Design Techniques for Manufacturable 60GHz CMOS LNAs

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
Amneh M Akour, MSEE, BSEE
Graduate Program in Electrical and Computer Engineering

The Ohio State University
2011

Dissertation Committee:
Professor Mohammed Ismail, Advisor
Professor Waleed Khalil, Co-Advisor
Professor Patrick Roblin
Professor Donald Houser
ABSTRACT

Emerging broadband applications are pushing for the need to build high data rate wireless transceivers at 60GHz for high volume low cost mobile devices. Central to the success of implementing such transceivers is the robust design of 60GHz CMOS RF front ends, especially the low noise amplifiers (LNAs). In the future, CMOS technology is expected to enable low-cost mm-wave applications such as high data-rate communication links, passive and active imaging and sensor systems, and instrumentation and measurements equipment. Building highly integrated inexpensive mm-wave CMOS devices requires high quality factor lumped and distributed passives with accurate and scalable transistor and passives models.

My research work focuses on demonstrating a methodology for generating a scalable compact model for on-chip transmission lines and interconnects on lossy silicon substrate. The model is demonstrated over 20-60 GHz frequency band using two types of transmission lines built in TSMC’s 0.18 µm CMOS technology. Several CPWs and Microstrip lines are designed with different lengths to verify the accuracy and scalability of the extracted model. The compact model shows an excellent agreement with measured data with maximum deviation in $S_{11}$ magnitude and phase of 9.2% and 5.6%, respectively, and maximum deviation in $S_{21}$ magnitude and phase of 10.1% and 6.6%, respectively. Compared to existing model extraction methodologies, the required time for generating the compact model and simulating transmission lines is reduced significantly.
The generated models are fully compatible with all commercial circuit simulators.

This work presents key design techniques for different CMOS mm-wave LNA topologies. The proposed LNA topologies are, the three-stage cascode RF NMOS configuration and four-stage Common Gate followed by Common Source configuration. Simulation results for 60GHz LNAs show that the first topology can achieve a peak gain of 16.67 dB with a 3-dB bandwidth of 7 GHz, and a noise figure less than 11.04 dB over the entire bandwidth. The achieved peak gain from the second topology is 9.7 dB with a 3-dB bandwidth of 7 GHz and a noise figure less than 13.06 dB over the entire bandwidth. Simulation results for Sub-THz LNAs show that the first topology can achieve a peak gain of 23.5 dB at 92.1 GHz with a 3-dB bandwidth of 25 GHz, and a noise figure less than 5.5 dB over the entire bandwidth. The achieved peak gain from the second topology is 23.2 dB at 105 GHz with a 3-dB bandwidth of 15 GHz and a noise figure less than 6.1 dB over the entire bandwidth. The LNAs are designed and tested in 90nm RF CMOS.

Moreover in today’s radio design environment, the front-end analog devices in the transceiver require several silicon spins before they meet the specifications and they have relatively low yields. My work aims to propose a digital self-calibration technique for LNAs’ to enhance the yield to at least 90%. The proposed technique is shown to maintain typical specified performance at worst case corners in the presence of random process, supply and temperature variations, and hence allowing for manufacturable 60GHz RF CMOS design for high volume applications without leading to over-design or increasing power consumption.
DEDICATION

To my great mother Khawlah, my brother Mutazz, my son Laith, and my daughters

Yasmeen, Maysa, and Nisreen
ACKNOWLEDGMENTS

I would like to thank the many friends and colleagues who have supported and challenged me during my tenure at The Ohio State University. First and foremost, I wish to start by expressing my thank and gratitude to my advisor and friend professor Mohammed Ismail, who gave me the opportunity to join his group at the VLSI lab and he always provided the encouragement and help I needed during these years. He taught me how to do research systematically and with enthusiasm. More importantly in my case, I could not be able to continue in this new research topic without his infinite support and guidance during my PhD journey.

I would also like to express my gratitude to my excellent co-advisor Professor Waleed Khalil for his guidance and continuous encouragement during my Ph. D. research work. Waleed has been an exceptional role model with his insistence on high academic standards, his great accessibility to students, his incomparable work ethic, and his natural charisma. I am also grateful to Professor Patrick Roblin who has been a welcome source of sage advice. I would like to thank him for taking the time to serve on my oral exam committee and for reading and improving my dissertation. I would like also to thank Professor John Volakis, who gave me the support and advice during our group meeting at the Electro-Science Lab. I have benefited many times from his technical insight and his years of experience, and I thank him for his time and advice, given so generously. I am also indebted to Professor Roberto Rojas-Teran for his help and guidance in passives
modeling and measurement.

I am honored to call myself a member of two great labs (VLSI lab and Electro-science Lab) at OSU. I want to thank Gursharan Reehal, James Wilson, John Hu, Khalid Obeidat, Shadi Shunnaq, Will Moulder, and Sleiman Bou Sleiman for their friendship. Your discussions have led me to valuable insights into the directions I should take in my research. In particular, I would like to thank Sleiman Bou Sleiman who has been my constant friend and research partner throughout my years at VLSI lab.

One person who contributes disproportionately to making ECE department a wonderful place is LuCinda Flores. Her administrative competence, contagious enthusiasm, and warm feeling ease many of the trails of graduate student life, and I am grateful to her assistance, often under great time pressure, and for her positive attitude. In addition to LuCinda, there are many staff members with ECE department who deserve my thanks. These include Patricia Toothman and Stephanie Muldrow who are always available to guide and help student at our department. Don Gibb, Edwin Lim, Vincent Juodvalkis, and Aaron Aufderheide, who work diligently behind the success to keep the computer systems running smoothly, with special thanks to Don Gibb who always tries to make things perfect and always gives you the best solutions.

My great gratitude is to my best friend Ghassan Bin Hummam and his wife Fatin. They are always beside me when I need any help, encouragement, and support at each stage during my PhD years. Their friendship is one the best thing I have in my life.

In addition to those in the Ohio State University community, I want to thank several people outside OSU who contributed directly to my work. In particular, I thank
Semiconductor Research Corporation (SRC) and Texas Instruments for partially funding my research work. My thanks also go to Jad Atallah and Saul Rodriguez at the Royal Institute of Technology (KTH) in Sweden for their help with the tape out of the 60 GHz test structures.

I also need to thank my master advisor, Professor John Basart from Iowa State University, for his mentorship and his encouragement to pursue graduate work and for his valuable discussions during my PhD work. A special thank to my friend Dr. Osama Shanaa for his help in selecting this topic and for his continuous encouragement and support. I thank my many friends at OSU and elsewhere for their support and friendship. Among those that I have not already mentioned are my closest friends Balsam Dakhil and Nahla Al-Huraibi in taking care of my kids in many places and replacing me in critical cases.

One of the many popular idioms says “Behind a successful man, there is a great woman”. In my case, I would like to say “Behind a successful woman, there is a great man”. This great man in my life is my friend and brother Mutazz Akour. He is always beside me giving his support and advice. Without his support and encouragement, I do not think I would have been able to finish this long journey. He has had, and continuous to have a positive influence on my decisions, behavior, and life in general. Thank you so much, Mutazz.

I must reserve my special thank to my kids, Laith, Yasmeen, Maysa, and Nisreen Alqaisi, who have enriched my life in so many ways. They were so patient, cooperative, and supportive to make this success. Their presence in my life gives me a meaningful
reason to continue my work even during the tough times in our lives.

Finally, I would like to dedicate this work to my great mother, Khawlah Al-Banna who gave me a strong home and always emphasized the importance of education to her children. Her major investment in me has led to the production of this work, and without her, this work would have never been coincided. I wish she would be with me now to share these moments of success. But, she will always be remembered as the major pillar, supporting me in times of naught, and cheering me up in times of sorrow. I hope she will somehow receive my tremendous appreciation for the great role she played in my life.
VITA

1989……………………... B. S. in Electrical Engineering, University of Jordan
Amman, Jordan

1997……………………... M. S. in Electrical Engineering, Iowa State University
Ames, Iowa

PUBLICATIONS

JOURNALS

1. Amneh Akour and Mohammed Ismail, “Manufacturable 60GHz CMOS LNAs”, (Submitted to IEEE Transactions on VLSI Systems”, in 2011)

2. Amneh Akour, Waleed Kahlil, John Volakis, and Mohammad Ismail “A Compact Model for mm-Wave Transmission Lines and Interconnects on Lossy CMOS Substrates” (In preparation to be Submitted to IEEE Trans on Microwave Theory and Techniques SPECIAL ISSUE ON "mm-Wave Circuits and Systems", in July 2011)


PATENT


CONFERENCE PROCEEDINGS


6. S. Bou-Sleiman, A. Akour, W. Khalil, M. Ismail, “mm-Wave BiST and BiSC using a High-Definition Detector in 90nm CMOS”, MWSCAS 2010.

**FIELDS OF STUDY**

Major Field: Electrical and Computer Engineering

Studies in:

  Analog Microelectronics, Radio Frequency IC Design
# TABLE OF CONTENTS

ABSTRACT........................................................................................................... ii  
DEDICATION.......................................................................................................... iv  
ACKNOWLEDGMENTS ......................................................................................... v  
VITA........................................................................................................................ vi  
PUBLICATIONS ...................................................................................................... ix  
LIST OF TABLES ..................................................................................................... xiii  
LIST OF FIGURES ................................................................................................... xiv  

CHAPTER 1 ............................................................................................................. 1  
INTRODUCTION ...................................................................................................... 1  
1.1 CMOS mm-Wave Transceivers Motivation and Challenges ....................... 2  
1.2 First Pass Silicon and SoC ............................................................................. 5  
1.3 Self Awareness .............................................................................................. 7  
1.4 Thesis Outline ............................................................................................... 9  

CHAPTER 2 ......................................................................................................... 12  
LITERATURE REVIEW – 60GHZ TRANSCIEVERS ........................................... 12  
2.1 Motivations, Challenges and Channel Properties at 60 GHz ..................... 15  
2.2 CMOS 60 GHz Transceivers Challenges ..................................................... 20  
2.3 Transceiver Topologies ............................................................................... 28  
2.4 Conventional CMOS Transistor Models ..................................................... 34  
2.5 Modeling Passive Devices ......................................................................... 37  
2.6 Conclusions .................................................................................................. 38  

CHAPTER 3 ......................................................................................................... 40  
PASSIVE DEVICES MODELING ....................................................................... 40  
3.1 Transmission Lines Types ......................................................................... 42  
3.2 Transmission Line Characterization ............................................................ 48  
3.3 De-embedding Transmission Line Measurements and Momentum .......... 51  
  Simulator Calibration ......................................................................................... 51  
3.4 Circuit Model Extraction ............................................................................ 60  
3.5 Verification Results ..................................................................................... 64  
3.6 Conclusions .................................................................................................. 67  

CHAPTER 4 ......................................................................................................... 68  
DESIGN MM-WAVE LNAS ............................................................................... 68  
4.1 LNA Fundamentals ...................................................................................... 69  
4.1.1 Gain ......................................................................................................... 69  
4.1.2 Input/Output Matching .......................................................................... 70  
4.1.3 Noise ....................................................................................................... 72  
4.1.4 Linearity .................................................................................................. 74  
4.1.5 Stability ................................................................................................... 75  
4.1.6 Reverse Isolation ..................................................................................... 76
LIST OF TABLES

Table 1.1 Block measured parameters for self awareness ..................................................8
Table 2.1 Examples of Short-Range Wireless Multimedia Applications ............................13
Table 2.2 Comparison of small signal parameters for a round-table layout versus a regular multi-finger layout .................................................................25
Table 4.1 Reported performance of CMOS amplifiers in 60 GHz frequency band ...........97
Table 4.2 Reported performance of CMOS amplifiers in Sub-THz frequency band ........97
Table 5.1 Simulation Results for the Reference 60 GHz LNAs .....................................120
Table 5.2 Simulation Results for the Calibrated 60 GHz LNAs ......................................121
Table 5.3 Corner Simulation Results for the two Sub-THz LNAs ....................................122
Table 5.4 Simulation Results for the Calibrated Sub-THz LNAs ...................................122
LIST OF FIGURES

