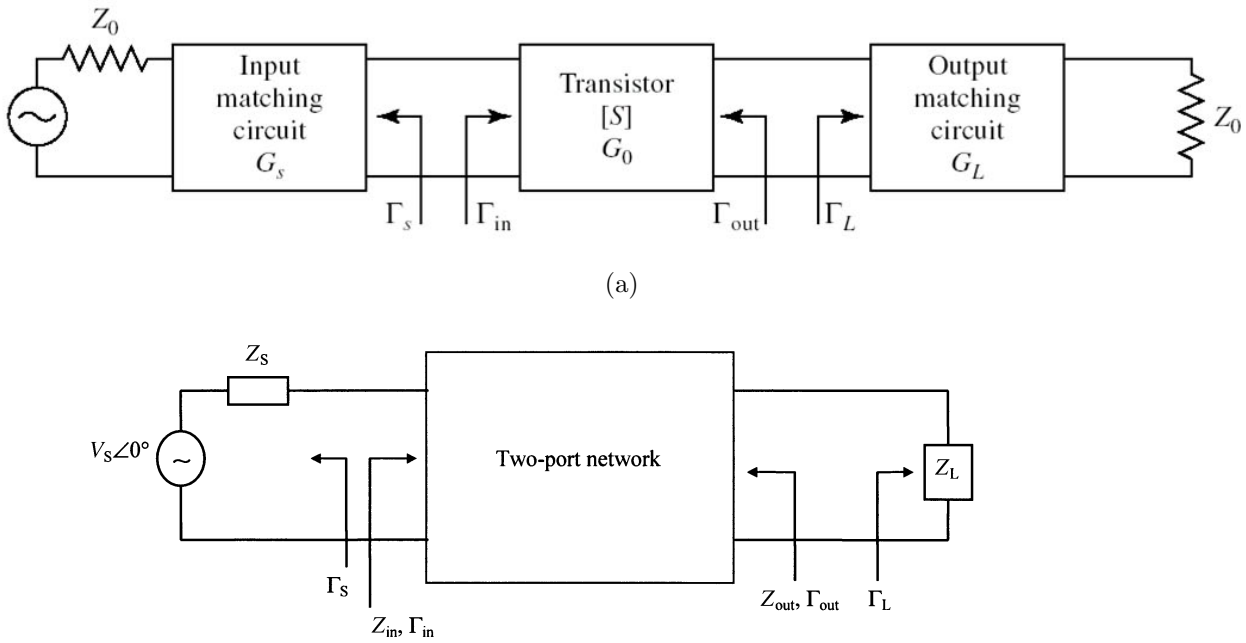


Figure 5.1: (a) The block diagram of a typical mm-wave amplifier including matching networks. (b) Equivalent diagram of the same amplifier.

Mm-wave amplifier design needs consideration of several factors. Choosing an efficient and high performance device is a critical issue both for the transistors and also passive devices used in matching networks. Having accurate models of these devices as discussed in chapter 2 is also of a great importance. The amplifier should exploit the maximum possible gain of the device while watching for potential instabilities. Depending on the application, noise and linearity considerations must also be included. Device optimization and modeling as well as noise considerations were discussed in previous chapters. Here in this chapter we consider other relevant issues. Fig.5.1(a) shows the general schematic of a high frequency single stage amplifier. The device is connected to the input and

output loads, usually 50Ω loads, using two matching networks. These matching networks, depending on the application, are designed to increase the power gain of the amplifier, minimize the noise figure or increase the power efficiency at the output. In mm-wave, an important part of amplifier design is coming up with proper input and output matching networks to optimize these parameters.

5.1.1 Stability Analysis

The stability of an amplifier is a critical consideration. An amplifier could easily turn into an oscillator (not a good one though!) if enough care is not taken. The oscillation possibility of a two port network can be determined from its S-parameters (or any other two port parameter) and the input and output matching networks. The stability analysis becomes easier if the amplifier schematic is simplified to what is shown in Fig.5.1(b). Potentially, oscillation is possible if either the input or the output have a negative resistance. This translates to having both $|\Gamma_{in}|$ and $|\Gamma_{out}|$ larger than one(80). Considering a matched two-port network as shown in fig.5.1(b), it is clear that the stability of the circuit depends on the value of Z_s and Z_L as these values are added to the input and output impedance of the network. If a two port network is stable for only some values of the source and load impedances, its stability is conditional. Likewise,unconditional stability happens when the circuit is stable for all positive real values of Z_s and Z_L. In

this case, these conditions should be satisfied:

$$|\Gamma_s| < 1 \tag{5.1}$$

$$|\Gamma_L| < 1 \tag{5.2}$$

$$|\Gamma_{in}| < |S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}| < 1 \tag{5.3}$$

$$|\Gamma_{out}| < |S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s}| < 1 \tag{5.4}$$

Since the input impedance of the network is a function of the load impedance,

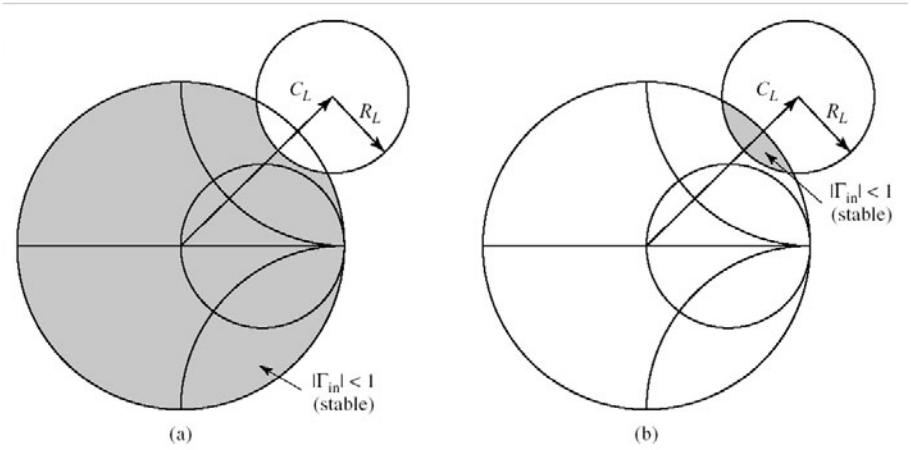


Figure 5.2: Stable region on the Smith Chart (a) if $|S_{11}| < 1$ (b) if $|S_{11}| > 1$

the input stability depends on the load. This determines certain values of Z_L that makes the input unconditionally stable. On the Smith Chart, the boundary between these two regions is a circle and is called stability circle(72). The magnitude of S_{11} determines whether the stability happens for load values inside or outside the stability circle. If $|S_{11}| > 1$, which is the case for most of practical mm-wave CMOS amplifiers, the area outside the circle ensures stability as shown on figure5.2. Same argument holds for source stability circles.

To make a network stable for any value of the source and load impedance, both these stability circles need to fall outside the unit circle on the Smith Chart, assuming positive $|S_{11}|$ and $|S_{22}|$. Solving equations 5.4 gives a necessary and sufficient condition for stability:

$$k > 1 \tag{5.5}$$

$$|\Delta| < 1 \tag{5.6}$$

Where

$$k = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|} \tag{5.7}$$

And

$$\Delta = S_{11}S_{22} - S_{12}S_{21} \tag{5.8}$$

k is called the stability factor and is a function of frequency. As the frequency increases, the gain of the device drops, making the device more and more stable. This presents itself in the form of stability circles moving outside of the unit circle and k approaching one. Figure 5.3 shows load and source stability circles for a typical device at 30,60,90 and 120 GHz and shows that the device is unconditionally stable at 120 GHz where both stability circles are outside. For frequencies where k is smaller than one, matching networks have to be designed to keep a maximum distance from stability circles to ensure stability. There are other stabilizing methods such as adding extra resistance to the input and output. But these methods have noise and power gain penalty.

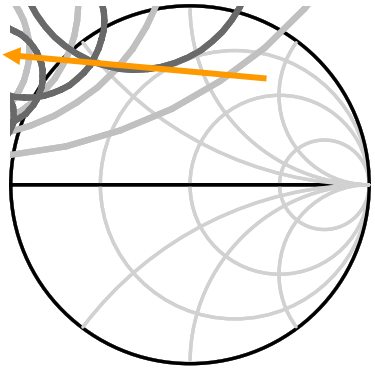


Figure 5.3: Input(light) and output(dark) stability circles for 30, 60, 90 and 120 GHz for a typical 90nm common-source device, circles move toward out of the unit circle as the frequency increases.

5.1.2 Bilateral Amplifier Design Procedure

If the amplifier device is unilateral,i.e. S_{12}=0, then the input and output matching networks can be designed independently as the input and output impedances are not functions of the load and source. In this case, the Γ_s and Γ_L are set equal to S_{11}^* and S_{22}^* respectively [1]

In practice CMOS devices are far from being unilateral at mm-wave frequencies. In this case, the source and load matching networks are interdependent and should be designed simultaneously to satisfy these conditions:

$$\Gamma_s = \Gamma_{in}^* \tag{5.9}$$

$$\Gamma_L = \Gamma_{out}^* \tag{5.10}$$

[1]This is assuming that the magnitudes of S_{11} and S_{22} are less than one which is usually the case for CMOS devices at mm-wave frequencies.

Substituting for Γ_s and Γ_L, we can write

$$\Gamma_s^* = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \tag{5.11}$$

$$\Gamma_{out}^* = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1 - S_{11}\Gamma_s} \tag{5.12}$$

These two equations give the values of Γ_L and Γ_s required for a simultaneous conjugate match. If the device is unconditionally stable, then the design is basically solving for these two equations and designing the appropriate matching networks. Some design tools such as ADS would also provide assistance with these equations.

For the case where the device is conditionally stable, the design procedure is somewhat different. Here, the load matching network is first designed based to optimize the G_p. G_p is the ratio of the power deleivered to the load by the available power at the input of the two port network and is independent from the impedance at the source.

$$G_p = \frac{|Y_{21}|^2}{|Y_L + Y_{22}|^2} \frac{\Re(Y_L)}{\Re(Y_{in})} \tag{5.13}$$

Load impedances that reslut in a certain value of G_p form a circle on the Smith Chart. Using G_p circles that are source independent the proper load impedance is determined to maximize the power gain. Then the value of Γ_{in} is calculated based on the selected load impedanc and the input matching network is designed to form a conjugate match at the input. Note that the load impedance may need to change if it generates instability at the input.

5.1.3 Matching Networks Passive Devices[1]

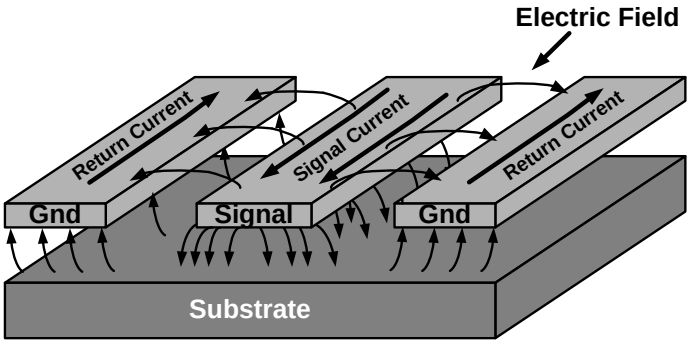


Figure 5.4: A Coplanar transmission line(77).

The matching networks in mm-wave frequencies can be best realized using transmission lines. The relatively small wavelength allows on-chip long structures such as quarter wave lines. The main advantage of transmission lines in high frequency is that there is no ambiguity in the definition of reference planes since the signal and ground are always well defined. Using transmission lines also allows to use them also to implement interconnection wirings and incorporate them as a part of matching networks.

Among several different proposed structures for transmission lines, coplanar waveguides (CPWs) were selected and implemented. The main advantage of CPW, as shown in Fig.5.4, over microstrip line - the other commonly used on-chip transmission line - is its considerably higher Q_L that makes it a better choice for tunning of device capacitors in a matching network (16). The adjustment of Z_0 can also be easily done by adjusting the signal to gap spacing of the line (77).

[1]This modeling of passive devices in this section is mainly based on the works of my colleague, Mounir Bohsali.

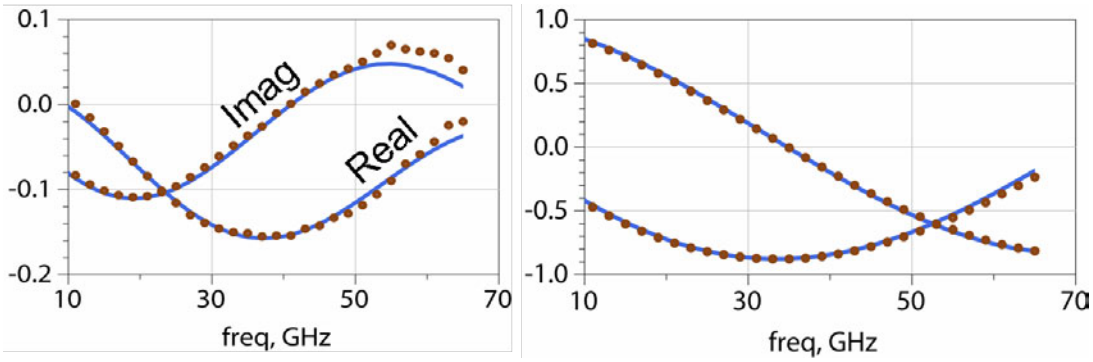


Figure 5.5: Simulated and measured S-parameter for a transmission line with W=10μm and S=7μm.

Transmission lines were designed using $3D$ electromagnetic simulators and a length scalable model was made using a generic lossy transmission line model in Agilent ADS. This model is measurement based and needs to be adjusted separately for each signal to ground gap spacing value. Figure.5.5 shows a sample S-parameter measurement versus simulation for a CPW with W=10μm and S=7μm.

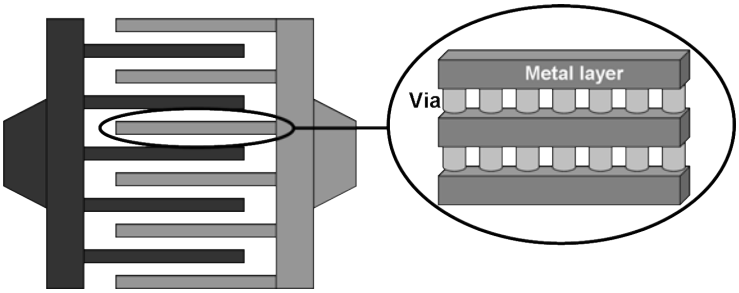
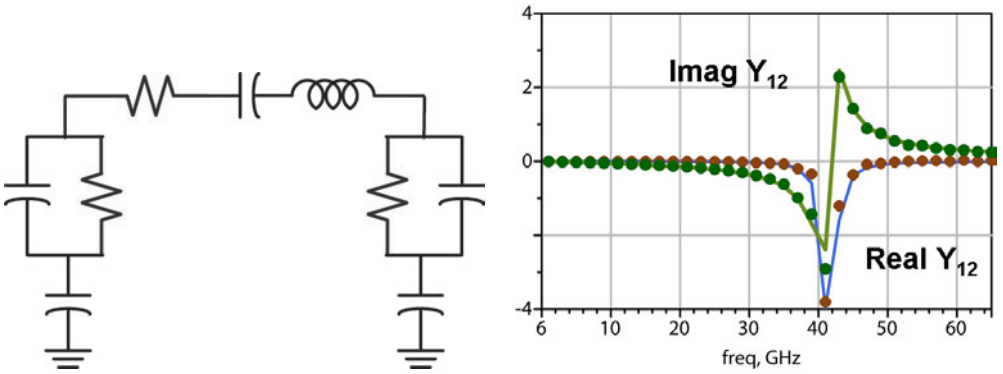


Figure 5.6: The conceptual picture of a finger capacitor.

Another important matching network component is capacitor. Capacitors are used mainly for DC-bypass and AC-coupling purposes. Finger capacitors, also referred to as "MOM" (metal-oxide-metal) capacitors, were used in all designs.

These capacitors can easily be fabricated in any modern CMOS process due to the abundance and high density of metal layers. Finger capacitors, as shown in Fig.5.6, consist of several fingers, each using all available metal layers connected together with the maximum number of vias allowed in the process. Given that now special layer is needed for such kind of capacitors, they are less expensive than Metal-Insulator-Metal (MIM) capacitors.



(a) (b)

Figure 5.7: (a) Equivalent circuit for a MOM capacitor. (b) Y-parameters measurement and model for a 1pF capacitor.

The model for these capacitors is also measurement based and is implemented using the equivalent circuit as shown in Figure.5.7(a). Figure.5.7(b) shows the accuracy of the model by comparing Y-parameters of a measurement versus the model. As evident from this figure, the self resosnace frequency of these capacitors happen somewhere in the mm-wave band. The implication is that the series inductance needs to be carefully modeled and be incorporated as a part of matching networks. This will shorten the physical length of transmission lines connected to such capacitors.