Figure 1.1 Block level self awareness ................................................................. 7
Figure 2.1 5/60 GHz System Scenario in Office and Home Environments .......... 17
Figure 2.2 5/60 GHz Radio ................................................................................. 18
Figure 2.3 60GHz Transmitter and Receiver .......................................................... 23
Figure 2.4 The Phased Antenna Array ..................................................................... 24
Figure 2.5 Conceptual Picture of Round-Table Device .............................................. 25
Figure 2.6 Block Diagrams of the 60-GHz Super-Heterodyne Receiver and Transmitter 30
Figure 2.7 Convention Homodyne Receiver ............................................................. 32
Figure 2.8 Small-Signal Transistor Model for an NMOS Device Showing the Important Parasitic Elements ............................................................................. 34
Figure 2.9 Small-Signal Equivalent Circuit of a Common-Source CMOS Device .... 36
Figure 3.1 Thin Film Microstrip Transmission Line .................................................... 44
Figure 3.2 Folded Microstrip Transmission Line ......................................................... 45
Figure 3.3 Coplanar Waveguide Transmission Line ...................................................... 46
Figure 3.4: Vertical-Ground-Plane Transmission Line .................................................. 47
Figure 3.5 Per-unit length RLCG model of TL ............................................................ 48
Figure 3.6 Die photograph A) CPW transmission lines B) Microstrip transmission lines and C) De-embedding structures ........................................................................ 52
Figure 3.7 Signal flow graph representing test fixtures and DUT ............................... 53
Figure 3.8 Measured and calibrated S11 for CPW and Microstrip line ...................... 59
Figure 3.9 Measured and calibrated S21 for CPW and Microstrip line ...................... 60
Figure 3.10 Flow-chart for extracting high frequency response of line ..................... 61
Figure 3.11 Extracted Series Inductance per 100µm .................................................. 62
Figure 3.12 Extracted Series Resistance per 100 µm .................................................. 62
Figure 3.13 Extracted shunt capacitance per 100µm .................................................. 63
Figure 3.14 Extracted shunt conductance per 100µm .................................................. 63
Figure 3.15 Simulated MoM, measured, and segmented S11(dB) .............................. 65
Figure 3.16 Simulated MoM, measured, and segmented S11(phase) ......................... 66
Figure 3.17 Simulated MoM, measured, and segmented S21(dB) .............................. 66
Figure 3.18 Simulated MoM, measured, and segmented S21(phase) ......................... 67
Figure 4.1 Common-Source LNA ............................................................................ 79
Figure 4.2 Common-Gate LNA .............................................................................. 80
Figure 4.3 Inductive Degeneration Cascode Common-Source LNA ......................... 81
Figure 4.4 Two stage CG – CS LNA ....................................................................... 84
Figure 4.5 Three-Stage Common-Source LNA ......................................................... 85
Figure 4.6 RF-Front End with Modified Inductive Cascode Source Degeneration LNA 86
Figure 4.7 Cascode Common-Source LNA ............................................................... 87
Figure 4.8 Effect of Transistor Width on fT and fMAX .............................................. 88
CHAPTER 1

INTRODUCTION

Developing and designing wireless access and wireless LAN systems operating at 60GHz have experienced a remarkable renaissance in the last few years. It is considered a major part of any fourth-generation system. The Federal Communication Commission (FCC) has designated the 7GHz band around 60 GHz (59 – 66 GHz) for general unlicensed use for ultra-high speed wireless indoor LAN supporting up to several gigabits per second. The advantage of this high data rate is that it reduces the energy dissipated per bit. In addition to that, the millimeter wavelength permits designs with high levels of integration since it permits the integration of multiple antennas in one chip. However, designing the 60 GHz transceiver presents many challenges in building passive and active devices. The techniques used at one digit Giga frequencies need to be modified to overcome the new limitations at mm-wave frequencies.

Analog circuits require several silicon cycles to meet their specifications, because they suffer from digital noise coupling and have low yield due to process, temperature and power variations. For these reasons, designing RFIC blocks in platform baseband Systems on Chip (SoC) has been restricted. Therefore, it is important to have some design techniques to handle the problems of parasitic elements and process shift. This requires several kinds of self calibration for RF blocks after fabrication to compensate for
these variations.

This chapter introduces the motivations and challenges in designing mm-wave CMOS RF blocks. Different solutions are proposed to overcome these challenges at both device and circuit techniques levels will be discussed in the next chapters. This chapter also shows the importance and motivations behind what is called “first pass silicon” especially for the mm-wave transceivers using nano-meter design technologies. Designing techniques leading to first pass success of mm-wave LNA are introduced. At the end of this chapter, the thesis outline is presented.

1.1 CMOS mm-Wave Transceivers Motivation and Challenges

The primary goal for this dissertation is to explore techniques for designing a manufacturable mm-wave LNA in an inexpensive complementary metal-oxide-semiconductor (CMOS) technology. Although the techniques developed apply somewhat generally across many wireless systems, the specific focus of this work is on the mm-wave wireless systems. Because the 7 GHz band around 60 GHz provides a convenient vehicle for supporting ultra-high speed and high-data rate systems.

For further cost reduction, CMOS technology implementation for this system is considered as a unique solution. Due to the huge investment in CMOS, It is natural to consider whether the technology’s deficiencies can be mitigated, making it attractive in the domain that historically has been dominated by more expensive silicon bipolar and GaAs technologies.

Meeting the goal of receiver integration in inferior technology especially at this high
frequency requires innovation in architectures, circuit and device modeling. In addition, some self calibration techniques are needed to improve the RF and analog block for analog mixed signal integration. Overall, the scope of the problem is broad, but a successful approach will bring clear benefits for wireless customers. These considerations motivate this research into highly integrated CMOS mm-wave receivers where the subject of this dissertation (LNA) is the major and first block of this receiver.

The three major challenges in using the 60GHz transceivers for high speed LAN are:
The first challenge is the specific attenuation characteristic due to atmospheric oxygen of 10-15dB/Km. This factor limits the usage of the 7 GHz band around 60 GHz only for all kinds of short-range wireless communication. In addition to that, the free loss which increases quadratically with the frequency is much more at this high frequency. This loss can be compensated for by using a special antenna with pattern directivity and small dimension.

The substantial drop of the received power because of the antenna obstruction and mispointing due to the weak diffraction of millimeter waves is the second challenge in building a 60GHz transceiver. Beside this, the transmittivity also depends strongly on material properties and thickness. Therefore, it is necessary to create a reliable shared medium and hot spot communication by having at least one access point per indoor environment.

The channel dispersion at 60 GHz is much smaller than the value at low frequencies because of the confinement to smaller cells and the shorter echo paths. Moreover, Doppler effects are proportional with frequency. Therefore, at 60 GHz, the channel
suffers from relatively severe Doppler effects due to the movement of the portable station and the movements of the object.

In spite of all these challenges, the possibility to design a high data rate wireless system using simpler modulation formats attracts researcher to explore the idea of designing 60 GHz wireless systems. The first generation of the 60 GHz transceiver has been implemented by GaAs PHMET technology, or by SiGe Bipolar technology, but the CMOS performance improvement due to technology scaling increases the possibility to build these transceivers using inexpensive CMOS technologies. Designing 60 GHz CMOS transceivers has more challenges and limitations. CMOS substrate suffers from excessive high path losses due to oxygen absorption, the high sheet resistance of the silicide gate, the high gate leakage, the low breakdown voltage, and the channel-length modulation due to the thinner gate oxide thickness and short channel length. To overcome these limitations, new CMOS technologies need to be used with double-sided narrow finger and with close substrate contacts to the device. Actual performance of a device at mm-wave frequency is highly layout dependent, so large-signal performance can still be maintained if care is taken through layout. While FET with $f_T$, and $f_{MAX}$ greater than 300 GHz are now available in 90nm CMOS nodes, accurate and scalable FET models in the mm-wave regime are challenging tasks. The present FET CMOS models have been designed and tested only up to 10 GHz frequency.
1.2 First Pass Silicon and SoC

Current and future trends look for the highest levels of integration to realize low cost and low power for handheld wireless devices. During the last few years, intense research has been performed on the integration of Analog Mixed Signal (AMS) circuits in System-on-Chip (SoC) and Network-on-Chip (NoS) platforms. However, designing the radio part of a wireless solution remains a major bottleneck. The integration of digital and analog circuits in one chip presents several challenges such as digital noise coupling through substrate and power lines. Radio frequency circuits are strongly affected by random variations in process, temperature, operating conditions and power supply and this leads to relatively low yield. Several design cycles are required in order to meet all product specifications and to reach an acceptable yield. To overcome these problems, designers build their circuit based on worst-case corners simulations, which leads to over-design and increased power consumption. On the other hand, digital circuits are designed to achieve high yields above 99% at all process corners in only one design spin. Moreover, RF models, package models and design kits are only valid under certain conditions and assumptions, which limit design space exploration severely. Therefore, first pass silicon is important in designing mixed signal circuits.

First pass silicon requires meeting all block level goals, meeting the overall system specifications, performing only one tape-out, and increasing the chip yield to an acceptable level. The one silicon cycle is important in order to decrease the Non-Recurring Engineering (NRE) cost, decrease the mask set cost, and save time to market. The design cycle passes through five basic steps: system and design specifications, block
design, block layout, fabrication, and testing. The cost for a design cycle increases due to the increase of the mask set cost in every new process generation. This design cycle starts all over again if the chip does not meet specifications. Moreover, each re-spin requires about six extra months of work. Therefore, first pass silicon is so important in designing mixed signal systems.

Integrating RF blocks have all the issues of analog integration in addition to several new problems. The most important problem in designing RF blocks is the load variations during the design process. Designing the current block cannot wait for the next block to be finished. Therefore, the current block makes some assumptions about the next block input impedance to begin its design cycle. Information about the next block input impedance is updated and the test bench is adjusted to account for the new changes. At the end, the current block needs to be re-simulated using lumped elements to represent the final value of the next block input impedance and also using the actual next block. The lumped equivalent circuit is only valid for the frequency at which the values were extracted and is not able to show the out of band performance.

Moreover, the top level interconnects form a part of the load tank and need to be included in the simulation for accurate results. To model these interconnects between blocks; a 3D field solver is often used. The interconnection routings have significant inductance depending on their length and the size of the coupling (represented by both inductive and capacitive elements) depending on the separation between them. Another interconnection effect is the mutual inductive coupling between the routing and the load tank. This effect might introduce some imbalance between the positive and negative
paths of the differential blocks. If this effect is not considered and modeled properly, the performance of the block varies and shifts from the desired goal, especially the desired gain and resonance frequency. Without knowing the actual resonance peak, it becomes very difficult to estimate and hunt down the parasitic elements that shifted the load. More details about SoC and integration issues will be discussed in chapter 5.

1.3 Self Awareness

Webster defines self awareness as “An awareness of one’s own personality or individuality.” Another definition for self awareness by L. Ron Hubbard is “Self-awareness includes recognition of our personality, our strengths and weaknesses, our likes and dislikes. Developing self-awareness can help us to recognize when we are stressed or under pressure.” What is the relationship between these definitions and an RFIC? Self awareness in RFIC is interpreted at two different levels: system level self awareness and block level self awareness. Self awareness at block level means that each block needs to evaluate and to quantify its performance. Each block requires being able to measure its input and output characteristics as shown in Figure 1.1 [1]

Figure 1.1 Block level self awareness [1]
Now, which parameters need to be measured? And how are these parameters measured by the block? Answering the first question depends on the block itself. Each block has some critical parameters, which are considered key parameters for its performance, need to be measured. For example, the gain, linearity, noise and input match are the key parameters for the LNA performance. While, I/Q imbalance, phase noise, output power or DC offset are not. Table 1.1 displays some common blocks and the required measured parameters for each one.