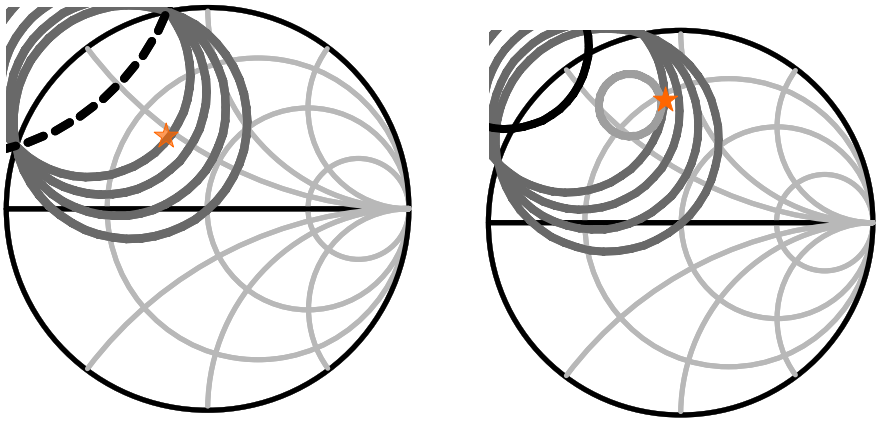
5.2 A 60 GHz Low Power, Low Noise Amplifier

5.2.1 Design Issues

The low noise amplifier consumes a considerable part of the front end power and any power reduction of this block would directly affect the total power dissipation of the system. For mm-wave LNA block , there are mainly two reasons for the high power dissipation. First of all, due to the limited gain of each device at 60 GHz, the designer is forced to increase the number of stages in the amplifier. On the other hand, since the minimum noise figure of the device is inversely proportional to the size of the transistor, large devices need to be selected and the high power consumption is the result of multi-stage of large devices. In this design, a low noise, low power 60GHz amplifier is designed using an optimized round-table common-source device. On one hand, benefiting from the high available gain of the round-table device can reduce the number of stages of the amplifier. Since the round-table device has a considerably smaller gate resistance compared to other structures, it can provide the same noise performance as regular devices with twice the size.

Another difference between this design and many other 60 GHz amplifiers is that it uses common-soruce devices instead of the commonly used cascode structures. Previously reported mm-wave CMOS amplifiers use mostly cascode devices. One advantage of cascode devices is that they can be made unconditionally stable at the operating frequency, making the design more robust while simplifying the design of matching networks. However the noise figure of a cascode is higher compared to common source device at high frequency due to the reduced degeneration on the cascode device. As a result one will need to select

larger devices and increase the dissipated power. The R_n of cascode devices also show an increase with frequency unlike common source devices. On the other hand, as will be seen in the next chapter, cascode devices do not offer much gain advantage in mm-wave as they normally do for lower frequencies and in fact after a certain frequency, their available gain could be even lower than a common-sorce device with the similar size. Linearity of common-source devices is also often superior to cascodes due to the more voltage headroom available at the output. Moreover, the maximum stable gain of common source devices with similar layout methods is more or less independent of the width, enabling low power design by using smaller devices. As a result, common-source amplifiers are more efficient in terms of gain and noise perforamnce. However one should be exteremely careful with the stability issues as most of high gain devices are also susceptible to oscillation.



(a) (b)

Figure 5.8: Output and input impedance selection method at 60 GHz for the LNA device. (a) Gain (G_p) circles and output stability circle. (b) Gain (G_s) circles (dark gray),Noise circles for NF_{min} and $NF_{min}+3$ (light Gray) and input stability circle (black).

Round-table common source devices used in this design were conditionally stable up to 95GHz, requiring careful design of the source and load networks. Consequently for the 60GHz LNA, the source and load impedances for each stage should be selected at a safe distance from the instability regions while maintaining a good gain and noise performances (80). As shown in the previous chapter, in mm-wave, the simultaneous noise and power match can be achieved without using the inductance degeneration technique as the optimal source impedance for noise and input match become very close to each other at the frequencies close to the f_t of the device. Figures 5.8(a) and 5.8(b) show the impedance selection method for the device based on the stability, gain and noise circles. The load impedance is selected on the maximum G_p circle and keeps a distance from the stability circle. Accordingly, in the input, stability, gain and noise circles are drawn. As shown, the optimal gain can be achieved by choosing the intersection point between the maximum gain circle and the $NF_{min} + 0.3$dB circle.

CPW transmission lines are used extensively in the design for impedance matching, interconnect wiring, and the bias networks. Transmission line lengths are kept much shorter than $\lambda/4$ in order to reduce losses and minimize the noise contributions at the input of the amplifier. Finger MOM capacitors are used for AC-coupling and as bypass capacitors for the DC feed lines. These capacitors are non-ideal and behave inductively at 60 GHz due to their self-resonance frequency (47 GHz). This does not cause any issue as long as they are well modeled and the associated inductance is used as a part of the matching network to shorten the length of transmission lines.

5.2.2 Experimental Results

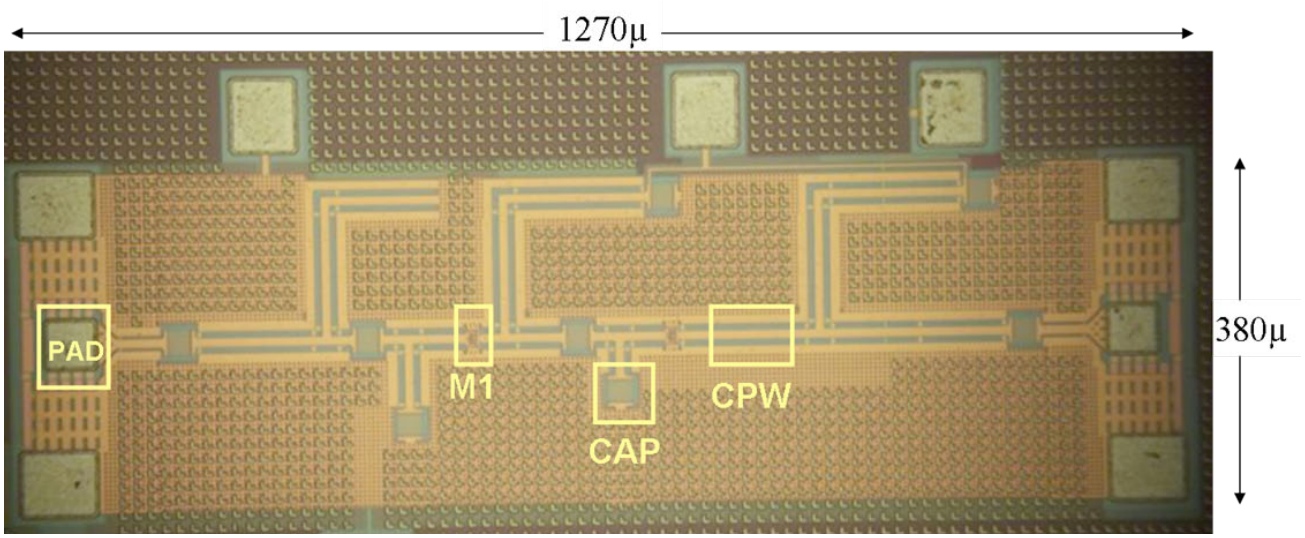


Figure 5.9: The die photo of the 60GHz amplifier.

The chip micrograph of the 60 GHz LNA is shown in Fig.5.9. On the figure, various building blocks of the circuit are highlighted. The input and output pads are also modeled and used as a part of matching networks. This enables us to rely on the measured data without any de-embedding process. Most of the surface of the chip is covered by a slotted sheet of the top metal as can be seen on the figure. This is a mandatory design rule about the minimum and maximum densitiy of the top metal layer per any unit area. Metal slots in the ground planes should have a safe distance, several times larger than the skin depth of the metal at the operating frequency, to avoid increasing losses of transmission lines.

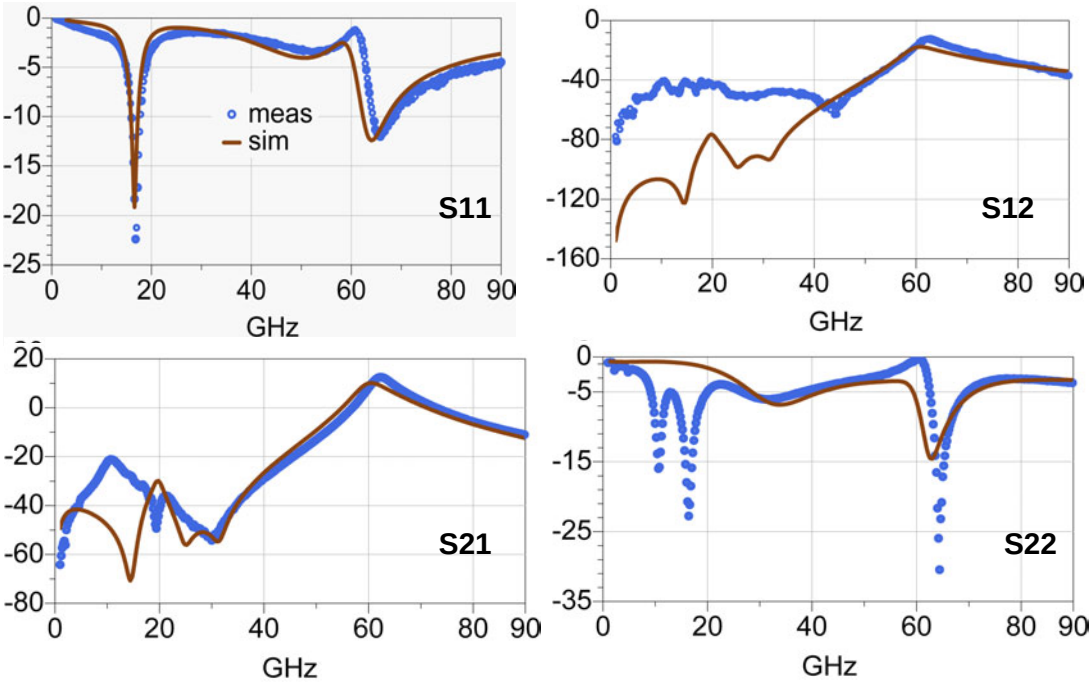
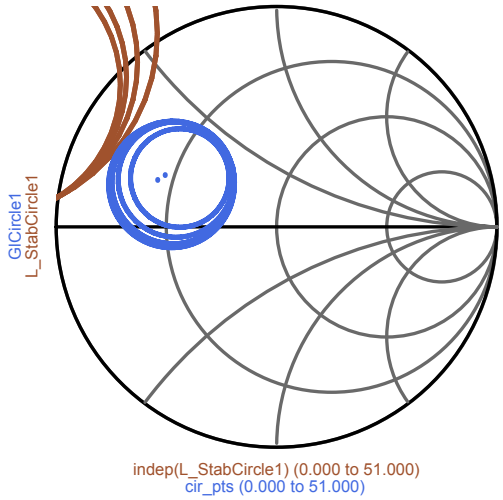


Figure 5.10: Measurement(dotted) and simulation (solid) S-parameters (in dB) of the 60GHz LNA.

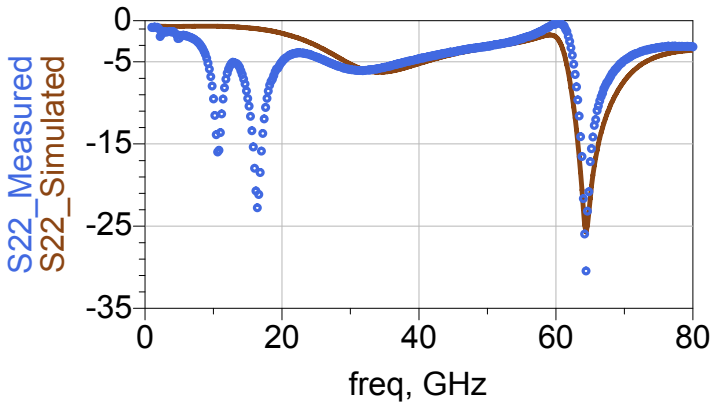
5.2.3 Experimental Results

The circuit was probe-measured up to 65 GHz and the measured and modeled S-parameters for the 60-GHz amplifier are shown in Fig.5.10. The amplifier achieves a peak power gain of 12.2 dB at 63 GHz and the input and output return losses are -13 dB and -25 dB respectively. The measurement is in a good agreement with the simulation with a 1 GHz of frequency mismatch. Also the output match of the circuit shows some discrepancy with the simulated result, pushing the circuit to the edge of instability. Process variation could explain the frequency mismatch, and based on the post measurement simulations, it was verified that a 5 percent reduction in t_{ox} which increases the gate to source and drain capac-

itances, could match the measured center frequency with the simulation. The source of mismatch in the S_{22} is not completely clear; however a variation in the drain to body capacitance could also explain this discrepancy. This also causes output gain circles and load stability circles to approach each other as shown in Fig.5.11(a). A model of the transistor with t_{ox} and C_{db} variation was compared to the measurements results and shown in Fig.5.11(b) which qualitatively resembles the observed deviations in S_{11} and S_{22}. Valleys observed in both S_{11} and S_{22} around 17GHz are observed in all measurements and are attributed to the internal frequency response of the calibrated VNA. In Fig.**??** we plot the power measurements of the LNA, which shows a measured +4dBm 1-dB compression point which matches the predicted value based on the large signal model of the device. This translates into 23 % of power added efficiency, making it suitable as a pre-driver for a power amplifier or the output stage of a short-range transmitter. The simulated noise figure based on the model is 6 dB in the pass band of the amplifier and is shown by Fig.**??**. Measurements of noise figure were performed with a front-end with an internally matched amplifier and mixer. The amplifier is identical with the presented LNA except a small degeneration inductance was added for stability. The simulated noise performance of both LNAs is very similar. The gain of each stage was recorded separately and then the noise figure of the entire setup is performed. This measurement confirms that the noise figure is 6dB $\pm 0.5dB$.



(a)



(b)

Figure 5.11: Effect of C_{db} variation on the gain and stability circles. (b) Measurement and simulated values of S_{22} after adjustment of t_{ox} and C_{db}.

5.3 A 60GHz Cascode Amplifier with Interstage Matching

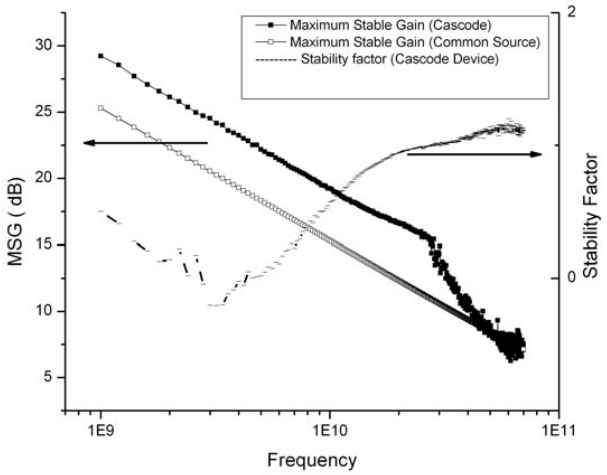


Figure 5.13: Measured MSG of a $20\mu m$ shared-junction cascode and a $20\mu m$ common-source device.

Cascode devices are widely used in amplifier design due to their higher gain and stability compared with common-source devices. These advantages come with no power dissipation penalty as the input and cascode devices share the same current. One potential problem with cascode devices in higher frequencies is the finite capacitance that exist at the junction of the two devices. This capacitance, can potentially reduce the gain and add to the noise (74). To overcome this problem, cascode devices are usually implemented in a shared-junction fashion as shown in chapter two. However, the shared-junction cascode still has a non-negligible parasitic capacitance at the interstage node. This capacitance together with the reverse feedback of the cascode device as will be discussed in chapter

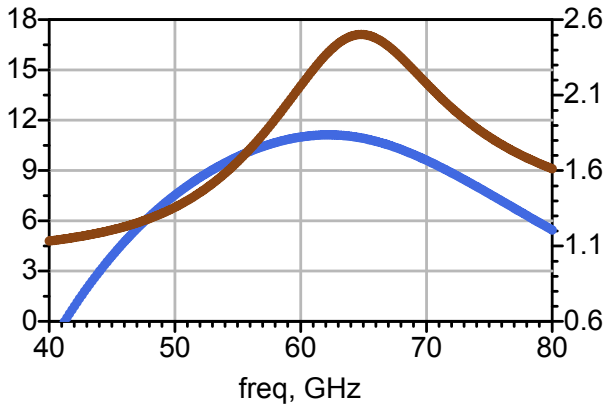


Figure 5.14: Simulated MSG(light) and k-factor(dark).

6, reduces the gain of the device at higher frequencies, making the cascode less beneficial at mm-wave. This can clearly be seen in Figure. 5.13. Indeed, at 60GHz, the MSG of the round-table common-source transistor is the same as the MAG of the shared-junction cascode device.

A different approach, followed in this design, is to create a cascode structure with two round-table transistors, and to place an interstage matching network between the common-source and the common-gate stage. This circuit combines the excellent mm-wave performance of the round-table layout with the DC current re-use of a cascode topology. This solution also achieves a lower DC power consumption compared to a two-stage cascade amplifier with two round-table common-source transistors. It also achieves more gain for the same DC power consumption when compared to a shared-junction cascode transistor.

Figure 5.14 shows the simulated MSG and k-factor of a cascode amplifier, composed of two 40 μm round-table devices, with an interstage matching. At 60GHz, a gain of 11.5dB is achieved with a DC power consumption of 6.7mW. As a comparison, a two-stage cascaded amplifier achieves a similar gain of 12dB,

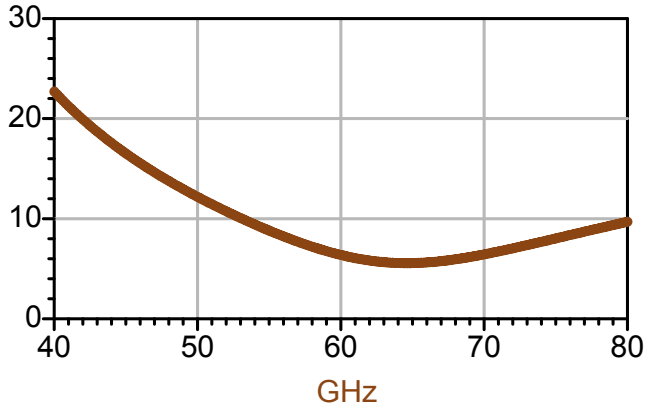


Figure 5.15: Simulated noise figure in dB.

but at a higher power consumption of 10.4mW (22). Also, this compound device compares favorable to a regular shared-junction cascode device that achieves a gain of 8.5dB for the same power consumption of 6.7mW. Clearly, the current re-use and interstage matching allows one to achieve more gain for a given DC power consumption. The simulated noise figure is 6.5dB as can be seen in figure 5.15.

The proposed schematic, using this compound cascode device, is shown in figure 5.16. Both coplanar transmission lines were used for input, output and interstage matching. Figure 5.18 shows the micrograph of the amplifier. The total area, including probepads is 750 μm by 450μm.

The S21 and S11 measurements are shown in figure 5.17. The amplifier achieves a peak gain of 7.5dB at 60GHz. The S_{11} achieves a minimum of -11dB at 62GHz, which is slightly off frequency. The output match S_{22} is -5dB at 60GHz.

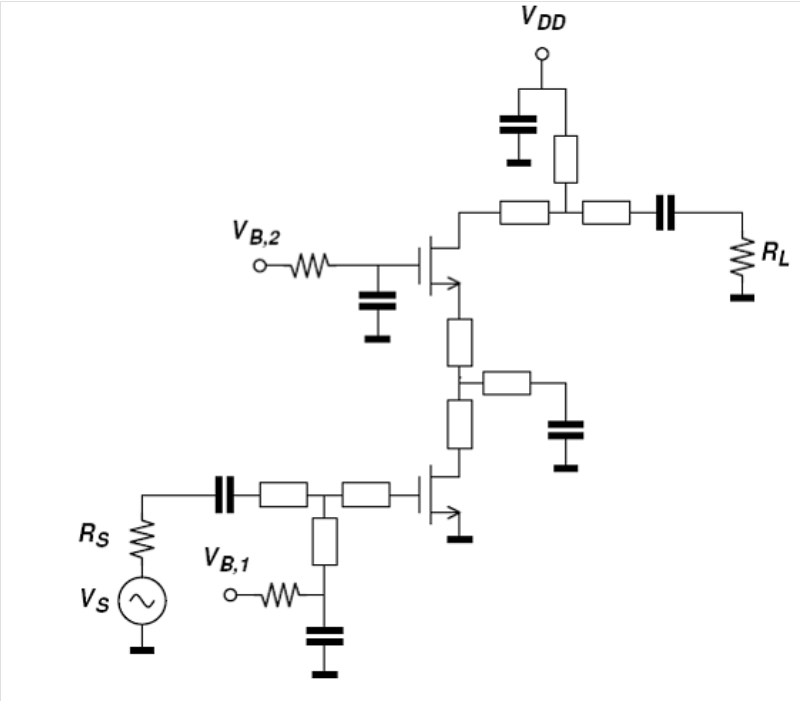


Figure 5.16: Schematic of the cascode amplifier.

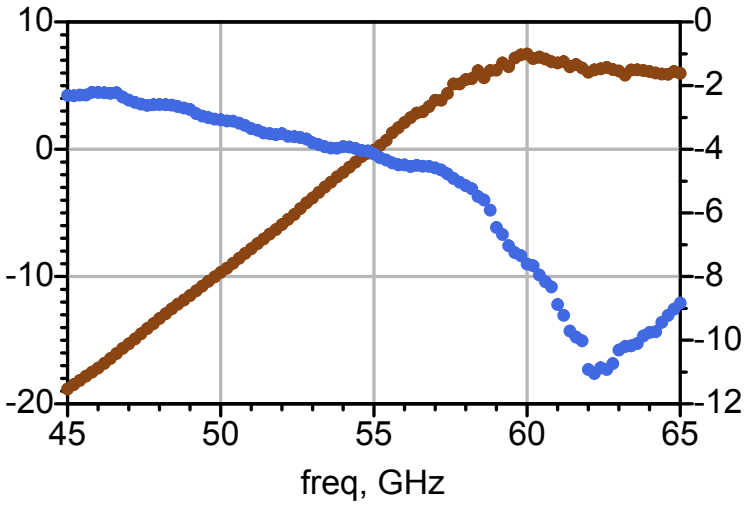


Figure 5.17: Measured S21 and S11 of the cascode amplifier.

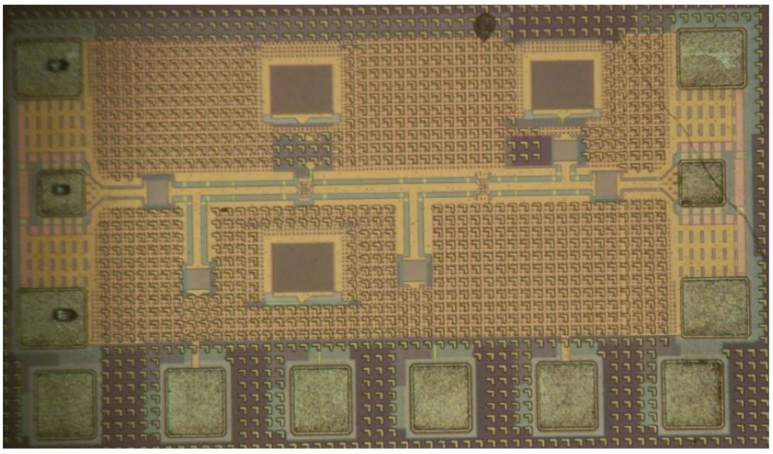


Figure 5.18: Micrograph of the cascode amplifier.

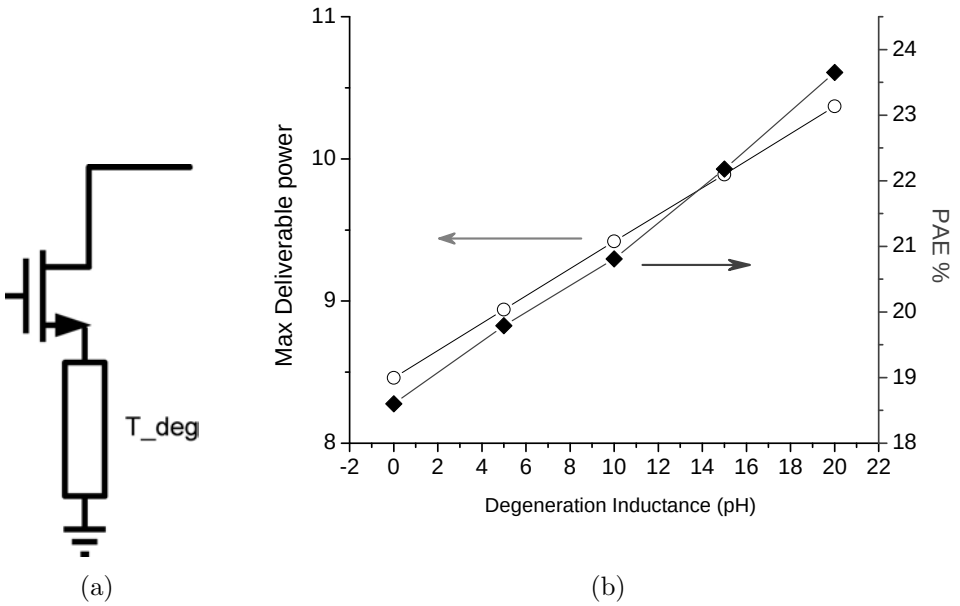
5.4 A 60GHz High Efficiency Power Amplifier

The two stage 60 GHz LNA that was presented in this chapter had a very good linearity with +4 dBm of output power. As the output matching network of that amplifier had been designed to maximize the gain, with a redesign of the output network, this amplifier can be turned into a two stage power amplfier. Apart from the size of the output stage device, the main difference between the power amplifier design and the LNA design is that in the former the output matching network is designed to maximize the power efficiency.

5.4.1 Output Stage Design

The most important part of PA design, is designing the output stage. This usually translates to designing the output matching network for the optimal drain efficiency. The value of load to maximize the efficiency is determined either by direct measurement or through simulation using an accurate large signal model of the device.

In the direct measurement method, the device is characterized as a function of terminating impedance as well as a function of drive level. This technique is called "Load Pull" and has the advantage that the actual behavior of the device is emulated. However, this method needs a large number of discrete measurement points of frequency and power level. The load values that provide the same output power level, form a series of contours on the Smith Chart and these contours are used as the basis of the design of the output matching network. If an accurate large signal model of the device is available, power contours can be generated through simulation and their values are very close to the actual load pull measurements (36).



(a) (b)

Figure 5.19: (a) Transmission line degenerated device used in the output stage. (b) The output power and efficiency of the device versus the inductance degeneration value.

To linearize the output stage device, a shorted transmission line degeneration

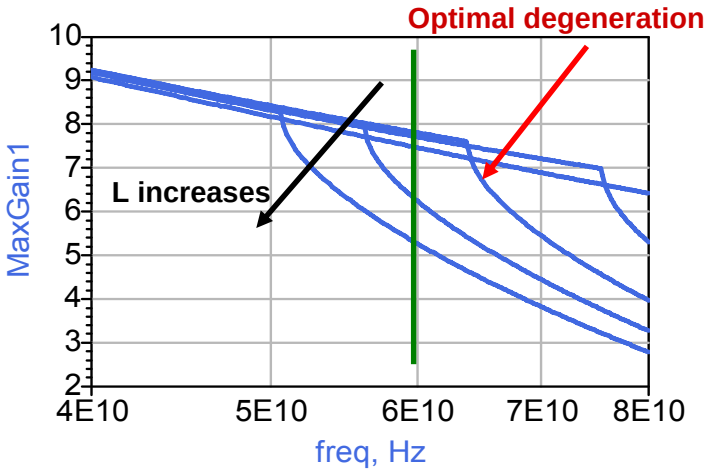


Figure 5.20: The change in the maximum stable gain versus the lenght of the degeneration transmission line.

was used at the source of the device as shown in Fig.5.19(a). This shorted transmission line acts as an inductive degeneration that helps to increase the maximum output power and the power added efficiency of the device. Fig. 5.19(b) shows the increase in the output power and efficiency of an $80\mu m$ device that was used in the output stage of the power amplifier. Increasing the value of the inductor improves the linearity of the device. However, as the device enters the unconditionally stable region, any increase in the inductance value has a negative effect on the maximum available gain. The optimum length of the transmission line is determined based on these two effects. The stabilizing and gain reducing effect is shown in Fig.5.20.

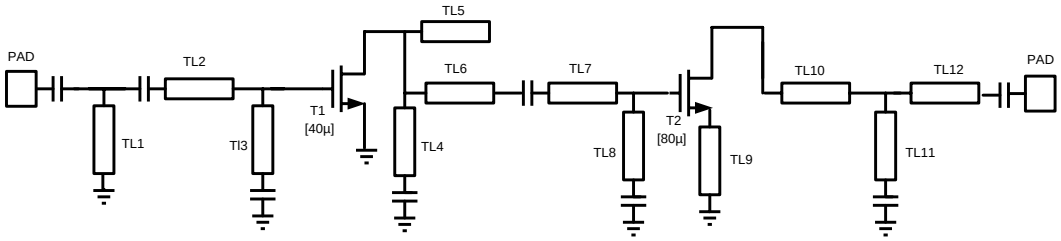


Figure 5.21: Schematic of two-stage power amplifier.

5.4.2 Amplifier Design

A two-stage amplifier was designed using "round table" layout transistors with different sizes for the input and output stages. The output device is 80μm and the input stage is 40μm. The choice of common-source output stage unlike most previous works, allows for more swing at the output and increases the output power and PAE as a result. The output stage is biased with 17.5 mA of current whereas the driver stage is biased with 6 mA of current. To realize a higher power, in this design the output matching network converts the load to a lower impedance of 10 Ω using a transmission line matching network.

Note that the output cannot drive the 50Ω load directly since this would result in very poor gain. A compromise is made between gain and output power when selecting the optimum load impedance. An interstage matching network is also utilized between the driver stage and the output stage. A two-stage low Q matching network is utilized so that the frequency response is determined mainly by the output network. Finally an input matching network is incorporated for a power match. The most important matching network is the output since this de-

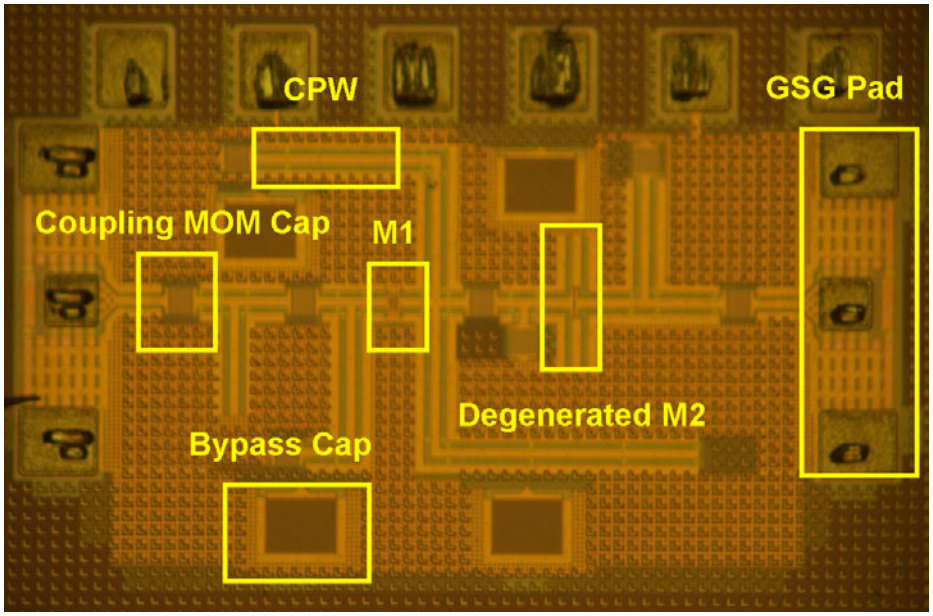


Figure 5.22: Die photo of power amplifier.

termines the output power capability. The input and interstage matching network should be broadband in order to make the design more robust. Extensive bypass capacitors are employed to decouple the bias and supply. The gate bias networks are de-Q'ed in order to suppress potential instability at lower frequencies.

5.5 Experimental Results

The prototype PA was fabricated in a digital CMOS process. The chip die photo is shown in Fig.5.22. Co-planar transmission lines, finger capacitors, and poly resistors are employed in the layout of the structure. The input and output pads parasitics are also absorbed into the design and measured results include pad losses. The measured S-parameters are shown in Fig.5.23. The input is match at 57 GHz, slightly below the targeted frequency. The gain also peaks at 9 dB at

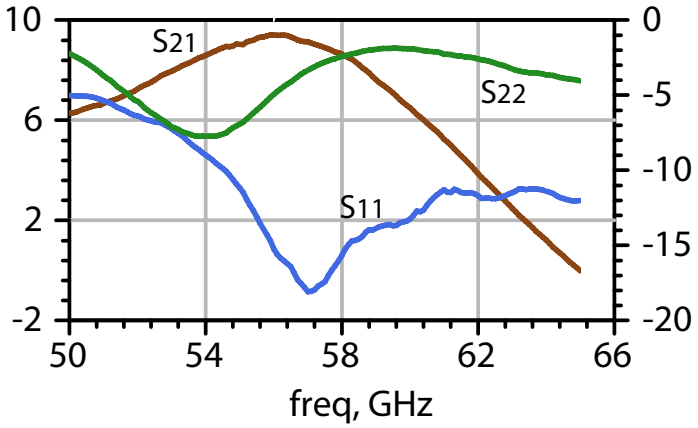


Figure 5.23: Measured S Parameters of power amplifier.

56 GHz and the gain is larger than 8 dB over a 6 GHz bandwidth. The output is matched for power (not gain), as evident in the figure. In simulations, the overall design is unconditionally stable, but the stability is marginal at 40 GHz in this particular design. In the actual measurements, potential instability is observed around this frequency but this did not cause any oscillations. The mismatch between measurement and simulations is mainly due to the improperly substrate connection of the output transistor that creates a feedback path between the input and output. The amplifier power measurements are shown in Fig.5.24. The input power is varied and the output power is measured using power meter. The linear behavior between input/output power indicates stable behavior. In the measurement setup the output was not directly observable around 40 GHz but the power sensor was used to verify that the circuit is stable, despite the potential instability observed in the small-signal measurements. The output power has a $P_{-1dB} = 6.5$ dBm, 1 dB lower than predicated by simulation. This corresponds

to a PAE of nearly 20%. Although not many mm-Wave CMOS PAs have been reported, a previously reported CMOS PAs demonstrated 6.7% power added efficiency but a higher output power of 10 mW (18).