Table 1.1 Block measured parameters for self awareness

<table>
<thead>
<tr>
<th>Block</th>
<th>Gain</th>
<th>Linearity</th>
<th>Noise</th>
<th>Input Match</th>
<th>I/Q</th>
<th>DC Offset</th>
<th>Filter Corner</th>
<th>Phase Noise</th>
<th>Output Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNA</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Mixer</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>BaseBand</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>PLL</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>PA</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Answering the second question depends on the block itself too. For the RF blocks, LNA, mixer and PA, an RF detector is used to measure the input and output for the gain and linearity parameters. For the baseband circuits, a digital demodulator is used to measure these parameters. While for the PLL, an RF detector is used to measure the output power and a special phase noise circuit in the baseband is used to measure the phase noise. More details on RF detector will be discussed in chapter 5.

The most critical parameters in designing an RF detector are its low power consumption, its flat frequency response, and its minimal loading effect on the circuit.
under test. The absolute accuracy of the detector is not important to measure the block gain, but it needs to be calibrated first to measure an absolute value. This can be done by measuring the power of a reference source such as the PA output before the real measurements of the device under test (DUT). The linearity measurement is similar to the gain one. The only difference in the linearity measurement is in using the PA to provide a swept power. By sweeping the PA output power over certain range, both the gain at each step and the compression power can be extracted.

After building a self awareness block using an implemented mechanism to track their performance, the calibration loop needs to be closed by identifying the cause of the impairments and using circuit techniques to turn the performance of the block.

1.4 Thesis Outline

The following chapters delve into the problem of mm-wave receiver design in general and of LNA design in specific in details. The ultimate goal is the design and implementation of a manufacturable 60 GHz CMOS LNA that meets the specifications under all process, temperature, and supply variations in IBM-90nm CMOS process. The techniques developed along the way are, however, broadly applicable to other wireless systems.

Chapter 2 begins by discussing the motivations, challenges, and channel properties at 60 GHz. Then in chapter 2, the extra challenges in designing mm-wave CMOS transceivers are introduced. In addition, a review of the current state of the art in non-CMOS and CMOS receiver research establish a context for the present work. At the end
of chapter 2, an overview of radio receiver architectures, the proposed modifications on active device models, and the modeling of passive are presented in general.

Chapter 3 tackles the subject of generating an accurate and efficient electrical model for two types of transmission line; microstrip and coplanar waveguide, using Agilent’s Momentum simulator in great details. This includes the advantages and advantages of each type to select best one for mm-wave applications. Then, de-embedding and calibration of the Momentum simulator is demonstrated. At the end of chapter 3, the proposed method for extracting compact RLCG circuit models for the two types of transmission lines using the calibrated simulator is described and compared with the measurements.

Chapter 4 discusses the subject of CMOS mm-wave low-noise amplifiers in great details. Chapter 4 begins with an overview of LNA fundamental concepts, parameters, and low frequency architectures with special attention paid into common-source inductive degeneration LNA. Lastly, the most suitable modified mm-wave LNAs are designed to maximize the achievable gain, minimize the produced noise, and optimize the other LNA parameters. Simulation results for different 60 GHz and mm-wave LNAs are presented for two architectures.

Chapter 5 follows with an investigation of digital Built-In Self-Calibration (BIST) technique to design manufacturable mm-wave LNAs. This chapter focuses on how to design 60 GHz LNAs which are robust against process, temperature, and power supply variations and have a high yield for the purpose of System-on-Chip integration. To put these theoretical developments into practice, four self-calibrated LNAs (two for 60 GHz
receivers and two for sub-THz ones) are implemented in IBM CMOS 90nm. The experimental results demonstrate a high level of performance and integration that is comparable to or better than existing implementations, thereby confirming the value of the techniques presented in earlier chapters. Finally, chapter 6 concludes with a summary and some suggestions for future work.
CHAPTER 2

LITERATURE REVIEW – 60GHZ TRANSCIEVERS

Design and development of wireless access and wireless LAN system operating at 60 GHz is considered as part of the fourth-generation (4G) systems. There exist numerous multimedia applications calling for wireless transmission over short distances. Table 2.1 summarizes these applications with some estimates of data rates and cost requirements [2]. This table shows that the data rates ranges are very wide up to hundreds of megabits per second, and the information integrity and security play an important role.

The third-generation cellular systems are not built based on low-cost technology and may not be able to cope with data rates in excess of 2 Mb/s. Due to these limitations, WLAN systems come into the picture. WLAN products are based on either the IEEE 802.11b standard at 2.4 GHz band with a gross capacity up to 11 Mb/s, or the IEEE 802.11a/g at 5 – 6 GHz and 2.4 GHz with input data rates of 6 - 54 Mb/s. Because of the data rate limitation of this type of communication, WLAN does not support the required data rates for attractive services listed in table 2.1. To achieve higher network capacity, two approaches can be used: increasing the spectral efficiency, and/or using more bandwidth.

Increasing the spectral efficiency can be implemented using different methods. The first method is to use higher order modulation methods such as M-array quadrature
Table 2.1 Examples of Short-Range Wireless Multimedia Applications [2]

<table>
<thead>
<tr>
<th>Application</th>
<th>Capacity per user [Mb/s]</th>
<th>Low cost requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wireless LAN bridge, e.g., for interconnecting Gigabit LANs in different buildings</td>
<td>10-100</td>
<td>No</td>
</tr>
<tr>
<td>Wireless virtual reality allowing free body movements</td>
<td>450</td>
<td>Yes</td>
</tr>
<tr>
<td>Wireless IEEE 1394</td>
<td>100, 200, 400</td>
<td>Yes</td>
</tr>
<tr>
<td>Wireless TV high-resolution recording camera</td>
<td>150-270</td>
<td>No</td>
</tr>
<tr>
<td>Wireless trading terminal having multiple video channels that can be viewed simultaneously for monitoring world news next to stock quote information</td>
<td>50-100</td>
<td>No</td>
</tr>
<tr>
<td>Wireless news tablet, a very thin, possibly flexible device that provides the user with a newspaper, e.g., the possibility of activating images to see video impressions</td>
<td>50-100</td>
<td>Yes</td>
</tr>
<tr>
<td>Wireless (high-quality) videoconferencing</td>
<td>10-100</td>
<td>Yes</td>
</tr>
<tr>
<td>Wireless Internet download of lengthy files</td>
<td>10-100</td>
<td>Yes</td>
</tr>
<tr>
<td>Wireless ad hoc communications, i.e., direct communication between notebooks, between notebook and nearby printer, etc.</td>
<td>0.1-100</td>
<td>Yes</td>
</tr>
<tr>
<td>Wireless interactive design</td>
<td>20-40</td>
<td>Yes</td>
</tr>
<tr>
<td>Hospital bedside application allowing wireless retrieval of patient’s status including X-ray pictures</td>
<td>10</td>
<td>Yes</td>
</tr>
<tr>
<td>Wireless surveillance cameras allowing face and number plate recognition at long distances</td>
<td>4-10</td>
<td>Yes</td>
</tr>
<tr>
<td>Patient monitoring (patients can walk freely around the hospital or even at home) with devices that transmit ECG, blood pressure information, etc.</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>Wireless videophone</td>
<td>1.5</td>
<td>Yes</td>
</tr>
<tr>
<td>Wireless connection between domestic appliances and the Internet (e.g., a refrigerator scans its contents and orders the nearby supermarket to deliver what is missing)</td>
<td>0.1</td>
<td>Yes</td>
</tr>
<tr>
<td>Wireless billing (e.g., automatic payment for petrol service via wireless connection between car and filling station)</td>
<td>0.1</td>
<td>Yes</td>
</tr>
<tr>
<td>Road pricing</td>
<td>0.1</td>
<td>Yes</td>
</tr>
<tr>
<td>Wireless burglar alarm (wireless window sensors, etc.)</td>
<td>0.01</td>
<td>Yes</td>
</tr>
<tr>
<td>Remote control (TV, lighting, door/window lock)</td>
<td>0.01</td>
<td>Yes</td>
</tr>
<tr>
<td>Wireless embedded systems (e.g., in car between oil filter and dashboard)</td>
<td>0.01</td>
<td>Yes</td>
</tr>
</tbody>
</table>

amplitude modulation (M-QAM) which comes at the cost of transmitting extra amount of power in the link budget to maintain the required bit error ratio. The second method is to utilize space by using multiple transmit and receive antennas and transmitting different data streams on the different transmit antennas simultaneously. Different techniques can be used to accomplish this goal. The first technique is to use space-time coding which encodes the data by channel code and then split them into parallel streams for
transmission on different antennas. The second technique is to use multipath scattering generated by appropriate processing architecture. Both techniques are paid with complexity in terms of processing power and number of transceivers. The third method for increasing the spectral efficiency is to create spatially orthogonal channels with highly aimed gain antennas which can be used widely in HIPERLAN family. This method is suitable for moderate spectral efficiency of many hundreds of megahertz at higher frequencies.

The European Advanced Communications Technologies and Services (ACTS) program, the Multimedia Mobile Access Communication (MMAC) Committee in Japan, and the Federal Communication Commission (FCC) address the 40 GHz and 59 - 66 GHz frequency bands for general unlicensed applications to get ultra-high speed wireless indoor LANs supporting 156 Mb/s [2], [3].

In this chapter, section 2.1 presents the motivations, challenges and channel properties at 60 GHz. The 60 GHz CMOS Transceivers Challenges are discussed in section 2.2. Section 2.2 also reviews and summarizes some of the researches’ results that have been done in the literature review toward designing 60 GHz wireless systems. In section 2.3, different transceiver topologies are presented to determine the best one for 60GHz applications. Section 2.4 discusses the active device modeling challenges and the proposed modifications on these models to consider some device mechanisms that are not well captured at low frequencies. However, section 2.5 covers briefly the modeling challenges and limitations for passive devices at mm-wave frequencies where this topic will be discusses in details in Chapter 3. At the end of this chapter, section 2.6
summarizes the key parameters that need to be considered in designing CMOS 60 GHz RF front-end for the future wireless LAN systems and highlights the major problems and the required improvements in the present 60 GHz transceivers.

2.1 Motivations, Challenges and Channel Properties at 60 GHz

The most important benefit of using the 60 GHz band is the availability of large block of universal unlicensed spectrum around this frequency band. This huge frequency covers 7 GHz frequency band in the USA and at least 3 – 4 GHz frequency band in the intersection of the set of all international standards. This large bandwidth translates into high throughput which supports a high data rate with low complexity modulation schemes. High data rate is necessary and important in many applications. The first application of high data rate is in the domain of personal area networks (PANs) where it is required to transfer huge content between multimedia portable devices and a computer. The second application includes transmitting uncompressed video between a multimedia device and an HD flat screen television. The third application can be in the gigabit Ethernet in the WLAN counterpart. If frequency goes up to 77 GHz, then the fourth application will be in automotive radar to detect collision and to control the speed and will be in mm-wave imaging for security.

Based on Shannon’s theorem, the channel capacity is proportional to the bandwidth and a logarithmic function of SNR. Since around 60 GHz frequency band, there is a large untapped bandwidth, this allows designer to use relatively inefficient and simpler modulation schemes and still realize high data rate. While, in low-frequency band (below
10 GHz), to achieve high data rates with smaller bandwidth requires more complex modulation schemes to compress as many bits/Hz as possible.

Another advantage of the 60 GHz band over any other frequency spectrums such as the UWB frequency band between 3 – 10 GHz is relaxing requirement of the FCC regulation on the maximum transmit power. FCC regulations limit the transmit power for the UWB systems to an average of about 0 dBm. In contrast, the maximum power transmission for 60 GHz applications is about 40 dBm. This adds a great amount of complexity in the UWB radios. In summary, power and bandwidth are the critical motivation in building communication systems in 60 GHz frequency band.