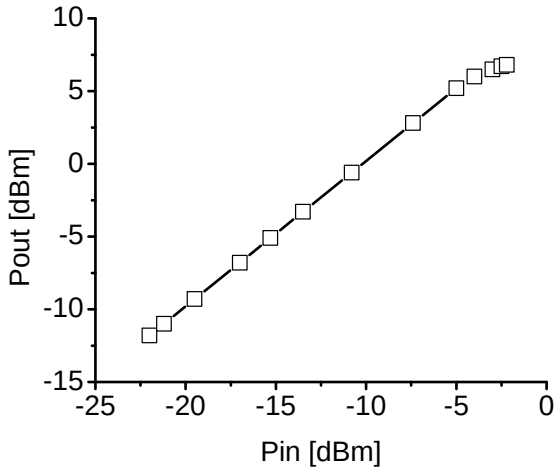


Figure 5.24: Measured output power versus input power.

5.6 Summary

Using the proposed high performance devices and their developed accurate models, few amplifiers were designed and implemented for the 60 GHz band as shown in this chapter. A low power, high linearity LNA were demonstrated with 10.5 mW of power dissipation, 6.5 dB of noise figure and $+4dBm$ of output P_{-1dB}. Based on the findings of Chapter 4, it was shown that unlike low GHz frequencies, there is no need for an inductive degenerated devices in LNA design when the frequency is close to f_t. It was shown that interstage matching between a common source and a common gate device can improve the gain and noise per-

formance of the overall circuit while keeping the power dissipation low through current sharing. A power amplifier was demonstrated that was designed for the maximum output power. The output power was improved by using an accurate large-signal model and linearizing the output device by inductively degenerating it at the source.

6

Unilateralization

The limited performance of transistors at high frequencies usually result in an increase in the number of gain stages which proportionally adds to the power dissipation of the system and also degradates the noise and linearity performance of the circuit. As a result, circuit techniques to improve the gain-power efficiency of devices at frequencies in the vicinity of the f_t of the device are highly valued.

On the other hand, it is of a theoretical value to construct a systematic method of boosting the gain of an N-port active networks and to determine the maximum possible stable gain achievable out of such a network. As the maximum stable gain is inversely proportional to the reverse feedback conductance of a network, minimizing this feedback path is a way to increase the potential gain of the network. In the extreme case, where this feedback is canceled out completely, the network becomes *unilateral*. At this point, the network is also very stable due to the lack of reverse feedback. As a result of these two effects, *unilateralization* techniques are highly valued for RF and mm-wave circuit design.

6.1 Theory Of Unilateralization

In 1953, when transistors were only 5 years old, people had started considering them for RF applications, limited in the VHF range for old-time devices. Mason started a goal to look for an invariant property of two port networks that could be used as a figure of merit for high frequency devices (47). The problem is defined as follow:

Consider a linear two port network as shown in Fig.6.1(a)[1]. An invariant metric has to be indifferent to any lossless transformation to the network. Any transformation of the device can be conceptualized as an embedding network, similar to Fig.6.1(b).The 4-port embedding network has to be linear, lossless and reciprocal.

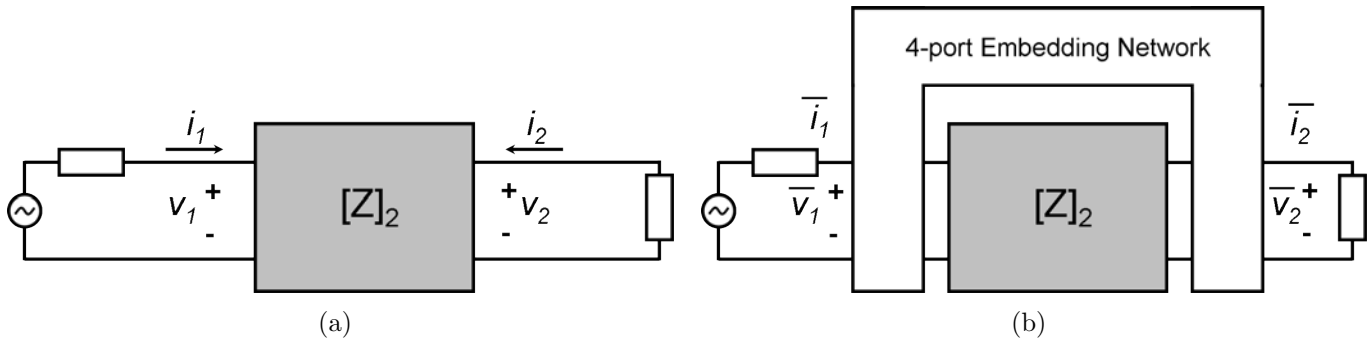


Figure 6.1: The linear two port device (a)used as an amplifier (b) embedded in a 4-port lossless reciprocal network

Mason showed that all conceivable transformations that satisfy the constraints in the four port network, could be realized from just three basic transformations. These transformations are Reactance Padding, Real Transformation and Inversion. In terms of impedance matrix they could be represented with these

[1]Both being two port and linear are essential constraints for the problem

equations:

1. Reactance padding:

$$\begin{bmatrix} \overline{Z_{11}} & \overline{Z_{12}} \\ \overline{Z_{21}} & \overline{Z_{22}} \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} + j \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

2. Real transformation

$$\begin{bmatrix} \overline{Z_{11}} & \overline{Z_{12}} \\ \overline{Z_{21}} & \overline{Z_{22}} \end{bmatrix} = \begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{bmatrix} \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{bmatrix}$$

3. Inversion

$$\begin{bmatrix} \overline{Z_{11}} & \overline{Z_{12}} \\ \overline{Z_{21}} & \overline{Z_{22}} \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}^{-1}$$

These three transformations could be realized with several different circuits. Now the problem reduces to finding an index in terms of the impedance matrix that remain intacts to these three transformation. The reactance padding keeps $[Z - Z^t]$ and $[Z + Z^*]$ unchanged while the real transformation reduces this to the determinant of $[Z - Z^t][Z + Z^*]^{-1}$. In the end, inversion transformation restricts only the magnitude of this matrix to be invariant. The resulting invariant term from all these three basic transformations is called U and can be writen as:

$$U = \frac{|det(Z - Z^t)|}{det(Z + Z^*)}$$

This could be written in more familiar forms as:

$$U = \frac{|Z_{12} - Z_{21}|^2}{4\{\Re(Z_{11})\Re(Z_{22}) - \Re(Z_{12})\Re(Z_{21})\}} \tag{6.1}$$

$$U = \frac{|Y_{12} - Y_{21}|^2}{4\{\Re(Y_{11})\Re(Y_{22}) - \Re(Y_{12})\Re(Y_{21})\}} \tag{6.2}$$

At this point, the desired invariant metric is found and this in fact is the major result of Mason's paper.

6.1.1 Mason Gain As A Gain Maximum

Other than being an invariant parameter, U implies a maximum gain under certain condition. If the original two port network is unilateralized using a 4-port linear lossless reciprocal network as shown earlier, then the value of U is equal to the value of maximum stable gain of that network [1].

Writing the maximum stable gain equation in a slightly different form we have

$$G_{max} = \frac{|Y_{21}|}{|Y_{12}|} \frac{1}{k + \sqrt{k^2 - 1}} \tag{6.3}$$

Now

$$G^u_{max} = \lim_{Y_{12} \to} G_{max} = \frac{|Y_{21}|^2}{4\Re(Y_{11})\Re(Y_{22})} \tag{6.4}$$

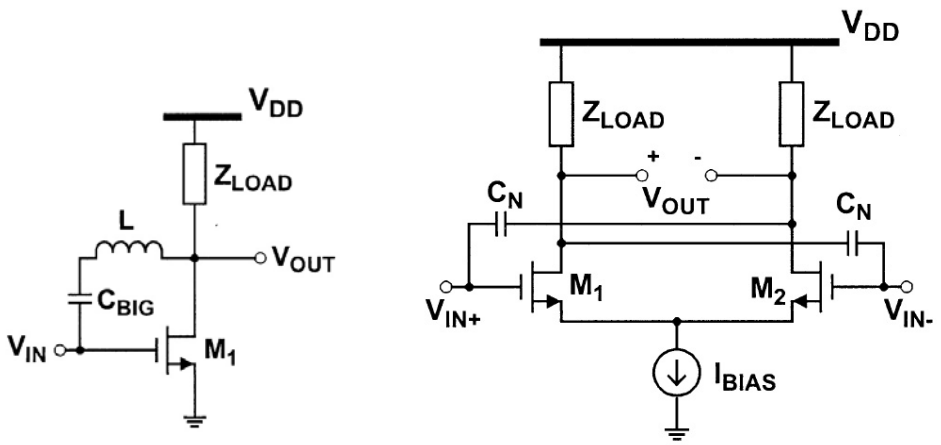
Setting Y_{12} to zero in equation 6.2 gives the same result proving that G^u_{max} is equal to U. This suggests that the stable gain of a unilateral network is bounded by mason gain.

It is however crucial to remember that U is not the maximum gain unless

[1] This is formally proven in the Appendix.B.

the device is first unilateralized using an embedding network. If the device is not unilateral, the maximum stable gain could be significantly higher than U as will be demonstrated in the following sections. In fact, as we pass the unilateral frequency, on one hand, K decreases since the network becomes less stable and makes the expression in the parenthesis in equation 6.3 increase. This in conjunction with the ratio of $|Y_{21}|$ and $|Y_{12}|$ determines the global maximum of the stable gain over the frequency. However, since the gain and stability increase coincide at the point where the device is unilateral, several circuit techniques to achieve unilateralization have been developed as will be discussed in the next section.

6.2 2-port Unilateralization Techniques



(a) (b)

Figure 6.2: (a)Neutralization using a resonating inductor. (b) Cross coupled capacitor neutralization.

Investigating unilateralization techniques has a long history. Cheng (67) in his classic paper in 1953, presented a general scheme and some circuit implemen-

tations for unilateralization of 2-port networks all require transformers. A more commonly used technique is neutralization which can be implemented more easily and often used as an alternative for unilateraliztion in moderate frequencies. Neutralizatin may be defined as the process of balancing out an undesirable effect (67).The technique is mostly investigated for common-source/common-emmiter devices. For this architecture, the dominant reverse feedback element is the C_μ of the device and neutralization goal is to cancel out the effect of this capacitor. This could be achieved by simply resonating out the capacitor, using an inductor as depicted in 6.2(a). A large series capacitor is needed to dc couple the gate and the drain. More over, this technique is narrow band due to its resonance nature.

A wide band neutralization could be achieved as shown in 6.2(b) (51). The idea is that C_N is sized in such a way that it injects a negative current equal to the magnitude of the feedback current passing through C_{gd} so that the total current returns to the input becomes zero. As the drain voltages of a differential pair have opposite phases, the negative current could be achieved with this architecture. For mm-wave application however, the parasitic inductance of the neutralization capacitance becomes significant and could limit the applicability of this method. It is also important to remember that neutralization is equivalent to unilateralization, only if the reverse feedback is pure imaginary. This is not the case for architectures other than common-source/base. The assumption of imaginary reverse feedback for common-source/base also breaks in frequencies close to f_t as other effects become important. A general condition for unilateralization in such frequencies could be achieved by looking at the desired circuit as a multi-port network.

6.3 N-Port Unilaterization

For an n-port network, the unilateralization technique translates into finding proper complex terminations for $n-2$ of the ports to make the remaining 2-port unilateral.

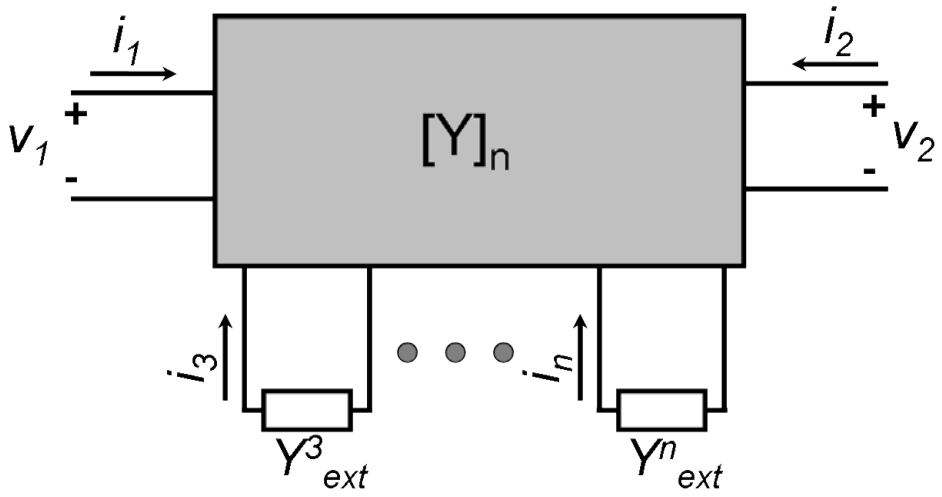


Figure 6.3: N port network with $N-2$ external complex termination.

Consider an n-port network as shown in 6.3 One can readily write the n dimensional admittance parameters relating input voltages and currents with designated signs shown on the picture:

$$I_n = Y_{1-n}^{1-n} V_n \tag{6.5}$$

This equation could be decomposed in this fashion:

$$I_{1-2} = Y_{1-2}^{1-2} V_{1-2} + Y_{1-2}^{3-n} V_{3-n} \tag{6.6}$$

$$I_{3-n} = Y_{3-n}^{3-n} V_{3-n} + Y_{3-n}^{1-2} V_{1-2} \tag{6.7}$$

For which $Y_{k_1-k_2}^{m_1-m_2}$ represents a $(k_2 - k_1 + 1) \times (m_2 - m_1 + 1)$matrix with y_{ij} elements for which i and j change from k_1 to k_2 and m_1 to m_2 respectively.

Now assume that ports 3 to n are terminated by a series of complex loads with admittances equal to $Y_{ext}^{i}(i : 3...n)$. These loads introduce another set of equations:

$$I_{3-n} = Y_{ext} V_{3-n} + Y_{3-n}^{1-2} V_{1-2} \tag{6.8}$$

In which

$$Y_{ext} = \begin{bmatrix} Y_{ext}^{3} & 0 & \cdots & 0 \\ 0 & Y_{ext}^{4} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & Y_{ext}^{n} \end{bmatrix} \tag{6.9}$$

Combining equations **??**, 6.7 and 6.8, the parameters for the resulted twoport network will be:

$$Y_{1-2}^{\bar{1}-2} = Y_{1-2}^{1-2} - Y_{1-2}^{3-n}(Y_{ext} + Y_{3-n}^{3-n})^{-1} Y_{3-n}^{1-2} \tag{6.10}$$

Using this equation, the set of Y_{ext}^{i}s to make $\bar{Y_{12}} = 0$ could be found to realize unilateralization.

To check the applicability of this technique, two other conditions must be tested. First, to take advantage of the benefits of this method, it is desirable to realize this gain boost using passive components to form the Y_{ext}, requiring its real part to be positive at all frequencies. The unilateral network needs also to

be stable. Since the $\overline{Y_{12}} = 0$, this condition translates to $Re(\overline{Y_{11}})$ and $Re(\overline{Y_{22}})$ be positive, putting a condition over the original Y parameters set. These will become more clear in the 3-port discussion.

6.3.1 Single Transistor Unilateralization

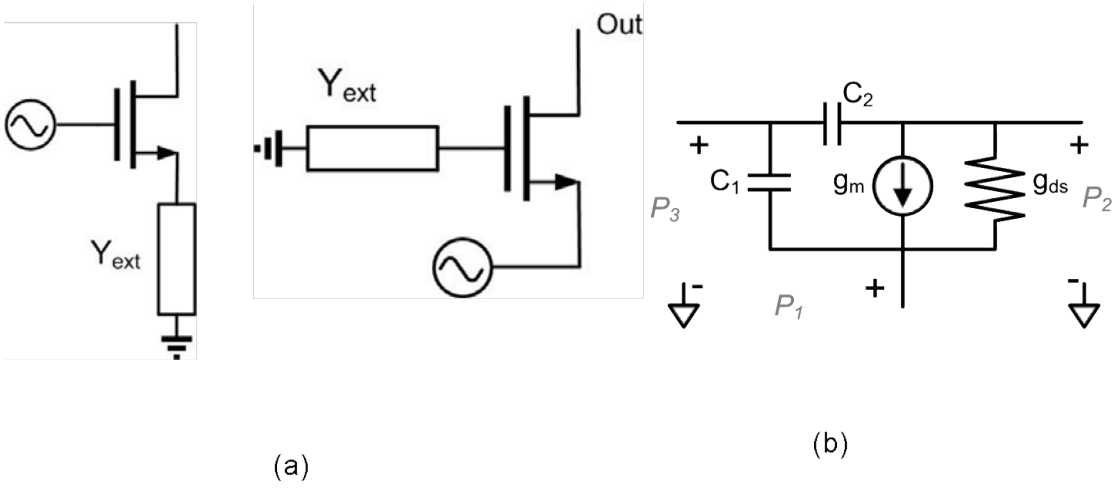


Figure 6.4: (a) Unilateral common-source and common gate. (b) Simplified small signal model for 3-port CMOS.