In principle, there is about 7 GHz bandwidth centered around 60 GHz available for dense wireless communications. Therefore, Federal Communications Commission (FCC) assigns the 59 – 64 GHz frequency band for general unlicensed applications. While the specific attenuation characteristic due to atmospheric oxygen of 10-15dB/Km makes the 60 GHz band unsuitable for long-distance (< 2 Km) wireless communication. Therefore, the 7 GHz band around 60 GHz can be used only for all kind of short-range wireless communications (indoor environment) as shown in Figure 2.1. In addition to that, the free loss increases quadratically with the frequency, so the free loss at 60 GHz is much more than the one at one-digit GHz band. This loss can be compensated by using special antenna with pattern directivity and small dimension.

The second challenge in designing the 60 GHz system is the substantial drop of the received power because of the antenna obstruction and mispointing due to the weak diffraction of millimeter waves. This causes a considerable drop of received power and a
possibility in nullify the provided gain by the antennas. To overcome the problem of blocking effect (obstruction), omnidirectional antennas can be used to collect contributions of reflected power in the event of line of sight (LoS) obstruction.

The third challenge in building mm-Wave transceivers is the strong dependency of the transmitivity on material properties and thickness. For example, the transmitivity of glass is 3 – 7 dB, while for concrete wall is more than 36 dB (depends on its thickness). This means the concrete walls and floors will be considered as cell boundaries. Therefore, it is necessary to create a reliable shared medium and hot spot communication by having at least one access point per indoor environment. For best coverage point, the

Figure 2.1 5/60 GHz System Scenario in Office and Home Environments [2]
access point should be near the center of the room at a high position near the ceiling [2]. A small antenna(s) will be mounted on a wall and it can be connected to the existing LAN cabling. To overcome the drop in the received power below certain threshold due to LoS obstruction by object is to apply macro diversity by switching to another access point, but this requires more than one access point in a room and increases the cost significantly. A better solution for this problem is to use low- or medium-gain antennas with particular antenna patterns.

Another solution for coverage limitations due to wall attenuation and the severe shadowing is to operate the 60 GHz system in combination with lower frequency system as shown in Figure 2.2 [2] where the 60 GHz system is combined with 5 GHz WLAN system. At nominal propagation conditions, the system tries firstly to reach the 60 GHz users. When the channel conditions get worse, the system connection switches to the 5 GHz band. The radio frequency selection will determined based on channel conditions, channel availability, user preferences, and connection parameters. The other system radio frequency can be used as an intermediate frequency (IF) for the 60GHz system.

Figure 2.2 5/60 GHz Radio [2]
The fourth challenge is the channel dispersion at 60 GHz. The channel dispersion in mm-Wave communications is much smaller than the ones at low frequencies due to the confinement to smaller cells and the shorter echo paths. This causes higher RMS delay spread in the range from few to 100ns especially if omnidirectional antennas are used for large reflective indoor environment. On the other hand, if high gain antennas are used, the RMS delay is less than few nanoseconds.

Lastly, Doppler effects are proportional with frequency. Therefore at 60 GHz, the channel suffers from relatively severe Doppler effects due to the movement of the portable station and the movements of object.

To enable reliable high-speed transmission in an indoor environment when low or medium gain antennas are used, a single-carrier channel equalization or multicarrier transmission scheme has to be used. Therefore, the orthogonal frequency-division multiplex (OFDM) will be most suitable technique for high-speed transmission at 60 GHz. OFDM technique has many attractive properties over other techniques. Firstly, the OFDM modem complexity vs. bit rate does not grow as fast as the complexity of a single-carrier system with an equalizer which makes it easier to implement these modems. Secondly, OFDM has easy scalability to different environments, bandwidths, or bit rates which gives the possibility to use for different applications at the same time. On the other hand, many factors have to be considered carefully in OFDM technique. The first factor is the Doppler shift effect. To avoid experiencing significant loss of coherent receiver, the data rate per OFDM subcarrier must be above certain minimum level. The second factor is phase noise produced by the local oscillators in the transceiver. This
influence become more serious if the OFDM is combined with higher order modulation schemes such as 16- and/or 64-QAM. To minimize the oscillator noise, the OFDM subcarrier spacing should be as high as possible.

In spite of all of the above challenges, the wide bandwidth available around 60 GHz frequency with the possibility to get multi-giga bits per second data rates using variety modulation formats attracts designers to build 60 GHz wireless systems. Therefore, for practical high-volume and low-cost application in the 60 GHz ISM band, the radio chipset should achieve the highest possible integration level avoiding both lossy off-chip interconnects at millimeter-wave frequencies and expensive packaging technology.

2.2 CMOS 60 GHz Transceivers Challenges

With all of the above motivations in using 60 GHz radios, many challenges and limitations need to be overcome to build mm-wave circuits. The Japanese were the early innovators in exploring this field and demonstrating working transceivers [3]. The first challenge in building any 60 GHz circuit is to check the validity of lumped circuit theory at mm-wave frequencies. Two basic components need to be checked in any circuit: active devices (transistors) and passive devices (matching circuits and interconnects). Active devices are still considered small and their dimension is still a fraction of the wavelength. Therefore, they can be treated as lumped circuits. While, passives have large dimensions and need to be treated as distributed circuits.

The second challenge is designing 60 GHz circuits is the lack of infrastructure, CAD tools, and models. Noise measurements at these frequencies are considered as a
challenging and difficult task even with expertise people. All measurement facilities need to be upgraded to support all circuit measurements (such as small-signal, large-signal and noise measurements) up to 65 GHz. The process nodes need to be characterized through extensive test chips, and then can be used build a library of active and passive devices for mm-wave design.

Recent advances in SiGe, BiCMOS, and Si CMOS technologies increase the possibility to build a low-cost radio receivers and transmitters at mm-Wave frequencies. The first generations of 60 GHz transceiver have been implemented by GaAs PHMET technology [3] – [4], or by SiGe Bipolar technology [5] - [11]. For high-volume and low-cost 60-GHz applications, high integration level is required in the radio chipset to avoid both lossy off-chip interconnections at mm-Wave frequencies and expensive package technologies. The most highly integrated 60 GHz blocks reported in the literature is shown in [3] – [11]. Ohata et al. [3] report on a 1.25Gb/s 60GHz-band compact transceiver module uses ASK modulation. They used 0.15um-Gate AIGaAsAnGaAs Hetero-junction FET with f_{max} 0f 240 GHz. They used a transmitter filter to cut higher order spurious to reduce the occupied bandwidth. Their transmitter exhibits 9.6dBm output power, while their receiver shows -50dBm minimum received power for 1.25 Gb/s error-free transmissions. Gunnarson et al. in [4] also report 60GHz Transmitter and Receiver MMICs on a 0.15µm GaAs PHEMT Technology. The LO operates in the range 7 – 8 GHz, and then is multiplied using an integrated by-eight (8X) multiplier. A three-stage power amplifier were used to achieve a 3.7 dBm output power over the RF frequency range of 54 – 61 GHz with peak power of 5.2 dBm at 57 GHz. On the receiver
side, three-stage low-noise amplifier (LNA) with an image reject mixer has been built. The receive achieves a gain of 7.5 dB in the frequency range 55 – 63 GHz, an image rejection ratio more than 20 dB between 59.5 – 64.5 GHz, a noise figure of 10.5 dB, and an input-referred third-order intercept point of 11 dB.

S. Reynold et al. and B. Floyd et al. in [5] – [7] report their double-conversion superheterodyne receiver and transmitter 60 GHz chipset using 0.13 μm SiGe BiCMOS technology. Their receiver chipset includes LNA, RF-to-IF and IF-to-baseband mixers, IF amplifier strip, phase-locked loop (PLL), and frequency tripler. The RX noise figure was about 6 dB, the IIP3 was -30 dBm, and RX power consumption was 500mW. The transmitter chip includes power amplifier (PA), image-reject driver, IF-to-RF and quadrature baseband-to-IF up-mixers, IF amplifier strip, PLL, and frequency tripler. The TX P1dB was 10 – 12 dBm, the Psat was 15 – 17 dBm, and the power consumption was 800mW. Their chipset have packaged with planar antenna. The wireless data link for this system was 630 Mb/s data rate at 10m as shown in Figure 2.3.

The availability of wider spectrum, the need for higher data rate, and the potential for high volume applications are generating interest in investigating the applicability of advanced CMOS technologies for solutions in X-band (8-12 GHz), K-band (18-26 GHz) and V-band (40-70 GHz). CMOS transceivers potentially provide highly integrated inexpensive mm-wave devices and can eventually be transformed into main stream high volume applications. FET fT, fMAX, and NFMIN will improve with every new process generation, although the concurrent constraints on supply voltages will present an
increasing challenge.

Designing 60 GHz CMOS transceivers has many extra challenges and limitations. CMOS substrate suffers from excessive high path losses due to oxygen absorption which occur precisely in this frequency range. Furthermore, the limitation in transmitting proper power level to overcome this path loss due to the reduced supply voltages of nanoscale CMOS adds extra challenge in designing circuits at this frequencies. This problem can be resolved by using CMOS strength in integrating large number in transceivers on one die. A phased antenna array is used to improve the antenna gain and the efficiency of the radiated power as shown in Figure 2.4. At mm-wave spectrum, the antenna element size will be very small and can be integrated into the package. This allows beam forming to improve antenna gain and increase spatial diversity. This allows spatial power combining that simplifies the power amplifier design. Since the path loss drops with smaller wavelength, a simple dipole antenna with medium gain can be employed to capture a smaller portion of the radiated power. For mobile system and to move the antenna beam in the antenna array, a phase shifter in the transmitter and receiver can be used.

Figure 2.3 60 GHz Transmitters and Receiver [6]
While FET with $f_T$, and $f_{\text{MAX}}$ greater than 300 GHz are now available in 90nm CMOS nodes, accurate and scalable FET models in mm-wave regime are a challenging task. The present FET CMOS models have been designed and tested only up to 10 GHz frequency. Therefore, these models need to be modified for better characterization of these devices at mm-wave frequencies. Active device models will be discussed in section 2.4, while passive device models will be discussed briefly in section 2.5 and in details in chapter 3.

Other challenges in designing mm-wave circuit in CMOS are its high sheet resistance of silicide gate, high gate leakage, low breakdown voltage, and the channel-length modulation because of the thinner gate oxide thickness and short channel length. To overcome these limitations, new CMOS technologies need to be used with double-sided narrow finger and close substrate contacts to the device. Actual performance of devices
at mm-wave frequency is highly layout dependent, so large-signal performance can still be maintained if care is taken through careful layout. The array layout, dubbed the "round table" layout as shown in Figure (2.5), is preferred to reduce importance of external parasitics on determining the maximum frequency of activity [12]. Table 2.2 shows a comparison of small signal parameters for a round-table layout versus a regular multi-finger layout.

Table 2.2 Comparison of small signal parameters for a round-table layout versus a regular multi-finger layout [12]

<table>
<thead>
<tr>
<th></th>
<th>Regular 1</th>
<th>Round Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_e$ (Ω) (total)</td>
<td>4.46</td>
<td>2.23</td>
</tr>
<tr>
<td>$R_d$ (Ω)</td>
<td>3.54</td>
<td>2.42</td>
</tr>
<tr>
<td>$R_s$ (Ω)</td>
<td>672m</td>
<td>438m</td>
</tr>
<tr>
<td>$C_{in}$ (FF)</td>
<td>35.7</td>
<td>57.1</td>
</tr>
<tr>
<td>$C_{gf}$ (FF)</td>
<td>21.3</td>
<td>17.2</td>
</tr>
</tbody>
</table>

Figure 2.5 Conceptual Picture of Round-Table Device [12]
On the other hand, CMOS has some attractive benefits over other technologies. Some of these benefits are the size and the cost of the CMOS solutions. The digital CMOS technology is the lowest-cost option. The fast silicon scaling in CMOS decreases the device sizes and increases the possibility to integrate the full system in a single chip in what is named System on Chip (SoC).