The first natural candidate for testing the theory is a single transistor. Ignoring the body terminal for simplification, a transistor is a three terminal device which can be terminated with a proper external impedance to increase the gain as described earlier. In order to see the possibility of this method, we test it for the two widely used gain stages, the common-source and common-gate devices as shown in 6.4-a . For the hybrid-pi model of transistor as depicted on Figure

6.4-b, the admittance metric is:

$$Y_3 = \begin{bmatrix} g_m + g_{ds} + j\omega C_1 & -g_{ds} & -g_m - j\omega C_1 \\ -(g_m + g_{ds}) & g_{ds} + j\omega C_2 & g_m - j\omega C_2 \\ -j\omega C_1 & -j\omega C_2 & j\omega(C_1 + C_2) \end{bmatrix} \tag{6.11}$$

Using these parameters in **??**, the required Y_{ext} could be determined for these two structures. For the Common-Source structure, it can be readily seen that the unilateralization is not achievable using passive devices since the real part of Y_{ext} is negative over all frequencies:

$$Y_{ext} = -g_{ds}(1 + \frac{C_1}{C_2}) - g_m - j\omega C_1 \tag{6.12}$$

The common-gate is a more interesting. The required Y_{ext} for this structure is as follow:

$$\begin{aligned} \Re(Y_{ext}) &= \frac{\omega^2 C_1 C_2}{g_{ds}} \\ \Im(Y_{ext}) &= -\omega(C_1 + C_2) - \frac{g_m C_2}{g_{ds}} \end{aligned} \tag{6.13}$$

Implying that unilateralization is achievable with a passive inductive impedance. Using a series RL circuit to realize the load, the required inductance and resistance is

$$L_{ext} = \frac{g_{ds}}{\omega^2((C_1+C_2)g_{ds}+g_mC_2+\frac{(C_1C_2\omega)^2}{(C_1+C_2)g_{ds}+g_mC_2})} \quad (6.15)$$

$$R_{ext} = \frac{L_{ext}C_1C_2\omega^2}{(C_1+C_2)g_{ds}+g_mC_2} \quad (6.16)$$

Since an inductor at the gate is a classical method to build an oscillator, it is natural to question the stability of this structure. For the unilaterizied common-gate structure, the real parts of $\overline{Y_{11}}$ and $\overline{Y_{22}}$ are as follow

$$\Re(\overline{Y_{11}}) = g_m + g_{ds}(1+\frac{C_1}{C_2}) \geq 0 \quad (6.17)$$

$$\Re(\overline{Y_{22}}) = \frac{\omega^2(C_1-C_2)C_1}{g_m^2+\omega^2C_1^2}g_{ds} \geq 0 \quad (6.18)$$

The first condition is always met. To satisfy the stability at the output, it is necessary that C_1 be larger than C_2 with a proper margin. With a reasonable control over external parasitics, this is almost always the case for CMOS processes hence ensuring the stability of the structure. In fact, the condition for oscillation is the opposite of what we are looking for, requiring the Y_{12} to be equal to infinity. This translates to $Y_{ext}+Y_{33}$ equaling zero. Considering an inductor at the gate, this condition yields to the familiar frequency of oscillation:

$$\omega_{osc} = \frac{1}{\sqrt{(c_1+c_2)L_{ext}}} \quad (6.19)$$

Using 6.19 and 6.16 it can be shown that the difference between the two

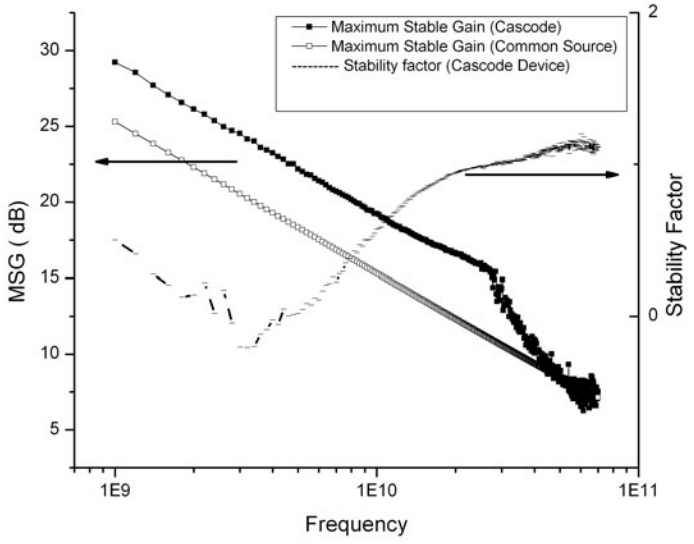


Figure 6.5: Comparison of maximum stable gain (MSG) between a common source and cascode structure using similar current.

frequencies can be shown as follow:

$$\omega_{osc} - \omega_{uni} = \frac{(\omega_{osc}\omega_{uni})^2}{\omega_{osc} + \omega_{uni}}(L_{ext}c_2\beta + L_{ext}\frac{(c_1c_2\omega_{uni})^2}{(c_1 + c_2) + c_2\beta}) \tag{6.20}$$

Since this difference is always positive, it implies that the unilateralization frequency always happens first and the difference is a function of the β as well as internal capacitances of the device.

6.3.2 Cascode Device Unilateralization

Cascode devices are used extensively due to their larger gain and the input-output isolation.The gain benefit that normally associated with cascode structures is not necessarily available at mm-wave frequencies as the unilateral assumption breaks. Fig. 6.5 compares the measured maximum available gain (MSG) between

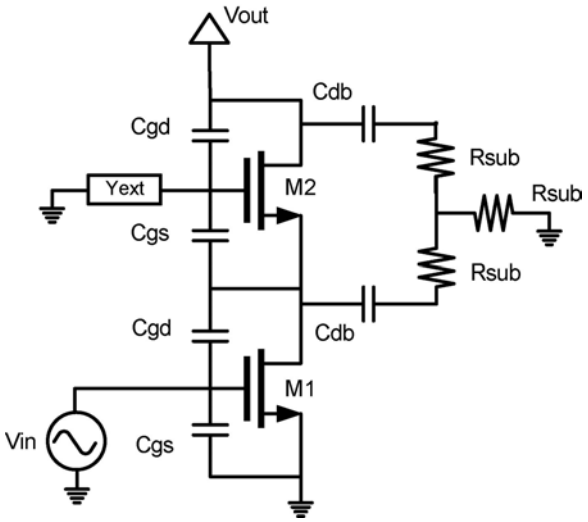


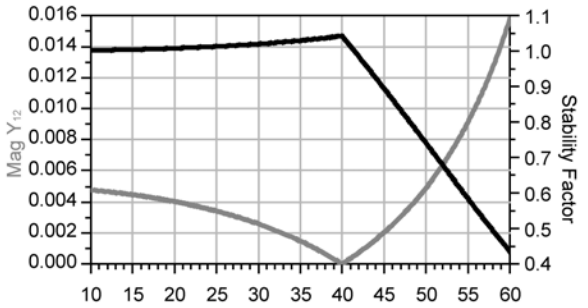
Figure 6.6: Small-signal model of a cascode structure used in hand calculations.

a normal common source device and a cascode device. As evident in the figure, the maximum available gain of the cascode deivce is considerably higher at high frequencies, but gets close to the value for the common source device at mm-wave frequencies. At 60 GHz for example, both devices show similar MSG of around 7.5 dB.

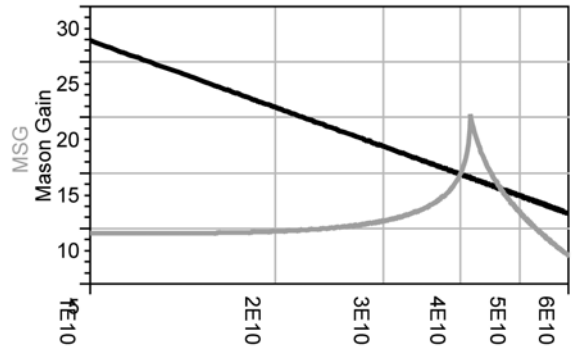
The described technique could essentially be used for cascode structures by placing the required network at the second gate of the device. To calculate the required impedance, the input device and the shared substrate network needs to be counted in the calculations. Using the small-signal model of the device as shown in Fig.6.6, the required admittance at the second gate could be calculated as follows :

$$\Re(Y_{ext}) = \frac{\omega^2 C_{gd}\left(G_{sub}^2((3C_{gs}+C_{db})g_{m2}-3C_{gs}G_{sub})+C_{db}C_{gs}\omega^2(2C_{db}g_{m2}+C_{gs}G_{sub}-2C_{db}g_{m2})\right)}{\left(G_{sub}^2+\omega^2 C_{db}^2\right)\left(G_{sub}^2+\omega^2 C_{gs}^2\right)}$$

$$\frac{\Im(Y_{ext})}{\omega} = 3C_{gd}-C_{gs}+\frac{C_{gd}G_{sub}(C_{db}g_{m2}-C_{gs}G_{sub})}{\left(C_{db}^3\omega^2+C_{db}G_{sub}^2-C_{gd}G_{sub}^2-C_{db}^2C_{gd}\omega^2\right)}-\frac{G_{sub}C_{gd}\left(2C_{db}-3C_{gd}\right)(G_{sub}-g_{m2})}{(C_{db}-C_{gd})(G_{sub}^2+C_{gd}^2\omega^2)}$$



(a)



(b)

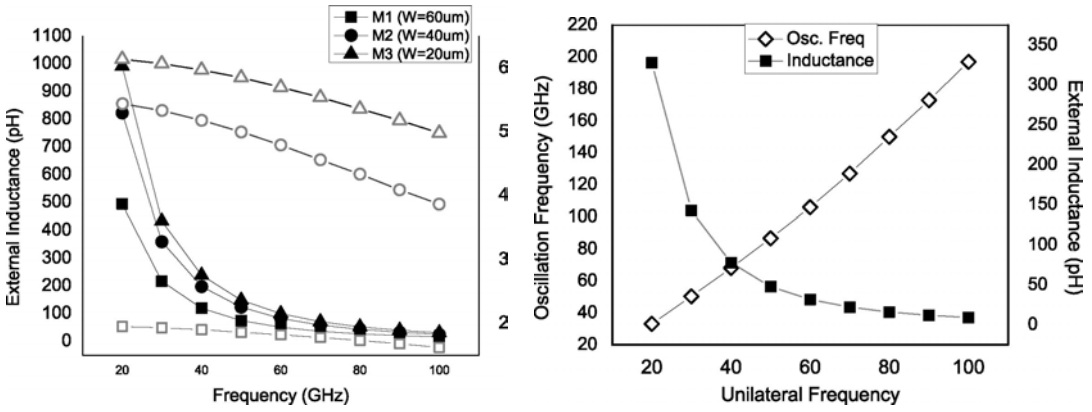
Figure 6.7: (a) Simulated magnitude of Y_{12} and the stability factor (k) and (b) device maximum stable gain and maximum unilateral gain (Mason gain).

6.4 Simulated Results

The value of the required L_{ext} and R_{ext} were calculated for NMOS devices of various size in the 90nm Technology. These devices were modeled up to 100GHz to extract the required device parameters based on the method described in chapter 2.

For a sample $40\mu m$, Fig. 6.8(a) shows the magnitude of Y_{12} and the stability factor. The gate network is set to make Y_{12} zero at 40GHz. The stability factor is above one confirming the stability of the device as was predicted in the previous

section. Fig. 6.7-b compares the maximum stable gain as opposed to the maximum unilateral(Mason) gain. The significant difference between the two types of the gain is apparent in frequencies far below the unilateralization frequency. At this frequency, the two gains become equal as suggested by the definition of the Mason gain. Interestingly, the stable gain goes beyond this value. As un-intuitive as it might seem at first, it actually does not contradict with the Mason theory. The theory in fact suggests that the U is the maximum unilateral gain and is invariant to the type of the external network that has been used to realize the unilateralization (48). The gain boost resulted from this technique could be up to 7 dB at the unilateral point and up to 13 dB at the peak gain depending on the frequency of operation and the size of the device.



(a) (b)

Figure 6.8: (a)Required L_{ext}(filled) and R_{ext} (empty) values for unilateralzation for NMOS devices of W=20μ,40μ and 60μ, L=90nm (b) Simulated oscillation frequency vs. unilateralization frequency for W=40μ.

Fig. 6.8(a) shows the corresponding L_{ext} and R_{ext} for three sizes of an NMOS device with the same multi-finger layout structure. The needed inductance is very large at lower frequencies, but quickly drops to reasonable integrable values as

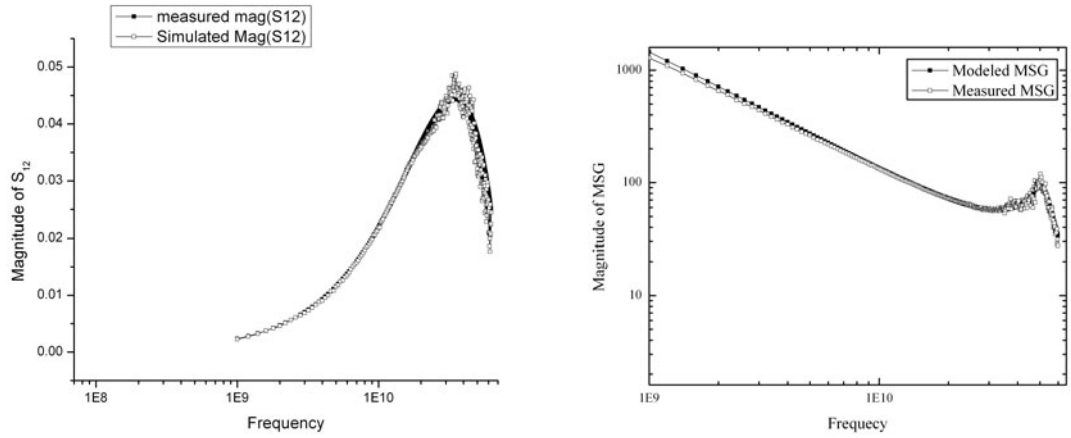
the frequency increases to mm-wave region. The required inductor also decreases for larger device sizes due to larger internal capacitances. These two suggest that the method is mostly suitable for relatively large devices and frequencies beyond 5GHz. Resistance value on the other hand, is quite steady versus frequency and is on the order of few ohms for the device sizes in the range of interest. Fig. 6.8(b) illustrates how the oscillation frequency changes versus unilateralization frequency for different values of external inductances and shows a significant distance between these two frequencies as predicted by equation 6.20.

Practically, a bypass capacitance is usually used at the gate of common-gate devices to minimize the effects of biasing lines. With a value of few pFs, the self resonance frequency of such capacitors usually happen somewhere in the mm-wave region implying the existence of a series inductance of few pHs with the capacitor. An external inductance in a form of a short transmission line could be used to set a net inductive impedance based on equation 6.16. The line could also be designed to set the required resistance.

6.5 Implementation And Experimental Results

The unilateralization procedure was implemented on a sample 80μm/90nm cascode device using an integrated 2 pF external capacitor in series with a total of 50 pH of external inductance. The 2 pF finger "MOM" capacitor was modeled and its internal inductance was used in the modeling process. The cascode device was carefully modeled based on the equivalent circuit that was given in chapter 2, considering the effect of the shared substrate node.

Fig. 6.9(a) and 6.9(b)the measured and modeled S_{12} of the device as well

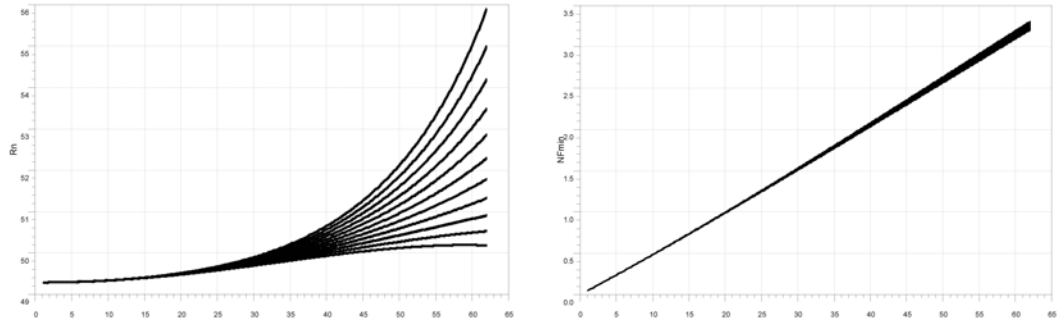


(a) (b)

Figure 6.9: (a) Device S_{12} (b) Device maximum stable gain

as the maximum available gain up to 65 GHz. The fall-off of the magnitude of the reverse reflection parameter, S_{12} after 20 GHz corresponds to the increase in the maximum available gain in the same frequency range. Both these effects are well in agreement with the modeling results. The peak gain at 50 GHz is close to 20 dB which is drastically larger than the 8 dB gain for the case of normal cascode, as shown previously in Fig. 6.5.

The noise performance of the device was simulated using Pospieszalski noise model with the aid of the proposed small-signal circuit (57). The equivalent noise parameter γ was set to 1.3 as was discussed in chapter 4. The effect of unilateralization technique was investigated on the minimum noise figure and equivalent noise resistance (R_n) by changing the series inductance at the second gate LC tank. The effect is minor as shown in Fig. 6.10(a). While NF_{min} decreases slightly when the device goes into unilateral region, R_n increases and no clear overall benefit or disadvantage is associated with technique in terms of noise



(a) (b)

Figure 6.10: (a) Simulated device noise resistance parameter R_n (b) Minimum achievable noise figure NF_{min}.

performance. However, the technique enables the designer to trade-off gain for noise optimization which results in an overall more power efficient amplifier.