Despite of all of these challenges and limitation in using CMOS, getting the benefits of wide unlicensed band, high data rate at 60 GHz, and fast silicon scaling in CMOS attracts few researchers to start building some 60GHz CMOS transceivers to achieve high volume of integration in CMOS systems in the last few years [12]–[24]. C. Doan et al, in [12] – [15] discuss the design and modeling of CMOS transistors, integrated passives, and circuit blocks at mm-wave frequencies. They characterize the high frequency performance of 130-nm CMOS transistors and used this to design amplifiers at 40 GHz and 60 GHz. Their amplifier at 40 GHz achieves a peak $S_{21}$ of 19 dB, output $P_{1dB}$ of -0.9 dBm, $I_{IP3}$ of -7.4 dBm, and consumes 24 mA from a 1.5V power supply. Their 60 GHz amplifier exhibits a peak $S_{21}$ of 12 dB, output $P_{1dB}$ of +2.0 dBm, NF of 8.8 dB, and consumes 36 mA from a 1.5V power supply.

B. Heydari et al. [16] – [19] proposes a systematic methodology for layout optimization of active devices that can be used fro mm-wave applications and they develop a hybrid mm-wave modeling technique to extend the validity of the device models. Afterward, they design three stages common source amplifiers and a Colpitts oscillator. The first amplifier is centered around 60 GHz and provides a gain of 12.2 dB, an output $P_{1dB}$ of +4 dBm, and consumes 10.5 mW. The second amplifier is centered at
104 GHz and has a peak gain of 9.3 dB, input and output reflection coefficients of -9.8 dB and -5.5 dB respectively, and consumes 22 mA from a 1 V power supply. The Colpitts oscillator operates at 104 GHz while delivering up to -5 dBm of output power and consuming 6.5 mW.

C. Wang et al. in [20] report a 60 GHz low-power six-port transceiver. Their transmitter is implemented using a reflection-type I/Q modulator which include a 90° coupler to generate the quadrature-phase signals and an in-phase combiner. The receiver path includes an LNA and the I/Q modulator, the transmitter includes a mixer, LO, and buffer amplifier (BA). The LNA and BA is implemented using six-stage cascode devices. Both amplifiers provide a gain of 20 dB and noise figure of 8 dB between 60 – 64 GHz. The BA has P1dB of -2 dBm and consumes 36.9 mW, while LNA consumes 31 mW. Their VCO demonstrates a tuning range from 61.5 to 62.3 GHz with output power of -12.2 dBm at 62 GHz and phase noise of -92.2 dBc/Hz at 1MHz offset. The VCO consumes a total DC power of 30 mW.

Razavi in [21] – [23] reports different CMOS receivers and in [24] demonstrates a CMOS transceiver. Razavi uses two different topologies to build his receivers. He used direct-conversion and super heterodyne topologies. His receiver in [21] consists of a cascode LNA which is followed by a transconductance stage and a single-balanced mixer, IF mixers, local oscillator, divider, and IF amplifiers. Razavi uses an inductor at the cascode node to resonate with the capacitance at this node and overcome the bandwidth and noise limitations of the cascode architecture. His receiver exhibits a voltage gain of 26 – 31.5 dB, noise figure of 6.9 – 8.3 dB, P1dB of -15.5 dBm, image
rejection ratio of 44.5 dB, I/Q mismatch of 1.5 dB/6.5°, LO phase noise of -95 dBc/Hz at 1MHz offset, and a power consumption of 80 mV from 1.8 V power supply. In [22]-[23], he designed a common-gate LNA and an active mixer. The receiver front-end provides a voltage gain of 28 dB, a noise figure of 12.5 dB, a $P_{1dB}$ compression point of -22.5 dB, and consumes 9mW from a 1.2V power supply. His voltage gain was 8 dB lower than the expected, while his noise figure was 4 dB higher than expected. In [22], Razavi’s transmitter consists of an up-conversion mixer, an output stage, and an oscillator. The transmitter has been fabricated in 0.13 CMOS technology and consumes 44 mW.

2.3 Transceiver Topologies

Complexity, cost, power dissipation, and the number of external components are considered the essential criteria in selecting transceiver architectures. Some factors impose severe constraints upon transceiver design. The first factor is the limited spectrum allocated for each user. This factor limits the rate of transfer data and consequently uses sophisticated techniques such as coding, compression, and complex modulation schemes. This factor also impacts the design of RF blocks by employing narrowband modulation, amplification, and filtering to prevent the leakage of the adjacent channels at the transmitter, and by processing the desired channel while sufficiently rejecting strong interferers at the receiver.

The second factor is the dynamic range of the signals. The dynamic range requirement increases with multipath fading and path loss. Dynamic range has two limitations on the transceiver design. Firstly, the input noise of the receiver and the cross-
talk become critical as the minimum detectable signal decreases in the range of microvolt. This influence can noticed clearly in the FDD systems where the transmitted signal has a finite attenuation in the receiver band. In this situation, the leakage signal from the transmitter to the receiver might significantly desensitize the LNA. To overcome this problem, some FDD standards require high isolation or introduce an offset between transmit and receive time slots. Secondly, the receiver may experience large signals amplitudes as high as several hundred millivolts. The receiver must still process any signal correctly and this leads to the use of automatic gain control (AGC).

The last factor is related to the transmitter power amplifiers. Due to periodically turning on and off of the power amplifier to save power and due to the large current drawn by the power amplifier, a tremendous noise is introduced in the supplies. This might change the battery voltage by several millivolts. For this reason, noise immunity and supply rejection for all building blocks become so critical.

Generally speaking, two types of transceiver architectures are commonly used in wireless communication: heterodyne (low/high-IF) and homodyne or direct conversion (zero-IF). The major difference between these two architectures is the operating frequencies in analog baseband circuit. Super-heterodyne architecture; Figure (2.6), with a variable intermediate frequency (IF) has been used by many researchers in [7] and [24]. To bring center frequency from $\omega_1$ to $\omega_2$, the signal is mixed firstly with a signal of frequency $\omega_o$. The center frequency $\omega_2$ is called intermediate frequency (IF). IF plays an important rule in design a super-heterodyne transceivers and bearing trade-offs with any performance parameters. The selection of IF depends on trade-offs among the amount of
image noise, the spacing between the desired band and the image, and the loss of the image-reject filter [25]. The selection of IF also depends also on the availability and the physical size of filter at specific frequency. Heterodyne transceivers demonstrate a trade-off between image rejection which determine the sensitivity of the receiver and channel selection which determine the receiver selectivity.

Figure 2.6 Block Diagrams of the 60-GHz Super-Heterodyne Receiver and Transmitter

Heterodyne transceivers exhibit many drawbacks. An important drawback is the need for the image reject filter which is realized using external passive components. This requires the preceding LNA and the forgoing mixer to drive the 50-Ω impedance of the filter. Another drawback of heterodyne receiver is the complexity in selecting the NF, IP3, and the gain of each stage in the transceiver path which requires much iteration to reach to an acceptable gain distribution. The third drawback is its frequency planning to determine its overall performance.
In heterodyne 60 GHz transceiver, the input signal 59 - 64 GHz is amplified by an image-rejecting LNA. The amplified signal is mixed to approximately a 9-GHz IF signal, at which point it is filtered and amplified by the IF variable-gain amplifier (VGA). Then, the signal is split and fed to a pair of double-balanced mixers which covert the signal to in-phase (I) and quadrature-phase (Q) baseband. LO signals are generated by either tripling or halving the frequency of the voltage-controlled oscillator (VCO) signal. The IF signals was 8.4 - 9.1 GHz, while the VCO frequency was 16.8 - 18.3 GHz. The transmitter is a mirror-image of the receiver with up-converting mixers replacing the down-converting ones in the receivers. The image-rejecting pre-driver provides additional amplification before the final power amplifier stage. Heterodyne receiver avoids quadrature separation at mm-wave frequencies, easing the management of interconnect lengths, and allowing the integration of all high frequency building blocks.

The second commonly used transceiver topology is the homodyne (direct-conversion) architecture shown in Figure 2.7 and used in [12] – [20] and [22] – [24]. The circuit consists of low-noise amplifier (LNA), quadrature mixers, and base-band gain stages. A single-ended-to-differential (S/D) converter (balun) has been used due to the extremely difficult to externally generate and distribute differential local oscillator (LO) signals. The lower mixer loads the LNA without providing any additional information. The LO frequency in homodyne transceivers is equal the input carrier frequency. Therefore, only a low-filter with relatively sharp cutoff performance is required for channel selection.
Figure 2.7 Convention Homodyne Receiver

Homodyne architecture exhibits two advantages over the heterodyne counterpart. The first advantage is the simplicity in designing this transceiver since there is no image problem and no image filter is required any more. The second advantage is in the reduction in the components used. In homodyne transceiver, the low-pass filter with baseband amplifiers replace the IF filter and the subsequent down-conversion stages in the heterodyne transceiver. On the other hand, homodyne architecture faces many drawbacks. The first drawback is in designing the active low-pass channel select filter. Active filters suffer from sever noise-linearity power trade-offs. The second drawback is in the DC offset problem. This dc offset is generated from many sources in the transceiver such as LO leakage at the LNA port, or LNA large interfere leakage at the LO port. The extraneous offset voltage corrupts the signal and saturates the following stages. To solve this problem, transceiver must include some offset cancellation mechanisms. The most commonly used techniques are the DC-free coding and idle time intervals in digital wireless standards to carry out offset cancellation. The third drawback is the I/Q mismatch. The phase error in the 90° phase shifter and the amplitude mismatch between I
and Q signal corrupt the down-converted signal constellation and this increases the bit error rate. The fourth drawback is the even-order distortion due to mixer mismatch and mixer finite direct feedthrough from the RF input to the IF output. The fifth drawback is the flicker noise since the down-converted spectrum extends to zero frequency. Therefore, the 1/f noise of devices substantially corrupts the signal. This problem can be minimized by using large devices for the stages following the mixer and by employing DC-free coding.

As discussed before the major setback of heterodyne architecture is the image rejection problem, which requires an image rejection filter to be placed in between the LNA and RF mixer. This filter is designed to have a relatively small loss in the desired band and a huge attenuation in the image band which will increase the complexity and power consumption of the receiver. Heterodyne architecture exhibits a trade-off between image rejection and channel selection or a trade-off between the sensitivity and selectivity. On the other hand, homodyne architecture does not require image filter since the input signal is directly converted into zero-IF output, but this direct conversion to zero frequency in homodyne architecture generates an extraneous dc offset voltage which can corrupt the signal and saturate the following stages and causes the issue of flicker noise. Therefore, homodyne receivers require some means of offset cancellation. Based on the above discussion, each transceiver exhibits some advantages and suffers from few drawbacks and designers need to consider some additional techniques to overcome these problems.
2.4 Conventional CMOS Transistor Models

The available used models by circuit designers are the "compact" models as shown in Figure 2.8. Compact models are the interface between the technology and the design. Several good compact models have been developed for digital, analog, and RF applications [26] - [30]. The behavior of the device is described by general equations built using a combination of physical and empirical methods. To capture the details of a given technology, several parameters are embedded in each equation. Most compact models describe the behavior of the device in all regions of operations and provide small and large signal analysis as well as noise analysis. These extracted parameters are valid over a fair range of geometry, width and length of the device. Moreover, the core equations have been derived under quasi-static assumptions and most of the extracted parameters are for low-frequency applications. This makes these models less desirable and accurate for mm-wave applications.