6.6 Summary

This chapter explored the possibility of increasing the power gain of the device beyond its two port MSG. A general theory of unilateralization for N-port networks were developed and the theory was successfully tested for the case of single transistor in the common-source structure. It was shown that the stable gain of the device can exceed the value of mason gain and U is not the maximum achievable gain of a two port as sometimes is mistakenly mentioned. The unilateralization theory was used for a cascode structure and the gain peaking was observed and successfully modeled.

7

CMOS Circuits Design Beyond f_t

In Chapter 3 we saw that the actual figure of merit for transistors is the f_{max} of the device rather than the f_t. Also it was shown that this figure of merit, together with the available gain of the device at any particular frequency can be significantly improved using optimization methods as described in chapter 3. The significantly larger f_{max} compared to f_t should make it possible to design circuits at or even beyond the f_t of the technology. In 90nm technology, the f_t is equal to 100GHz and the actual f_{max} of an optimized round-table device is close to 200GHz[1]. However the actual f_{max} is smaller than this as U is not a linear curve due to higher order poles of the device. Even considering all the losses in passive devices and matching networks, this essentially means that we should be able to build circuits beyond 100 GHz. For our experiment, measurement facilities however were only available up to 110 GHz. As a result, experimental circuits were designed at a frequency between the f_t of the device and the measurment capability limit. Two circuits at 104GHz were designed and implemented. One

[1]The extrapolated f_{max} is 300 GHz as shown in chapter 3

non-linear circuit, a high performance first harmonic oscillator and another, a linear circuit, a three stage amplifier. These two blocks are described in this chapter.

7.1 mm-wave Oscillator Operating at The Device f_t

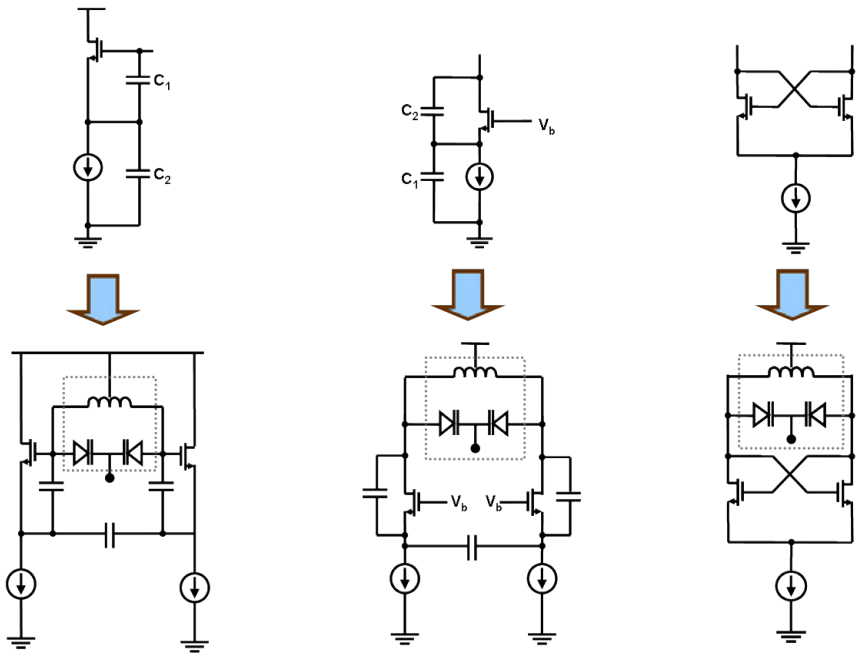


Figure 7.1: Negative resistance block of the mm-wave oscillator.

A voltage-controlled oscillator is a key component in mm-wave transceivers. Designing oscillators with high output power, low power dissipation and reasonable phase noise at mm-wave frequencies is a challenging task. CMOS oscillators at frequencies beyond 100GHz have been previously reported (82; 83; 84; 85). However, most published oscillators use a push-push topology or employ higher

harmonics and subsequently offer a low output power. In order to achieve high output power with reasonable power dissipation, fundamental mode oscillators are desirable. This however requires that the f_{max} of the device is substantially higher than the frequency of operation and the device has sufficient available power gain at the operating frequency.

An oscillator could be viewed as a a negative resistance stage, connected to a proper resonator at a particular frequency. At mm-wave, the most common negative resistance stages are either based on a single transistor Colpitts structure or differential negative g_m stage. Collpitts-based structures also could be combined to form a differential oscillator as shown in Fig.7.1. As the resulted negative resistance comes with an associated capacitive part, an shorted transmission line ususally serves as an inductor to form the resonator and set the frequency of oscillation.

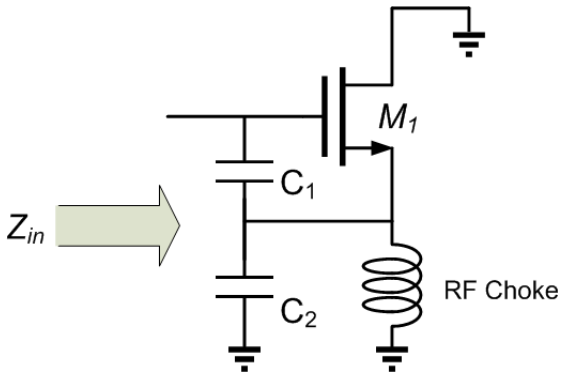


Figure 7.2: Simplified schematic of a collpitts oscillator.

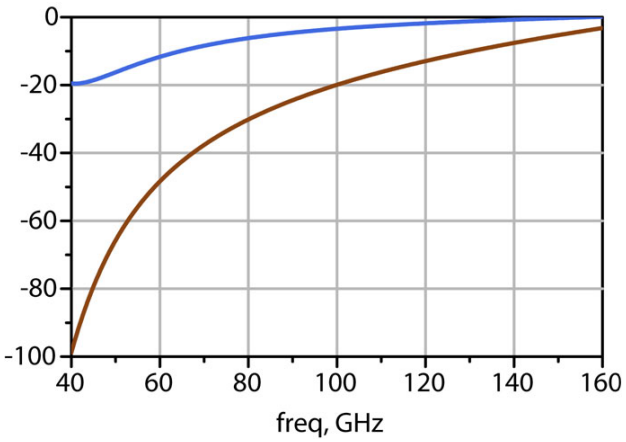
The optimization of the round-table structure device with its high maximum frequency of operation allows, in principle, for a simple single transistor oscillator. A member of the Colpitts family depicted in fig.7.2, was selected as the archi-

tecture of choice. This structure is compact, efficient and can form the core of a differential VCO. The capacitive feedback of the structure creates a negative real part and a capacitive imaginary impedance at the gate terminal. This imaginary resonates with an inductive impedance whose value is determined based on the desired oscillation frequency. Assuming a simple first order transistor model, the input impedance of this structure is given by

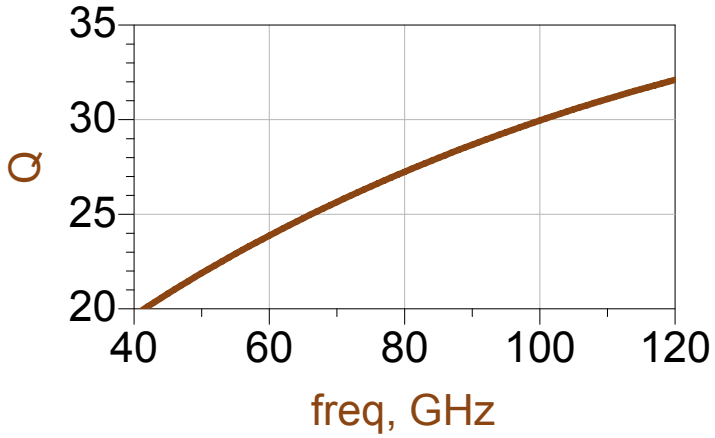
$$Z_{in} = -\frac{g_m}{C_1 C_2 \omega^2} + \frac{1}{j\omega C_1} + \frac{1}{j\omega C_2} \tag{7.1}$$

In this equation, g_m is the effective transconductance of the device and C_1 and C_2 are total capacitances including the internal device capacitors. While this equation is useful for design, in parctice due to the complexity of models at mm-wave frequencies beyond the device f_t, the actual real and imaginary parts need to be derived through simulation. Fig.7.3(a) shows the real and imaginary parts of the structure using physical models.

As clear from the figure, the circuit has negative resistance up to 160 GHz, and as the graph suggests, at 100 GHz an inductive load with a Q of at least 7.5 is needed to ensure oscillation. At mm-wave frequencies, a shorted transmission line could be used to serve as the inductive load instead of loop or spiral inductors that are more commonly employed in RF designs. CPW transmission lines have reasonably high quality factor at these frequencies, providing a Q of 30 at 100 GHz in this technology as Fig.7.3(b) suggests. The lenght of the transmission line could be used to set the frequency of oscillation. The bypass capacitor, C_8, has a capacitance of 1pF and a self-resonance frequency of 50 GHz, necessitating the inclusion of its series inductance into the resonant tank.



(a)



(b)

Figure 7.3: (a) The simulated input impedance of the negative resistance block using physical models for all the components. (b) Shorted transmission line quality factor versus frequency.

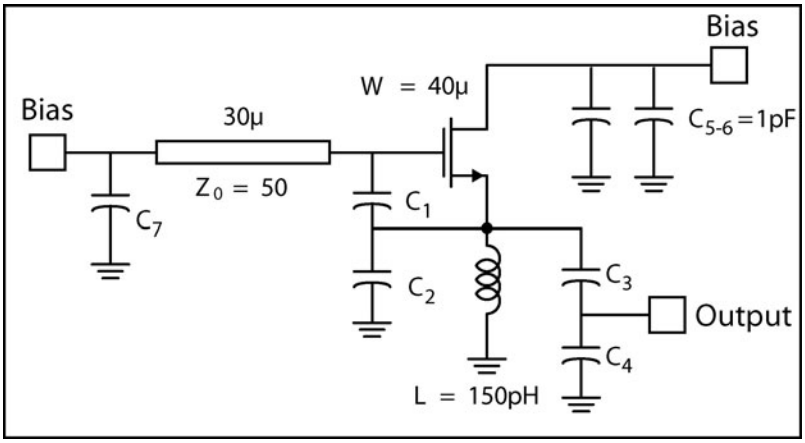


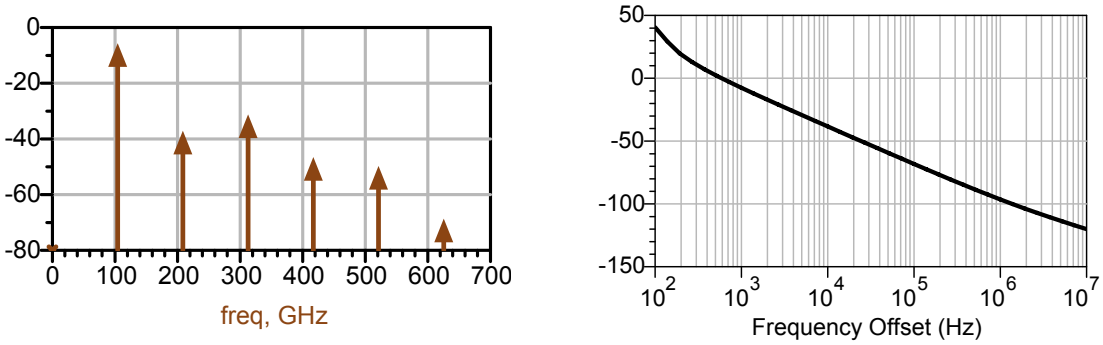
Figure 7.4: The complete schematic of the oscillator.

To control the oscillation frequency, a voltage dependent capacitor in the form of a varactor is ususally used as shown in figure 7.1. MOS varactor introduces a considerable loss to the circuit and thus limit the frequency operation of the oscillator. For applications that require a limited pulling range (up to 5%) the frequeny pulling could be done using the internal non-linear capacitance of the device. If the proper choice of C_2, the C_{gs} of the dervice could be used as C_1. The value of this capacitor is voltage dependent and varies between $C_{ox}/2$ and $2/3C_{ox}$ (81):

Figure.7.4 shows the complete schematic design of the Colpitts oscillator as implemented. Capacitors C_1 and C_2 form the feedback network, however capacitor C_1 is replaced by the internal C_{gs} of the device which is on the order of 50 fF. Removing C_1 reduces circuit size and excess phase losses due to external capacitance interconnections. Due to the non-linear nature of the gate to source capacitance, the the gate bias voltage can be employed as a rudimentary means of controlling the VCO frequency for ranges up to 5%. Needless to say that in a

real applications a better way is needed to control the VCO frequency, as varactor Q's are prohibitively low at 100 GHz.

In a low frequency Colpitts oscillator, it is typical to employ the drain node in order to derive a buffered version of the VCO signal. Surprisingly, in this design we found that the drain node to be excessively sensitive, and any additional loss shunted to ground at the drain quenched the oscillation. For this reason we chose not to load this node with the measurement load impedance. The device source node is employed as the output of the oscillator. A capacitive matching network, C_3 and C_4, transforms the real part of the device output impedance, $\frac{1}{g_m}$ from 25 to 50 Ω. Capacitors C_5 and C_6 are bypass capacitors with sizes close to 1pFA 150pH inductor is designed and modeled in a 3D EM simulator and used as a bias choke at the source. Output and bias pads are also modeled and embedded into the matching network and tuning circuitry.



(a) (b)

Figure 7.5: (a) The simulated output spectrum for the 104GHz oscillator. (b) The simulated phase noise.

Figure 7.5(a) shows the simulated output spectrum of the oscillator. This

simulation is based on an accurate large signal model of the round-table device. The simulation shows a fundamental frequency at 104.2 GHz with -7.3dBm output power. Interestingly the 3rd harmonic at 312 GHz has close to -30dBm of output power and is even stronger than the second harmonic. Implementing the same circuit with a higher performance device (a 45nm device for example) and using the third harmonic of oscillation can result in a silicon-based THz source. The simulated phase noise, as shown in Fig.7.5(b), is -96dBc at 1MHz offset.

7.1.1 Implementation and experimental results

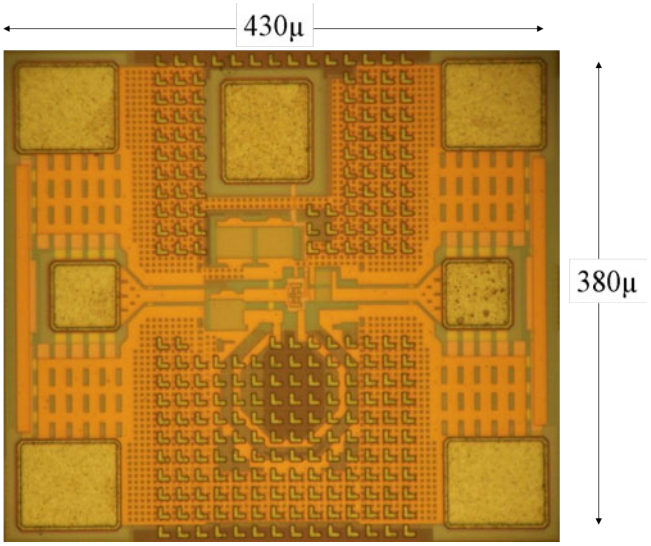


Figure 7.6: The chip micrograph for the 104GHz oscillator.

The circuit was fabricated in 90nm standard CMOS process. Figure.7.6 shows the micrograph of the implemented oscillator. The core area, ignoring bias and output pads, is extremely compact, 0.034 mm^2, and is dominated by the loop choke inductor. The output spectrum of the circuit was measured using a down

converter and a spectrum analyzer.

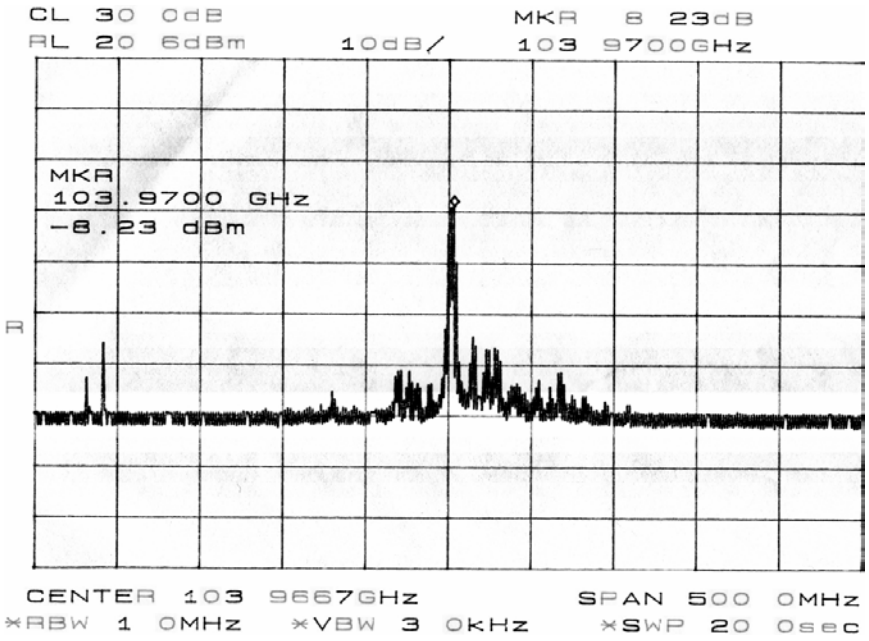


Figure 7.7: The output spectrum of the 104GHz oscillator.