![Figure 2.8 Small-Signal Transistor Model for an NMOS Device Showing the Important Parasitic Elements [15]](image-url)
FET models are of two types: firstly, the quasi-static models based on transistor dc nonlinearities, dc I-V curves, and charge-capacitor models. Secondly, the non-quasi models which are built from channel segments to consider the finite channel-charging time and the effect of instantaneous channel charge distributions. Most of the compact models used by the simulators are based only on the quasi-models. The quasi-models ignore the finite charging time of the carrier in the inversion layer and can not predict the frequency transadmittance.

A complete physical model for the transistor must include the source/drain resistances, the gate resistance, the non-quasi-static channel resistance, the resistive substrate network, the series inductor for all terminals to represent the signal delay, the layout-dependent interconnect capacitances, the extrinsic wiring capacitances, and the junction diodes.

**Proposed Modification on Compact Models**

As discussed in the previous section, most of the compact model parameters have been extracted for low-frequency applications. This causes inaccuracy in using these parameters for mm-wave applications. Some device mechanisms that are not well captured at low frequencies have considerable effect on the performance of the device at higher frequencies. The effect of substrate resistances and capacitances is an example of such effect. The induced gate noise is another example of a non-modeled effect in the present low-frequency compact model. At mm-wave frequencies, the layout effect causes another reason for inaccuracy in these models. Small inductors, resistances, and
capacitors due to the device interconnections to the outside world introduce additional components to the low-frequency model. These components change and dominate the performance of the device as the frequency increases.

A detailed full-wave electromagnetic simulation is required for an accurate prediction of these parasitics, but this simulation is difficult and lengthy. Therefore, each small finger of the transistor is modeled with the "intrinsic" transistor model as the core for a hybrid mm-wave model. Interconnects are captured by combination of selective electromagnetic simulation and experimental techniques. For a common-source CMOS device, a small-signal model as shown in Figure 2.9 is used [15] and [31]. Additional parasitic resistances and inductances as well as a substrate network have been added to the hybrid-JI model of the core transistor. The effect of the induced gate noise needs to be added to this model for accurate parameter extraction of the noise.

![Figure 2.9 Small-Signal Equivalent Circuit of a Common-Source CMOS Device [31]](image-url)
2.5 Modeling Passive Devices

Transmission lines are considered as the most important passive devices at this frequency band since they are used as a reactive element in the matching networks and resonator and they represent interconnects between different blocks. Transmission lines (TLs) have many benefits: TLs are scalable in length and can be used to realize small reactances, interconnects can be modeled directly when they are implemented using TLs, and TL has a well-defined ground return path that significantly reduces magnetic and electric field coupling. TL can be completely characterized by its equivalent frequency-dependent RLCG distributed circuit model.

The most commonly used transmission lines are the microstrip lines and the coplanar waveguide. Microstrip line uses the top metal layer as the signal line and the next metal layer as the ground plane. Microstrip has some advantages: microstrip has effective metal shield where no electric field penetrates into the substrate due to the lower metal ground plane, and microstrip has constant capacitive quality factor since the shunt loss (G) is due only to the loss tangent of the oxide. The major drawback of microstrip lines is the very small distributed inductance due to the close proximity of the ground plane to the signal line. This causes a degraded inductive quality factor. Coplanar waveguide (CPW) is the second option to build transmission lines. CPW is implemented with one signal line surrounded by two adjacent grounds on the same top metal layer. CPW characteristic is controlled by two parameters: W which represents the width of the signal line and it determines the conductor loss, and S which represents the spacing between the signal line and the ground plane and it controls the characteristic impedance of the line and the trade
off between the inductive and capacitive quality factors. CPW has some benefits over microstrip such as the higher inductive quality factor. The major drawback of the CPW is the undesired odd modes due to the presence of three conductors.

These transmission lines are characterized and studied in more details in chapter 3. Chapter 3 discusses in details the effort done in generating a compact model for these transmission lines. This model has been built using Momentum in ADS after calibrating the simulator using test structures. The simulator utilizes simulated calibrated S-parameters to generate an accurate frequency dependent distributed RLCG model library. The compact model shows an excellent agreement with measured data with maximum deviation in S11 magnitude and phase of 9.2% and 5.6%, respectively, and maximum deviation in S21 magnitude and phase of 10.1% and 6.6%, respectively. Compared to existing model extraction methodologies, the required time for generating the compact models and simulating transmission lines is reduced significantly. The generated models are fully compatible with all commercial circuit simulators.

2.6 Conclusions

In this chapter, different researchers’ works have been presented and summarized. The presented works at mm-wave frequencies show some differences between measurements and simulated results. This inconsistency or variation is due to many reasons. The first reason is the lack of accurate models for both passive and active device in these simulations at this high frequency band, since most of the present models are only valid up to 10 GHz. The second reason is the lack of testing the yield for these
designs. All of the presented works have been reported, performed and tested under nominal conditions. None of these works were discussing the simulation results under worst case coroners or under process, temperature or supply variations. Generally, RF and analog blocks have very low yield even at low frequencies where most of design issues are more mature. Therefore, we are expecting to give more attention to the yield of RF blocks at these very high frequencies. This means RF blocks need to be tested under critical corners and need to have some built-in self-calibration mechanisms which help the RF block to meet the desired performance under worst case scenarios. In the following chapters, a methodology for generating a scalable compact model for on-chip transmission lines and interconnects on lossy silicon substrate is presented. In addition, a method to provide built-in self-calibration and configuration for any RF block is introduced and is tested to design a 60 GHz LNAs. This method helps to achieve first pass silicon success and to optimize the performance of the radio.
CHAPTER 3

PASSIVE DEVICES MODELING

The exponential rate of reduction in minimum feature size of CMOS transistors has accelerated the pace of technology adoption for millimeter-wave (mm-wave) System-on-Chip (SoC) solutions. In comparison, CMOS circuits have specific advantages over SiGe or BiCMOS counterparts covering low-power active devices with high transition frequencies ($f_T$), and low-cost high volume manufacturing capabilities for mixed-signal/RF systems [32] – [33]. The major drawback of standard CMOS technology is the limited RF performance of its on-chip passive components and interconnects, which are deemed critical for mm-wave circuits [34], [35]. As such, the maximum achievable circuit speed and thus its yield are often constrained by the quality and accuracy of its interconnects. Existing parasitic extraction tools supplied by process design kits (PDKs) (e.g. simple low frequency modeling or RC-based modeling) provide inadequate and inaccurate modeling parameters at the post-layout simulation level. This is attributed to frequency dependent distributed effects of interconnects that are not considered in the model extraction flow. Therefore, an efficient and accurate model for transmission lines (passives or interconnects) is critically important to reduce the discrepancies between simulated and measured performance of mm-wave circuits.

Several modeling techniques have been proposed to characterize on-chip transmission lines (TL). In [36], transmission lines have been experimentally
characterized by the quasi-TEM, slow-wave, and skin-effect modes. In [37], a full-wave electromagnetic simulation based on finite physical layout (using HFSS tool) is proposed. To improve the run time in EM simulators, different assumptions are made: uniform doped silicon substrate, modeling multi-layer dielectric by two layers, and constant loss tangent. In general, the accuracy of EM simulators can be improved at the expense of long simulation time and large memory requirements. A third approach using conformal mapping techniques to derive simple analytical equations for transmission lines was proposed in [38]. These equations are built based on some invalid assumptions in modern silicon technologies such as an infinite substrate and/or single dielectric layer. While some of the above methods are compatible with commercial circuit simulators, they still suffer from either speed or accuracy problems.

In this chapter, an efficient and accurate electrical model for transmission lines based on a full-wave technique using Agilent’s Momentum simulator (MoM) is introduced. Generated models from Momentum can be easily extracted and then integrated in commonly-used simulators (such as HSPICE and Spectre) and optimized for less simulation time and memory requirements. To extract these models, S-parameter measurements have been performed on different microstrip and coplanar waveguide transmission lines to calibrate the Momentum simulator. Following, Momentum is used to generate S-parameters for variable lengths and widths transmission lines and to extract the corresponding compact per unit RLCG models. These models can then be used in the commonly-used circuit simulators for accurate and efficient full chip simulation.

Section 3.1 of this chapter reviews different kinds of on-chip transmission lines;
the advantages and disadvantages of each. Section 3.2 discusses the frequency-dependent characteristics of transmission lines. Section 3.3 discusses de-embedding transmission line measurements and the calibration of the Momentum simulator. Section 3.4 describes the proposed method for extracting circuit models. Section 3.5 summarizes the model results and compares the accuracy of the extracted RLCG model to the measured and the simulated S-parameters. Section 3.6 summarizes the accuracy of the proposed model.

3.1 Transmission Lines Types

The scaling in the new CMOS technologies can hardly scale down passive components in the circuits and passive elements still occupy most of the chip area. Therefore, compact and low passive components in silicon process are required especially with the induced loss of interconnections at high frequencies due to the conductive silicon substrate. There are two major reasons in introducing transmission line theory for mm-wave applications. Firstly, the electrical size of a network is the critical key that determines whether circuit theory or transmission line theory will be used in treating interconnecting wires. If the physical dimensions of a network are much smaller than the electrical wavelength, circuit theory will be used to create a lumped model to handle these interconnections. While if the physical dimension is a considerable fraction of a wavelength, then transmission line theory needs to be used for modeling. In this case, a transmission line will be represented as a distributed-parameter network, where voltages and currents can carry in magnitude and phase over its length. Secondly, at mm-wave frequencies, the required inductors for matching networks and resonators become increasingly small in the order of 30 – 300 pH. Therefore, transmission lines are used for
passive devices and signal distribution to retain the signal fidelity in the mm-wave regime. Transmission lines have many advantages over spiral inductors in designing passive devices at such high frequencies:

a) Transmission lines are scalable in length and they have the capability in realizing small inductors accurately for the given quasi-transverse electromagnetic mode of propagation. On the other hand, spiral inductors have some limitations in implementing inductors below 1nH.

b) Transmission lines have a well-defined ground return path that confines the field and reduces the magnetic and electric field coupling.

c) The scaling ability of the transmission lines simplifies developing compact models and generating efficient simulation techniques.

d) Transmission lines can provide passive devices with higher quality factors than spiral inductors. This quality factor is independent on the transmission line length.

e) Designing and modeling spiral inductors require a strong knowledge of the substrate profile.

Many researches at mm-wave frequencies have discussed different kinds of transmission lines as introduced below. This section discusses the major types briefly. This discussion covers the structure, advantages, and limitations of each kind.

1. Thin Film Microstrip (TFMS) transmission line.

Microstrip line is one of the most popular types of planar transmission lines, primarily because of its ability to be fabricated by photolithographic process and its
simplicity to integrate with other passive and active microwave devices. The geometry of a microstrip line is shown in Figure 3.1, where a conductor of width \( W \) is printed on a thin, grounded dielectric substrate of thickness \( d \) and relative permittivity \( \varepsilon_r \). Most designers use the upper metal layer to build the signal conductor of width \( W \), while use the next metal layer to build the underneath ground plane. Because of that, microstrip line cannot support a pure TEM wave since its field lines are contained within non homogenous dielectric regions. However, in most practical applications, the dielectric substrate is electrically very thin compared to the wavelength and so the fields are considered as quasi-TEM wave.

![Figure 3.1 Thin Film Microstrip Transmission Line](image)

Microstrip lines have many advantages. They are easy to be meandered, have intrinsic shielding, have reduced shunt loss, have isolated ground plane so it is less sensitive to substrate process details, can be accurately predicted by field simulators, and have no electric field penetration into the substrate. However, microstrip lines suffer from degraded inductive quality factor due to small distributed inductance because of the close
proximity of the ground, constant capacitive and inductive quality factor regardless of the geometry, and difficult routing of signal and power lines due to its lengthy layout.

2. Folded Microstrip (FMS) transmission line.

Folded Microstrip lines are microstrip lines in their structure, but their ending terminals are folded for shorter transmission lines [21]. The geometry of folded microstrip transmission line is shown in Figure 3.2. As shown in this figure, the performance of FMS lines is determined by two parameters: the line width $W$ and the spacing between the legs $S$. The two legs of TMS lines must bear minimum spacing. On the other hand, if the spacing goes below a minimum value, it causes mutual magnetic coupling that lowers the overall inductance values and degrades the performance of the transmission line. The dimensions of FMS line must be chosen to maximize the resistance, minimize the capacitance, and maintain enough small inductance to resonate with active devices capacitance connected with them. The characteristic of FMS lines are the same as TFMS ones except they have half of the length of the corresponding TFMS. Therefore, FMS lines reduce the layout size and solve the difficulty problem in routing signal and power lines.

![Figure 3.2 Folded Microstrip Transmission Line [21]](image-url)
3. Coplanar Waveguide (CPW) transmission line.

Coplanar waveguide is formed from one signal conductor surrounded by a pair of ground planes, all on the same plane which is usually the top metal layer as shown in Figure 3.3. In ideal case, the thickness of the dielectric underneath the CPW is infinite, but in practical cases, it is assumed this dielectric is thick enough so that EM field die out before they get out the substrate. Because of this structure, the signal line in the middle can support even or odd quasi-TEM modes depending on whether the E-fields in the two slots are in the opposite direction, or the same direction. CPW lines are designed by proper selection for the signal width $W$ and slot size (spacing between the signal and ground conductors) $G$ to meet the required specifications. The signal width controls the conductor loss, while the slot size controls the characteristic impedance and the tradeoff between the inductive and capacitive quality factors.

![Figure 3.3 Coplanar Waveguide Transmission Line](image)

CPW lines have many advantages over the other transmission lines. First, CPW lines can provide extremely high frequency response (100 GHz or more) since connecting to CPW does not entail any parasitic discontinuities in the ground plane. Second, CPW lines can achieve higher characteristic impedances. Third, CPW lines can double the quality factor and have reduced capacitive quality factor because of the substrate
coupling. Fourth, CPW lines are less sensitive to process variations since signal and ground conductors are on the same plane. Finally, CPW lines have predictable and constant characteristic impedance. However, CPW lines are suffering from problems such as: odd parasitic modes, importance to force the two ground planes to the same potential, and large area because of the presence of the two ground conductors.

4. Vertical-Ground-Plane (VGP) transmission line.

Vertical-Ground-Plane transmission lines [39] can be considered as a modified version of CPW lines. VGP lines consist of one signal line surrounded by two vertical ground planes as shown in Figure 3.4. VGP lines have the same properties as the CPW except the ground plane are vertically places. VGP lines use the available metal layers as the ground-plane layers and use the top thick metal layer for the signal conductor. With more metal layers in the new silicon technologies, VGP lines can be designed with smaller ground-plane width, higher isolation between adjacent lines, and smaller area for passive device while maintaining the advantages of the classical CPW lines. Therefore, VPG lines have compact meandering, wide range of characteristic impedances, and high isolation between nearby lines.

Figure 3.4: Vertical-Ground-Plane Transmission Line [39]
In this work the most popular transmission lines (microstrip and coplanar waveguide) have been used, modeled, and tested. These two kinds have been characterized, designed, taped out, and tested to validate the compact model proposed in this work in order to minimize the error between the real measurements and the simulated results without slowing the simulator.

3.2 Transmission Line Characterization

At high frequency, the conventional $z$-, $y$-, and ABCD-parameters are not used to characterize transmission lines due to the difficulty in realizing short circuit termination (because of the magnetically induced inductance) and open circuit termination (because of electric field fringing). However, S-parameters are considered the most accurate method to characterize the response of transmission lines in the mm-wave band. Signal propagation in transmission lines is based on solving the Telegraph equations. The standard solution form for these equations is determined by the propagation constant $\gamma$ and characteristic impedance $Z$ of the transmission line, which can be described by distributed transmission line parameters. A lossy transmission line (coplanar waveguide or microstrip line) can be modeled accurately using frequency variant infinitely small distributed RLCG sections, as shown in Figure 3.5.

![Figure 3.5 Per-unit length RLCG model of TL.](image)

48
where

\[ \begin{align*}
R(\omega) & = \text{the distributed series resistance per unit length of both conductors, in } \Omega/m \\
L(\omega) & = \text{the distributed series inductance per unit length for both conductors due to magnetic field around the wires and self-inductance, in H/m.} \\
C(\omega) & = \text{the substrate coupling and junction capacitance per unit length, in F/m.} \\
G(\omega) & = \text{the dielectric shunt conductance per unit length due to dielectric constant of the material separating the two conductors.}
\end{align*} \]

This distributed model describes the per unit length values rather than the lumped elements values used by most simulators such as SPICE. In addition to that, the main advantage of the distributed RLCG model is that it is fully compatible with all available circuit simulators. The S-parameter matrix for a lossy transmission line with parameters \( \gamma \) and \( Z \) in Zo impedance systems is described in Equation 3.1 [40].

\[
[S] = \frac{1}{D} \begin{bmatrix}
(Z^2 - Z_0^2) \sinh \gamma l & 2ZZ_0 \\
2ZZ & (Z^2 - Z_0^2) \sinh \gamma l
\end{bmatrix}
\]

(3.1)

where

\[
D = 2ZZ_0 \cosh \gamma l + (Z^2 + Z_0^2) \sinh \gamma l
\]

The S-parameter matrix can be converted into ABCD parameter as given in [41].
\[ A = \frac{(1 + S_{11})(1 - S_{22}) + S_{12} S_{21}}{2 S_{21}} \]
\[ B = Z_0 \frac{(1 + S_{11})(1 + S_{22}) - S_{12} S_{21}}{2 S_{21}} \]
\[ C = \frac{1}{Z_0} \frac{(1 - S_{11})(1 - S_{22}) - S_{12} S_{21}}{2 S_{21}} \]
\[ A = \frac{(1 - S_{11})(1 + S_{22}) + S_{12} S_{21}}{2 S_{21}} \]  
\[ \text{(3.2)} \]

Using Equations (3.1) – (3.2), the equivalent ABCD will be

\[ [ABCD] = \begin{bmatrix} \cosh \gamma l & Z \sinh \gamma l \\ \sinh \gamma l & \cosh \gamma l \end{bmatrix} \]
\[ \text{(3.3)} \]

Combine Equations (3.1) – (3.3), the propagation constant and characteristic impedance of TLs can be extracted from S-parameters as

\[ Z = Z_0 \sqrt{\frac{(1 + S_{11})^2 - S_{21}^2}{(1 - S_{11})^2 - S_{21}^2}} \]  
\[ \text{(3.4)} \]
\[ e^{-\gamma l} = \frac{1 - S_{11}^2 + S_{21}^2}{2 S_{21}^2} \pm \sqrt{\frac{(S_{11}^2 - S_{21}^2 + 1)^2 - (2 S_{21}^2)^2}{S_{21}^2}} \]  
\[ \text{(3.5)} \]

Knowing the characteristic impedance and the propagation delay, the per-unit length RLCG transmission line parameters are derived as: [42]

\[ Z = \sqrt{\frac{[R(\omega) + j \omega L(\omega)]}{[G(\omega) + j \omega C(\omega)]}} \]  
\[ \gamma = \sqrt{[R(\omega) + j \omega L(\omega)][G(\omega) + j \omega C(\omega)]} \]  
\[ \text{(3.6)} \]
\[ \text{(3.7)} \]
Then
\[ R(\omega) = \text{Re}(\gamma Z) \]
\[ L(\omega) = \frac{\text{Im}(\gamma Z)}{\omega} \]
\[ G(\omega) = \text{Re}\left(\frac{\gamma}{Z}\right) \]
\[ C(\omega) = \frac{\text{Im}\left(\frac{\gamma}{Z}\right)}{\omega} \]  
(3.8)

These equations are used in the following sections to derive the distributed model using the S-parameters for both microstrip and coplanar waveguide transmission lines. Next sections discus the calibration, de-embedding, and extraction methods used to generate the compact model for any transmission line.

### 3.3 De-embedding Transmission Line Measurements and Momentum Simulator Calibration

To calibrate the Momentum simulator for a given CMOS process and to verify the accuracy of the extracted distributed model, test structures are designed and fabricated in TSMC’s 0.18\(\mu\)m CMOS technology. Two types of transmission lines have been fabricated: CPW and microstrip lines. All CPWs are designed using the top thick metal (metal 6) with signal plane width of 5.8\(\mu\)m, ground plane width of 30 \(\mu\)m, and signal-ground spacing of 3 \(\mu\)m. CPW lengths vary between 80 \(\mu\)m to 400 \(\mu\)m. All microstrip lines use metal 6 for the signal plane and metal 5 for the ground plane with signal plane width of 5 \(\mu\)m and ground plane width of 30 \(\mu\)m. Microstrip line length varies between 120 \(\mu\)m to 440 \(\mu\)m. Access to the transmission line conductor is provided by using a 70 x 70 \(\mu\)m RF signal pad constructed on metal 6 and 5. On either side of each signal, 70 x 70
µm pads are used to contact the ground plane using ground-signal-ground (GSG) device probes of 150 µm pitch as shown in Figure 3.6.

Figure 3.6 Die photograph A) CPW transmission lines B) Microstrip transmission lines and C) De-embedding structures

Before developing a mathematical process of de-embedding, it is important to represent the test fixture and the Device Under Test (DUT) in a convenient way. The fixture and DUT can be represented as three separate two-port networks as shown in the signal flow graph Figure 3.7. The test fixture is divided into two halves to represent the coaxial to non-coaxial interfaces on each side of the DUT. Fixture L and fixture R are used to designate the left-hand and right-hand sides of the fixture respectively. To directly multiply the matrices of the three networks, the S-parameter matrices need to be converted into scattering transfer matrices (T-parameter matrices) as follows:
Since the test fixture and DUT are defined as three cascaded networks, the equivalent T-parameter network can be obtained by easily multiplying their respective T-parameter networks, TL, TDUT and TR. This simple matrix equation can be written in this form:

\[
\begin{bmatrix}
    b_1 \\
    a_1 \\
    S_{11} & S_{12} \\
    S_{21} & S_{22} \\
    T_{11} & T_{12} \\
    T_{21} & T_{22}
\end{bmatrix}
= \begin{bmatrix}
    T_{12} \\
    T_{22} \\
    1 \\
    T_{22} \\
    T_{11}T_{22} - T_{12}T_{21} \\
    T_{22}
\end{bmatrix}
\begin{bmatrix}
    a_2 \\
    b_2 \\
    S_{11} \\
    S_{21} \\
    T_{11} \\
    T_{21}
\end{bmatrix}
\]

\[
\begin{bmatrix}
    T_{11} & T_{12} \\
    T_{21} & T_{22}
\end{bmatrix}
= \begin{bmatrix}
    -S_{11}S_{22} - S_{12}S_{21} & S_{11} \\
    S_{21} & S_{21} \\
    S_{22} & 1 \\
    S_{21}
\end{bmatrix}
\]

Figure 3.7 Signal flow graph representing test fixtures and DUT

This represents the T-parameters of the test fixture and DUT when measured by the VNA at the measurement plane. If a matrix determinate is not equal to zero, then the matrix has
an inverse and can be used to de-embed the two sides of the fixture TL and TR, and
gather the information from the DUT. Multiply each side of the measured results by the
inverse T-parameter matrix of the fixture and yield the T-parameter for the DUT only.
Then the T-parameter matrix can be converted back to the desired S-parameter matrix
using the above equations.

\[
\begin{bmatrix}
T_{a}^{-1} & T_{MEASURED} \times T_{b}^{-1} =
\end{bmatrix}
\begin{bmatrix}
T_{a}^{-1} & T_{DUT} \times T_{b}^{-1} =
\end{bmatrix}
\begin{bmatrix}
T_{DUT}
\end{bmatrix}
\]

The above process is typically implemented after the measurements are captured from the
VNA.