Figure.7.7 shows the measured output spectrum of the circuit which shows -8.2 dBm of output power at 103.97GHz with 6.5mW of power dissipation at 1V of supply voltage. Considering the loss in the output cable which is approximately 3dB, these results are in close agreement with the simulation results. As evident from the graph, the measured spectrum is noisy and contains many spurious tones. The bulk of these tones are undesired low frequency oscillations, up-converted and grouped around the main tone. This is mainly due to the lack of a large bypass capacitance that would suppress low frequency loop gain. The large noise figure of the down-converter together with the dirty spectrum makes it very hard to determine the phase noise based on measurement.

As mentioned earlier, the gate bias could be used as the control voltage for this

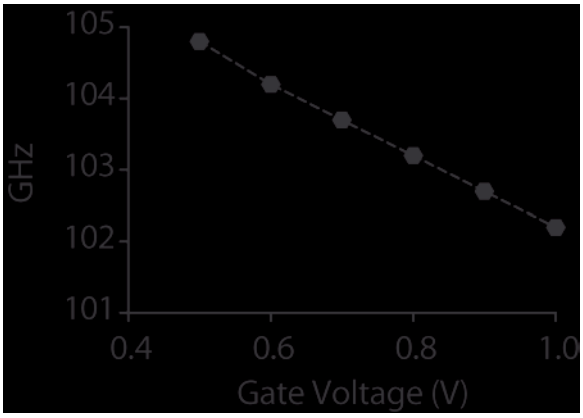


Figure 7.8: The simulated oscillation frequency versus gate voltage.

oscillator. Because of asymmetric distribution of the channel, the C_{gs} of the device increases in step with the V_{gs} and the device goes deeply into saturation. That is, an increase in bias voltage will decrease the oscillation frequency resulting in a tuning range of up to 3GHz, which is sufficient for some applications. Figure.7.8 illustrates the frequency pulling versus control voltage for this oscillator.

The overall performance of this oscillator in comparison with previously reported CMOS mm-wave oscillators is presented in a table. The power efficiency of this oscillator in particular is up to two orders of magnitude better than earlier work. This is mainly due to the use of an optimized high performance device, application of accurate large signal models and employing the fundamental frequency rather than harmonics.

7.2 An Amplifier at $f > f_t$ of The Technology

The 104 GHz oscillator demonstrates a non-linear operation of a circuit at the f_t of the technology. To show the linear functionality of the devices and models,

CMOS Process	Freq (GHz)	Power (mW)	Pout (dBm) (meas)	Core Area (mm2)	Ref
0.25µm	63	119	-4	0.351	ISSCC04[21]
90nm	100	120	-65	-	ISSCC04[19]
0.18µm	52.5	41	-8	0.8	ISSCC02[22]
0.13µm	114	8.4	-26	0.2	ISSCC05[20] (2nd harmonic)
90nm	104	6.5	-8.2	0.034	This Work

Figure 7.9: - Comparision between the fabricated oscillator and other reported mm-wave oscillators

a prototype amplifier was also designed at 104 GHz, a frequency larger than the f_t of the technology. The design of the 104 GHz amplifier is essentially very similar to the 60 GHz LNA as explained in chapter five. Unlike 60 GHz, devices at this frequency are unconditionally stable, and the optimal input and output impedances could be selected only based on gain considerations. The schematic of the amplifier is shown in Fig.7.10. Due to lower Maximum available gain at 100 GHz, one gain stage was added to the circuit compared to the 60GHz amplifier. Compared to the 60 GHz amplifier, the length of transmission lines are naturally shorter and a significant part of matching networks is now realized through the series inductance of coupling and bypass capacitors. This design was based on de-embedded measurement data of the active devices and models of passive devices.

Fig.7.11 shows the micrograph of the circuit. The dimention of the circuit is 940μm $\times$ 360μm that is quite compact for a three stage mm-wave amplifier. The

two long transmission lines are bent to reduce the area. The bend, as verified by EM simulations, does not have a significant effect on the charachtristics of transmission lines even at 100 GHz. The measured S-parameters of the 104 GHz amplifier are shown in Fig.7.12. The amplifier has the peak gain of 9.34 dB at 103.8 GHz. The input is well matched and its reflection coefficient is -9.8 dB. The output is not matched to 50 Ω and its refection coefficient is -5.5 dB. The circuit draws 22 mA from a 1 V power supply.

Summary This chapter explores the true limits of CMOS circuits at any given technology node. It was shown that one can consider the f_t of the transistor as a frequency limit by which he can design linear and non-linear circuits with an acceptable performance, considering the effects and losses of passive devices. To show this, two circuits were designed and implemented at a frequency larger than the f_t of the device: A 104 GHz amplifier as the worlds fastest CMOS linear circuit and a Collipitts based fundamental oscillator at the same frequency, also showing a record performance in terms of output power, power dissipation and chip area.

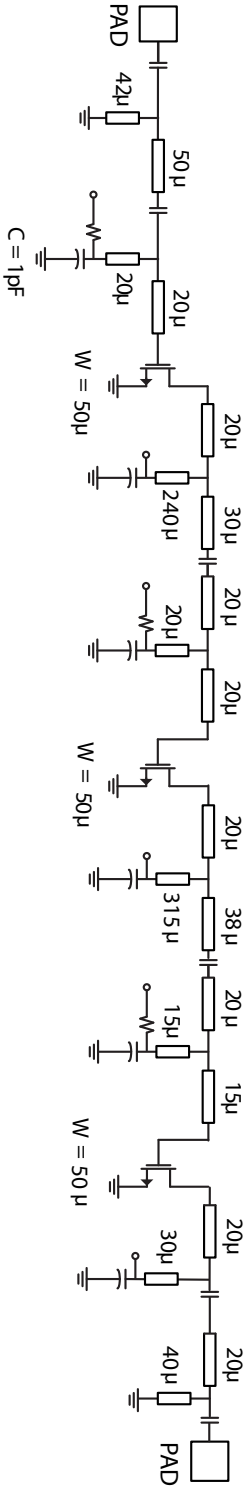


Figure 7.10: The schematic of the 100GHz amplifier.

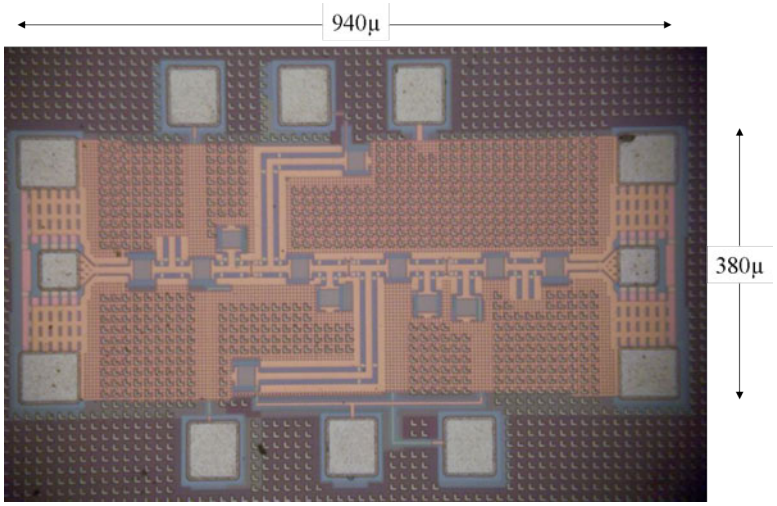


Figure 7.11: The chip micrograph of the 104GHz amplifier.

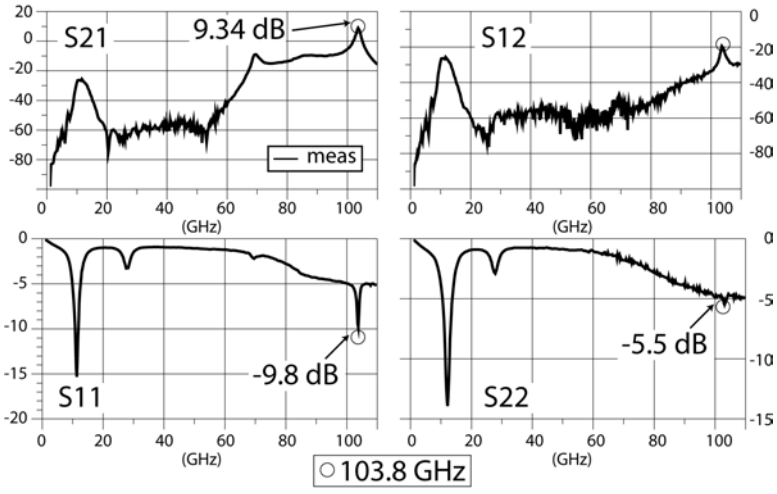


Figure 7.12: S-parameters measured results of the 104GHz amplifer.

8

Conclusions

This dissertation has explored CMOS device performances and circuit design up to 100 GHz, addressing issues ranging from performance optimization and modeling of CMOS devices to high performance circuit blocks up to 104 GHz. Important finding of this research can be summarized into few points:

Modeling of active and passive devices is an essential step in designing high performance mm-wave circuits. A CMOS transistor can successfully be modeled using a proper lumped equivalent circuit even beyond 100 GHz and available compact models can be used as the core of large-signal models once proper bias independent equivalent circuit of parasitics and substrate is added to it. On the measurement side, *recursive modeling* proves to be capable of accurately de-embedding the effect of measurement pads and extra connection wirings from the device. Cascode devices need special modeling due to their shared junction nature and their sensitivity to the parasitics on the second gate. All these were discussed in chapter 2.

The performance of mm-wave devices can be assessed based on several dif-

ferent parameters such as f_t, f_{max}, minimum noise figure and the amount of available gain. Most of these parameters, specially the f_{max} of the device are strong functions of the physical structure. The performance of these devices can be optimized by a study of the sensitivity of these parameters to the layout parasitics and improving the layout based on these findings. The *round-table* device is a special structure that shows a superior performance with an extrapolated f_{max} of 300 GHz. These optimized devices show a considerable power gain even beyond the f_t of the device. Large devices such as those used in the output stage of power amplifiers need different optimization procedure and the decision about the size and number of fingers and wiring methods should be made accordingly. The detail of device optimization can be find in chapter 3.

Chapter 4 talked about noise behavior of CMOS devices in the mm-wave frequency. It was shown that the noise could be modeled using two independent noise sources at the source and drain sides of the transistor with proper choices of source and drain temperatures. The optimal noise impedance at the source of the device approaches the optimal gain impedance as the frequency approaches the f_t of the device. On the other hand, the noise sensitivity of the device, R_n reduces with frequency, making the overall noise figure less sensitive to deviations from the optimal noise impedance. These imply that a noise figure close to NF_{min} is achievable once an amplifier is designed to maximize the gain. It was also shown that even in the optimal layouts, the gate resistance still contributes significantly to the overall noise figure of the device.

Using the proposed high performance devices and their developed accurate models, amplifiers were designed and implemented for the 60 GHz band as shown in chapter 5. A low power, high linearity LNA were demonstrated with 10.5 mW

of power dissipation, 6.5 dB of noise figure and $+4dBm$ of output P_{-1dB}. Based on the findings of chapter 4, it was shown that unlike low GHz frequencies, there is no need for an inductive degenerated devices in LNA design when the frequency is close to f_t. It was shown that interstage matching between a common source and a common gate device can improve the gain and noise performance of the overall circuit while keeping the power dissipation low through current sharing. A power amplifier was demonstrated that was designed for the maximum output power. The output power was improved by using an accurate large-signal model and linearizing the output device by inductively degenerating it at the source.

Chapter 6 explores the possibility of increasing the power gain of the device beyond its two port MSG. A general theory of unilateralization for N-port networks were developed and the theory was successfully tested for the case of single transistor in the common-source structure. It was shown that the stable gain of the device can exceed the value of mason gain and U is not the maximum achievable gain of a two port as sometimes is mistakenly mentioned. The unilateralization theory was used for a cascode structure and the gain peaking was observed and successfully modeled.

Chapter 7 explores the true limits of CMOS circuits at any given technology node. It was shown that one can consider the f_t of the transistor as a frequency limit by which linear and non-linear circuits can be designed with an acceptable performance, considering the effects and losses of passive devices. To show this, two circuits were designed and implemented at a frequency larger than the f_t of the device: A 104 GHz amplifier as the worlds fastest CMOS linear circuit and a Collipitts based fundamental oscillator at the same frequency, also showing a record performance in terms of output power, power dissipation and chip area.

As was mentioned in chapter one, mm-wave CMOS circuits are expected to become commercialized in the next few years, and will be a part of most of mobile devices including high-end cellular phones, game devices, digital cameras and wireless hard drives. They could also be used in wireless LAN systems where there is enough data speed to the local point through for example optical fibers. mm-wave car radars will also become ubiquitous and will increase the safety of the roads. High performance silicon technology will also find several applications in medicine once it can efficiently operate in the sub-THz range.

From research perspective, mm-wave circuit research will follow on two main path. On on path, as data communication in the 60 GHz band becomes more and more common, there should be a focus on reducing the power dissipation of 60 GHz CMOS to make it suitable for mobile devices. This should be realized using a combination of circuit and device techniques, similar to the *round table* transistor or unilateralization techniques as presented in this dissertation. To increase the effective range of such systems, high performance antenna arrays for 60 GHz will receive a great deal of attention. Apart from the mm-wave portions, the baseband is also critical when dealing with several Gbs of data. Realizing a reliable baseband with reasonable power dissipation needs a lot of creativity on the circuit side as well as the system side.

On another path, CMOS high frequency research will continue to push the boundaries of speed into the THz range. As it was shown in chapter one, CMOS circuits will be able to work at frequencies higher than 300 GHz in a few years. This however is only one of the pieces of the puzzle. Modeling and measurement techniques in order to effectively use this performance is a challenge that needs to be addressed. Also, unlike active devices, the performance of passive devices do

not improve with scaling and in the THz frequencies, most or all the gain of active devices might be lost as a result of poor passive devices. Benefiting from several metal layers in modern CMOS processes to build more efficient passive structures is an ongoing effort ans has been pursued in the last few years. Apart from that, co-design of devices and circuits could change the way we traditionally think about mm-wave circuits as transistor-matching network-transistor and reduce a significant part of potential performance losses.

Appendix A

A.1 Exact Calculation of R_n

To calculate R_n, the output voltage noise of the transistor for two cases are calculated. Once considering the noise contributions of r_{gs} and r_{ds} and once considering a voltage noise source , V_n, at the input of the transistor, considering the device as a noise free two port network. The value of R_n can then be calculated from the value of V_n.

For the case of noisy transistor, the output voltage noise can be written as the following equation:

$$v_2 = v_{n1}\frac{-g_m}{(1+j\omega c_{gs}r_{gs})(g_{ds}+j\omega c_{gd})} + v_{n2}\frac{g_{ds}}{g_{ds}+j\omega c_{gd}} \tag{A.1}$$

In this equation, v_{n1} and v_{n2} are noise sources associated with the r_{gs} and r_{ds} of the device respectively.

For the case of the noiseless two port with the value of v_2 can be calculated based on this equation:

$$v_2 = \frac{V_n}{A} \tag{A.2}$$

In which A is the first element of the $ABCD$ matrix of the device. The value of A can be calculated from the Y matrix of the device and it can be written as:

$$A = -\frac{(1 + r_{gs} j\omega c_{gs})\,(g_{ds} + j\omega c_{gd})}{g_m - j\omega c_{gd} + \omega^2 c_{gd} c_{gs} r_{gs}} \tag{A.3}$$

Using equations A.3,A.2 and A.1 the equation of R_n can be written as:

$$R_n = r_{gs}\frac{T_g}{T_0}\frac{g_m^2}{(g_m + c_{gs}c_{gd}r_{gs}\omega^2)^2 + c_{gd}^2\omega^2} + \frac{T_d}{T_0}\frac{g_{ds}(1 + r_{gs}^2 c_{gs}^2\omega^2)}{(g_m + c_{gs}c_{gd}r_{gs}\omega^2)^2 + c_{gd}^2\omega^2} \tag{A.4}$$

Appendix B

B.1 Calculation of noise parameters from the noise correlation matrix

Usually it's easier to start the noise calculations from the noise Y matrix. The Y matrix then should be transformed to the $ABCD$ noise matrix using this equation:

$$C_A^n = \begin{pmatrix} 0 & B \\ 1 & D \end{pmatrix} C_Y^n \begin{pmatrix} 0 & 1 \\ B^* & D^* \end{pmatrix} \tag{B.1}$$

In this equation, A, B, C and D are the elements of the two port $ABCD$ matrix. The $ABCD$ noise matrix of a two port network can be written as:

$$C_A^n = \begin{pmatrix} R_n & \frac{F_{min}-1}{2} - R_n Y_{opt}^* \\ \\ \frac{F_{min}-1}{2} - R_n Y_{opt} & R_n |Y_{opt}|^2 \end{pmatrix} \tag{B.2}$$

Accordingly, noise parameters could be calculated based on C_A^n matrix as follow:

$$Y_{opt} = \sqrt{\frac{C_A^{22}}{C_A^{11}} - (\Im\frac{C_A^{21}}{C_A^{11}})^2} + j\Im\frac{C_A^{21}}{C_A^{11}} \tag{B.3}$$

$$F_{min} = 1 + \frac{C_A^{21} + C_A^{22}Y_{opt}^*}{kT} \tag{B.4}$$

$$R_n = \frac{C_A^{11}}{kT} \tag{B.5}$$

Appendix C

C.1 The Equivalance Of Available And Mason Gain for Unilateral Networks

The Maximum Stable Gain(MSG) of a network can be calucluated from the Y parameters using this equation:

$$MSG = \frac{|Y_{21}|}{|Y_{12}|}\left(k - \sqrt{k^2 - 1}\right) \tag{C.1}$$

This can be written also as:

$$MSG = \frac{|Y_{21}|}{|Y_{12}|}\frac{1}{\left(k + \sqrt{k^2 - 1}\right)} \tag{C.2}$$

The value of k can be calculated using this equation:

$$k = \frac{2\Re(Y_{11})\Re(Y_{22}) - \Re(Y_{12}Y_{21})}{|Y_{12}Y_{21}|} \tag{C.3}$$

C.

If we define B as

$$B = 2\Re(Y_{11})\Re(Y_{22}) - \Re(Y_{12}Y_{21}) \tag{C.4}$$

For a unilateral network, the Y_{12} of the network approaches zero and we can write:

$$MSG_u = \lim_{y_{12}\to 0} MSG = \lim_{y_{12}\to 0} \frac{|Y_{21}|}{|Y_{12}|} \frac{|Y_{12}Y_{21}|}{B + \sqrt{B^2 - |Y_{12}Y_{21}|}} = \frac{|Y_{21}|^2}{4\Re(Y_{11})\Re(Y_{22})} \tag{C.5}$$

The last step is written based the fact that B simplifies to $2\Re(Y_{11})\Re(Y_{22})$ if the circuit becomes unilateral. The value of MSG_u is equal to the value of the Mason gain when Y_{12} is zero as was shown in chapter 6.

Bibliography

[1] J. F. Moore, "Predators and Prey: A New Ecology of Competition", *Harvard Buiesness Review*, vol.71 pp. 75-78, May-Jun. 1993 2

[2] J. Gowdy, "Coevolutionary economics: the economy, society, and the environment." Kluwer, Boston, 1994 2

[3] Special Report, "A Wireless World, Our Nomadic Future", *Economist Magazine*, April 10th, 2008 v, 4, 5

[4] Pew Research center, "http://people-press.org/reports/display.php3?PageID=1070" xi, 5

[5] Parks associates, "http://online.wsj.com/article/SB118885177843416106.html"

[6] iSuppli,"http://www.moconews.net/entry/419-mobile-game-revenues-hit-66-billion-mobile-video-14570-billion-by-2011-/" 6

[7] B. Rosenblatt, B. Trippe, S. Mooney, "Digital Rights Management: Buisness and Technology", First Edition, Wiley, Nov. 2001 6

[8] IEEE website, "http://standards.ieee.org/announcements/pr$_8$02153*wpanguide.html*"

[9] R. Lai et al. "Sub 50nm InP HEMT Device with Fmax Greater than 1 THz", *Digest of IEEE IEDM*, pp. 609-611, Dec. 2007 7

[10] X. Guan et al. ,"A fully integrated 24-GHz eight-element phased-array receiver in silicon",IEEE J. Solid-State Circuits, vol. 39, pp. 2311, Dec. 2004. 9

[11] A. Natarajan et al. "A 77 GHz phased array transmitter with local lo-path phase-shifting in silicon"', *IEEE ISSCC Dig. Tech. Papers*, vol. 49, Feb. 2006, p. 182. 9

[12] A. Babakhani et al."A 77 GHz four-element phased-array receiver with on-chip dipole antennas in silicon", *IEEE ISSCC Dig. Tech. Papers*, Feb. 2006, p. 180. 9

[13] A. D. Hellicar et al. "A 500-700 GHz system for exploring the THz frequency regime", *The 2nd International Conference on Wireless Broadband and Ultra Wideband Communications*, pp. 33-37, Aug. 2007 11

[14] L. Yujiri et al. "Passive mm-wave Imaging", *IEEE Microwave Magazine*,vol.4, pp.39-50, Sept. 2003 11

[15] J. C. Dickinson et al, "Terahertz Imaging of Subjects with Concealed Weapons", *Proceedings of SPIE*, vol. 6212, 2006.

[16] C. H. Doan, S. Emami. A. A. Niknejad, R. W. Brodersen, "Millimeter-Wave CMOS Design," *IEEE Journal of Solid-State Circuits*, vol. 40, , pp. 144-155,. Jan 2005.

[17] B. Razavi, "A 60-GHz CMOS Receiver Front-end," *IEEE Journal of Solid-State Circuits*,vol. 41, no.1, pp. 17-22, January 2006. 17, 80

[18] T. Yao, M. Gordon, K. Yau, M.T. Yang, and S.P. Voinigescu "60-GHz PA and LNA in 90-nm RF-CMOS," *IEEE RFIC Symposium Digest*, pp. 147-150, June 2006

[19] T. Mitomo et al. " A 60-GHz CMOS Receiver Front-End With Frequency Synthesizer", *IEEE Journal of Solid-State Circuit*, vol.43, pp. 1030-1037 April 2008 100

[20] S. Emami et al. "A Highly Integrated 60GHz CMOS Front-End Receiver", *Digest of ISSCC*, pp. 11-15, Feb. 2007

[21] B. Heydari, M.Bohsali, E.Adabi, A.M. Niknejad ' 'Low-Power mm-Wave Components up to 104GHz in 90nm CMOS," *ISSCC Dig. Tech. Papers*, pp. 200-201

[22] B. Heydari, M.Bohsali, E.Adabi, A.M. Niknejad ' 'mm-Wave Devices and Circuit blocks up to 104GHz in 90nm CMOS," *IEEE J. of Solid State Circuits*,vol.42 pp. 2893-2903 17, 39, 42, 92

[23] The BSIM 3v3 and 4.4.0 Website, BSIM Website 16

[24] MOS11 website, "MOS11 Website"' 16

[25] EKV model website, "EKV Website"' 16

[26] PSP model website, "PSP Website"' 16

[27] Y. Cheng, C. Hu, "MOSFET Modeling and BSIM3 User's Guide ", Springer ,1999 17

[28] O. J. Jia et al., " CMOS RF modeling for GHz communication IC's", *Digest of Technical papers, VLSI symposium*, pp. 94-95, Jun 1998 17

[29] T. F. Suet, K. Mayaram, " Substrate network modeling for CMOS RF circuit simulation"',*Proceeding of the IEEE Custom Integrated Circuits*, pp. 585-586, May 1999

[30] C. Enz, "'MOS Transistor Modeling for RF design 22
, *IEEE J. of Solid-State Circuits*, vol.35, pp. 186-201, Feb. 2000

[31] M. Mansun et al. " A robust and physical BSIM3 non-quasi-static transient and AC small-signal model for circuit simulation", *IEEE Transactions on Electron Devices*, vol.45, pp. 834-841, Apr. 1998

[32] X. Jin et al. " An Effective Gate Resistance Model for CMOS RF and Noise Modeling"',*IEEE Proceeding of IEDM*, Dec. 1998 47

[33] D. Lovelace et al. " Extracting small-signal model parameters of silicon MOSFETtransistors"',*Microwave Symposium Digest*,vol.2, pp. 865-868, May 1994 21, 23

[34] G. Dambrine et al. "A new method for determining the FET small-signal equivalentcircuit,*IEEE Transaction on Microwave Theory and Techniques*,vol.36, pp. 1151-1159, Jul 1988 21

[35] ICCAP website, ICCAP website 22, 25

[36] S. Emami, C. H. Doan, A. M. Niknejad, and R. W. Brodersen,"Large-signal millimeter-wave CMOS modeling with BSIM3," *IEEE RFIC Symp. Dig.*, June 2004, pp. 163-166 24, 95

[37] Y. Tsividis,"Operation and modeling of the MOS transistor",second edition, Oxford University Press, 2003 40, 51

[38] T. Ytterdal, Y. Cheng, T. A. Fjeldly, "Device Modeling for Analog and RF CMOS Circuit Design" 1st Edition, Wiley, 2003 24, 47

[39] Kolding T. E. "A four-step method for de-embedding gigahertz on-wafer CMOS measurements"',*IEEE Trans. Electron Devices*, April 2000, volume.47, pp.734740. 30

[40] Koolen M. C. A. M. et al. "An improved de-embedding Technique For on-wafer High-Frequency Characterization"', *IEEE 1991 Bipolar Circuits and Technology Meeting*, pp. 191194. 30

[41] Thomas H. Lee, "The Design of CMOS Radio-Frequency Integrated Circuits", second edition, Cambridge University Press, 2003. 35

[42] C.Y. Choong et al. "Small-signal substrate resistance effect in RF CMOS cascode amplifier", *IEEE Microwave and Wireless Components Letters*, vol.13, pp. 253-255, July 2003 35

[43] B. Heydari, E. Adabi, M. Bohsali, B. Afshar, M. A. Arbabian, A. M. Niknejad, "Internal Unilateralization Technique for CMOS mm-Wave Amplifiers", *RFIC Digest of Papers*, 2007, pp. 463-466. 36

[44] P. H. Woerlee et al. " RF-CMOS performance trends", *IEEE Transactions on Electron Devices*, vol.48, pp. 1776-1782, Aug 2001

[45] S. M. Sze, "High Speed Semiconductor Devices", Wiley, New York, 1990 42

[46] K. S. Cheon et al, " CMOS layout and bias optimization for RF IC design applications", *Digest of IEEE Microwave Symposium*, vol.2, pp.945-948, June 1997 39

[47] S . J. Mason, "Power gain in feedback amplifiers"', *Trans. IRE. Professional Group on Circuit Theory*, vol. CT-1, no. 2, pp. 20-25,. June 1954, June 1954. 43, 104

[48] M. S. Gupta, "Power gain in feedback amplifiers, a classic revisited,"' *IEEE Trans. Microw. Theory Tech.*, vol. 40, no. 5, pp. 864-879, May 1992. 117

[49] J. C. Guo, W.Y. Lien, M.C. Hung, C.C. Liu, C.W. Chen, C.M. Wu, Y.C. Sun, and Ping Yang, "Low-K/Cu CMOS Logic Based SoC Technology for 10Gb Transceiver with 115GHz fT, 80GHz fMAX RF CMOS, High-Q MiM Capacitor and Spiral Cu Inductor," *VLSI Digest of Technical Papers*, 2003, pp. 39-40. 39

[50] L. F. Tiemeijer et al., "Record RF performance of standard 90 nm CMOS technology," *IEDM Technical Digest*, 2004, pp. 441-444. 39

[51] A.M.Niknejad, "Electromagnetics for High-Speed Analog and Digital Communication Circuits",1st Edition. Cambridge press, 2007 42, 47, 108

[52] B. Razavi, " RF Microelectronics"', First Edition, Prentice Hall PTR, Nov. 1997 57

[53] H. T. Fries, "Noise Figures of Radio Receivers", *Proceedings of the Institute of Radio Engineers*, vol.32, pp.412-422, July 1944 57

[54] J. S. Goo, "High Frequency Noise in CMOS Low Noise Amplifiers", *PhD Thesis*, Stanford University, June 2001 57

[55] J. B. Johnson, "Thermal Agitation of Electricity in Conductors", *Nature*, vol. 119, pp. 50-51, 1927. 60

[56] A. van der Ziel, "Noise in Solid State Devices and Circuits"',John WileySons, New York, NY, 1986. 60, 62

[57] Pospieszalski, M.W, "Modeling of noise parameters of MESFETs and MODFETs and theirfrequency and temperature dependence," *IEEE Transactions on Microwave Theory and Techniques* vol.37, pp. 1340-1350 62, 119

[58] Scholten, A.J. Tiemeijer, L.F. van Langevelde, R. Havens, R.J. Zegers-van Duijnhoven, A.T.A. Venezia, V.C. "Noise modeling for RF CMOS circuit simulation," *IEEE Transactions on Electron Devices*, vol. 50, pp. 618-632, March 2003. 62

[59] P. Heymann et al. "Experimental Evaluation of Microwave Field-Effect Transistor Noise Models"', *IEEE Transactions On Microwave Theory and Techniques*, vol.47, pp. 156-163, Feb. 1999 66

[60] Jung-Suk Goo et al. " Physical Origin of the Excess Thermal Noise in Short Channel MOSFETs," *IEEE Electron Device Letters*, Vol. 22, No. 2, pp. 101-103, Feb. 2001. 62

[61] A. A. Abidi, "High-Frequency Noise Measurements on FETs with Small Dimentions", *IEEE Transactions on Electron Devices*,vol.33, no. 11, pp. 1801-1805, Nov.1986. 62

[62] RP. Jindal, "Hot-electron effects on channel thermal noise in fine-line NMOS field-effect transistors", *IEEE Transactions of Electron Devices*, vol.33, pp. 1395-1397, Sep 1986 62

[63] D. P. Triantis, " Thermal noise modeling for short-channel MOSFETs", *IEEE Transactions of Electron Devices*, vol. 43, pp. 1950-1955. Nov. 1996 62

[64] R. Navid et al, " The physical phenomena responsible for excess noise in short-channel MOS devices", *International SISPAD Conference*, pp. 75-78, 2002 62

[65] C. H. Doan, " CMOS Noise Modeling and Measurement at 60 GHz", *Berkeley Wireless Research Center, Winter Retreat Poster*, Jan. 2005 66

[66] H. Fukui, "The noise performance of microwave transistors", *IEEE Transactions on Electron Devices*, vol.13, pp. 329-341, March 1966 59

[67] C. C. Cheng, "Neutralization and unilateralization," *IRE Trans. Circuit Theory*, vol. CT-2, no. 2, pp. 138-145, June 1955. 107, 108

[68] M. P. Van der Heijden , L. C. N. De Vreede and J. N. Burghartz, ""On the design of unilateral dual-loop feedback low-noise amplifiers with simultaneous noise, impedance, and IIP3 match," *IEEE J. Solid-State Circuits*, vol. 39, pp. 1727, Oct. 2004.

[69] Z. M. Tsai et al, " A new feedback method for power amplifier with unilateralization and improved output return loss", *IEEE Transactions on Microwave Theory and Techniques*, vol.54, pp. 1590-1597, June 2006

[70] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. John Wiley and Sons, 2001.

[71] D. K. Shaeffer and T. H. Lee, "A 1.5-V 1.5-GHz CMOS Low Noise Amplifier," *IEEE J. Solid-State Circuits*, vol. 32, pp. 745-759, May. 1997.

[72] D.M. Pozar, " Microwave Engineering"', 3rd Edition, Wiley ,2004 18, 76

[73] A.M. Niknejad, H. Hashemi, " mm-wave devices and circuits"', 2008 9, 67

[74] H. Samavati, H. Rategh, and T. Lee, "A 5GHz CMOS Wireless-LAN Receiver Front-End," *IEEE J. Solid-State Circuits, vol. 35, pp 765-772*, May 2000. 90

[75] B. Afshar, A. M. Niknejad, "X/Ku Band CMOS LNA Design Techniques," *Proceedings of CICC*, 2006, pp. 389-392.

[76] E. Adabi, B. Heydari, M. Bohsali, and A. M. Niknejad,"30GHz CMOS Low Noise Amplifier," *Proceeding of RFIC*, 2007, pp.625-628.

[77] "Microwave Integrated Phased-Array Transmitters in Silicon", PhD Thesis, Caltech, Aug. 2006. v, viii, 10, 80

[78] G. D. Vandelin, A. M. Pavio, U. L. Rohde, " Microwave Circuit Design, using linear and non-linear techniques"',second edition, Wiley, 2005 59, 60

[79] C.-M. Lo, C.-S. Lin, H. Wang, "A Miniature V-band 3-Stage Cascode LNA in 0.13μm CMOS," *ISSCC Dig. Tech. Papers*, pp 322-323, Feb. 2006.

[80] G Gonzalez, "Microwave Transistor Amplifiers"', 2nd Edition , Prentice-Hall, Inc1996. 18, 75, 85

[81] C. Enz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications"', *Springer J. of Analog Integrated Circuits and Signal Processing*vol.8, pp. 83-114 126

[82] L. M. Franca-Neto, R. E. Bishop, B. A. Bloechel, "64GHz and 100GHz VCOs

in 90nm CMOS Using Optimum Pumping Method," *ISSCC Dig. Tech. Papers*, pp. 444-445, Feb. 2004. 122

[83] P.-C. Huang, M.-D. Tsai, H. Wang, C.-H. Chen, C.-S. Chang, "A 114GHz VCO in 0.13μm CMOS Technology," *ISSCC Dig. Tech. Papers*, pp. 404-405, Feb. 2005. 122

[84] R.-C. Liu, H.-Y. Chang, C.-H. Wang, and H. Wang, "A 63 GHz VCO using a standard 0.25 μm CMOS process," *IEEE ISSCC Dig. Tech. Papers*, Feb. 2004, pp. 446-447. 122

[85] M. Tiebout, H.-D. Wohlmuth, and W. Simbrger, "A 1 V 51 GHz fully integrated VCO in 0.12 μm CMOS," *IEEE ISSCC Dig. Tech. Papers*, Feb. 2002, pp. 238-239. 122