Before mathematically de-embed the test fixture from the measurements, each
fixture half needs to be modeled using the S-parameter of T-parameter. Creating an
accurate model of the test fixture is considered as the most difficult task in the whole
process of de-embedding. Simulation tools such as Advanced Design System (ADS) or
High Frequency Structure Simulator (HFSS) can help in creating the fixture model. Then
observing the physical structure of the test fixture and the measurements made on the
fixture can be used to optimize the model in an iterative manner. Simpler models are used
in firmware of many vector network analyzers to directly perform the de-embedding
without the requirement of the T-parameter matrix mathematics.

The most accurate model is the one that consider the complex effects of the coax-
to-non-coaxial transitions as well as the fixture losses and impedance differences. This
model is the hardest one since it includes all of the non-linear effects such as dispersion,
radiation and coupling that happen in the fixture. One way to determine the model is by
using a combination of measurements of known devices placed in fixture and a computer model whose values are optimized to the measurements. In this way, the coax-to-transmission-line transition is modeled by a lumped series inductance and shunt capacitance. The value for the inductance and capacitance can be calculated using the measured results. An EM simulator can be used to build the same lumped element component placed on each side to model the two test fixture transactions. This model requires an accurate value for dielectric constant and loss tangent for the substrate material used. The measured four S-parameters can be imported into the EM simulator (such as ADS) over the GPIB to optimize the values for inductance and capacitance until a good fit is obtained between measurements and the simulated results. All four S-parameters should be optimized and compared to the measured S-parameters to verify the accuracy of the model values. This lumped element model is only valid over a small frequency range due to the nonlinear effects in the transition. Therefore for broadband operation, an improved model must be implemented to incorporate the nonlinear behavior of the measured S-parameters as a function of frequency, or these lumped elements need to be modified over different frequency ranges. These optimized lumped element parameters will be saved for use by the de-embedding algorithm.

There are two main ways to implement the de-embedding process. The first technique (the static approach) uses measured data from a network analyzer and process the data using the T-parameter matrix calculations. The second technique (real-time approach) uses the network analyzer to directly perform the de-embedding calculations by modifying the calibration error terms in the analyzer’s memory. The static approach
for the process of de-embedding the test fixture includes five steps: 1) Create a mathematical model of the test fixture using S or T-parameter. 2) Calibrate the network analyzer using a standard coaxial calibration kit and measure the S-parameters of the device and fixture together. 3) Convert the measured S-parameters to T-parameters. 4) Apply the de-embedding equation to the measured T-parameters

\[ T_{DUT} = [T_A]^{-1}T_{MEASURED} [T_B]^{-1} \]. 5) Convert the final T-parameters back to S-parameters and then display the results which represent the S-parameters of the device only after removing the test fixture effects. On the other hand, the real-time approach incorporates the test fixture S-parameter model into the calibration errors terms in the vector analyzer. Then the analyzer is performing all the de-embedding calculations which allow the users to view real-time measurements of the DUT without the effects of the test fixture. Most network analyzers are capable of performing some modification to the error terms directly from the front panel which includes port extension and modifying the calibration “thru” definition.

The first stage of de-embedding is done during network analyzer calibration where the instrument measures actual, well-defined standards such as the open, short, load, thru, and compares the measurements to ideal models for each standard. The differences between the measurements and the models are used to compute error terms contained within the measurement setup. These error terms are used to correct the actual measurements of the device under test (DUT). A reference or calibration plane is created at the point where the standards are connected. Three offset characteristics for each standard are included in the calibration kit definition. They are Offset Delay, Offset Loss,
and Offset Impedance (Zo). These characteristics are used to accurately model each standard in order to establish a reference plane for each of the test ports. Fixture de-embedding can be performed by adjusting the calibration kit definition table to include the test fixture effects. The determined error terms during the coaxial calibration process can include some of the fixture characteristics. By the end of the calibration cycle, the analyzer will remove mathematically the three losses of the fixture.

To remove the systematic error of the test equipment, twelve–term error model is used inside the vector network analyzer (VNA). This is done during the VNA calibration where the system measures the magnitude and phase responses of known devices and compares the measurements with actual device characteristics. These results are used to characterize the instrument and to remove the system errors effectively from the measured data of a DUT.

The de-embedding process begins by creating a model of the test fixture on either side of the DUT. The second step is to perform a standard coaxial two-port calibration type such as SOLT (Short, Open, Load, Thru) or TRL (Thru, Reflect, Line). Then these calibration results are saved into the instrument memory. The third step is to download the twelve error terms into a computer and then modified them using the model(s) for each side of the test fixture. The modified twelve error terms are then placed back into the analyzer’s calibration memory. By the end of this step, the VNA displays the de-embedded response of the DUT including the pads and pad-TL interconnection effect. The last step is to de-embed the pads and the pad-TL interconnection using standard SOT (Open, Short, Thru) two-port calibration technique.
To de-embed the pads and metal line connections between the pads and the actual transmission lines, SOT (Short-Open-Thru) structures are fabricated on the same die. Figure 3.6 shows the die micrograph for the TL test structures and de-embedding structures. A Cascade Microtech Probe Station and an E8361 Agilent PNA Network Analyzer, which operates from 100 MHz to 67 GHz, are used to measure the S-parameters. The measured data is transmitted through GBIP and WinCal XE software to be used for Momentum calibration and model extraction. WinCal XE software comes with a vast library of functions that has the ability to mathematically manipulate S-parameter data by employing one or more mathematical functions. This post processing data manipulation is done inside WinCal XE using Math Scratch Pad (MSP) function which has the ability to utilize multiple functions, to save processed data to data items list, and to use the expression editor. The work flow of MSP starts by entering S-parameter data items into the Scratch Pad as variables, then processing these variables using functions with the Scratch Pad, and last saving the new post processed data items back to the Data Item List to display. Therefore, the open structure has been used in MSP to determine the pad and external interconnect parasitic capacitance, while the short structure has been used to extract the pad and interconnect parasitic inductance. The parasitic pad-interconnect capacitance varies from 23 fF to 32.9 fF while the parasitic pad-interconnect inductance varies from 63 pH to 118 pH over 20–65 GHz frequency band. The parasitic pad interconnects are de-embedded from all measured data yielding the actual S-parameters for the transmission lines.

Generating the TL compact model starts by using the measured data to calibrate
the Momentum simulator. These results need to be in a good agreement between measured and simulated results prior to subsequent simulations. The calibration process adjusts the physical technology parameters used in Momentum (such as substrate resistivity, metals conductivity, dielectric constants, tangent loss, etc). This is done by defining the substrate inside Momentum using the dimensions of each dielectric and metal layers provided by technology providers and then using measured data to define the physical parameters. Test structures comprised of CPW of length 80 µm and Microstrip line of length 120 µm have been used for the calibration purpose. Figure 3.8 and 3.9 show the measured and the calibrated Momentum S-parameters for CPW and microstrip lines. As shown, an excellent agreement is observed between the measured and the simulated S-parameters over the full range of frequencies.

Figure 3.8 Measured and calibrated S11 for CPW and Microstrip line.
3.4 Circuit Model Extraction

The process of characterizing and extracting the high frequency response of transmission lines is divided into three major steps, as shown in Figure 3.10. In the first step, the measured S-parameters from the test structures are used to calibrate the Momentum simulator for the given process technology. This step requires de-embedding the pads parasitic from the measured S-parameters of the TLs. In the second step, Momentum is used to calculate the TL characteristic impedance and propagation constant for a certain TL configuration using the calibrated S-parameters and then extracts the corresponding per-unit RLCG parameters. In the last step, a segmented frequency variant distributed RLCG model is generated for variable TL lengths/widths by cascading
After calibrating the Momentum simulator, then it can be used to generate S-parameters for a smaller set of transmission line configurations with variable widths and lengths. Subsequently, these S-parameters are used to extract the corresponding TL characteristic impedance and propagation delay and the per-unit length RLCG parameters. The RLCG parameters are extracted for CPWs of signal plane width of 5.8µm, ground plane width of 30µm, and signal-ground spacing of 3 µm, and for the microstrip lines of signal plane width of 5µm and ground plane width of 30µm. Figures 3.11 – 3.14 depicts the variations of the RLCG parameters per 100 µm unit length over the frequency band of 20 -65 GHz. As shown in the figures, the majority of line change across the frequency band is attributed to variations in R and G, which is expected due to the skin and dielectric loss variations over the frequency range.
Figure 3.11 Extracted Series Inductance per 100μm

Figure 3.12 Extracted Series Resistance per 100 μm
Figure 3.13 Extracted shunt capacitance per 100µm.

Figure 3.14 Extracted shunt conductance per 100µm.
As notices from these figures, the per 100 µm unit series inductance varies between 32 pH to 38 pH for CPW and between 45 pH to 55 pH for Microstrip line, the per 100 µm unit series resistance varies between 7Ω to 25Ω for CPW and between 20Ω to 130Ω for Microstrip line, the per 100 µm unit shunt capacitance varies between 52 fF to 45 fF for CPW and between 55 fF to 45 fF for Microstrip line, and the per 100 µm unit shunt conductance varies between 0.2 mS to 0.8 mS for CPW and between 0.2 mS to 3.6 mS for Microstrip line.

### 3.5 Verification Results

To verify the scalability and the accuracy of the generated RLCG distributed model, seven CPWs and seven Microstrip lines of variable length are built for testing and verification purposes, as shown in the above die micrograph. CPWs lengths varies between 80 µm to 400 µm while Microstrip lines lengths varies between 120 µm to 440 µm. Except for the variable TL lengths, the verification structures were based on the same TL configurations used in extracting the per-unit RLCG parameters described in section 3.4. Specifically, all CPWs have the same signal plane widths, ground plane width and signal-ground spacing and all microstrip lines have the widths as described in section 3.4. The extracted RLCG parameters are then used to determine the lumped model for a considerably small single segment. The lumped model for this segment is built using Rs = R(ω).SL, Ls = L(ω).SL, Cp = C(ω).SL, and Gp = G(ω). SL for each segment where R(ω), L(ω), C(ω), and G(ω) are the per unit RLCG calculated in section 3.4 and SL is the segment length. The distributed circuit model for each TL is constructed using cascaded stages of this small segment of length SL to obtain the required TL.
length. Figures 3.15 – 3.18 depicts the S-parameters for CPW320 of length 320 µm where S-parameters using RLCG model refers to the simulated results using cascaded RLCG segments, and S-parameters using MoM simulator refers to the simulated results for the CPW320 using the calibrated Momentum simulator. The displayed results indicate an excellent agreement between the measured and simulated S-parameters (both using the calibrated Momentum and using segmented RLCG models) with maximum deviation in S11 magnitude of 9.22%, maximum S11 phase difference of 5.6%, maximum deviation in S21 magnitude of 10.1%, and maximum S21 phase difference of 6.6%.

Figure 3.15 Simulated MoM, measured, and segmented S11(dB)
Figure 3.16 Simulated MoM, measured, and segmented S11(phase)

Figure 3.17 Simulated MoM, measured, and segmented S21(dB)
Figure 3.18 Simulated MoM, measured, and segmented S21(phase)

3.6 Conclusions

This chapter presents a simple and efficient technique to extract a frequency dependent distributed RLCG model for transmission lines in a lossy silicon substrate. The proposed technique relies on calibrated S-parameters models. Comparison of measured data versus the distributed model confirms the accuracy of the extracted model from 20-65 GHz. The proposed extraction method improves the accuracy of characterizing TLs compared to commonly used models based on standard CMOS technology files and reduces the required simulation time compared to full-wave electromagnetic simulators. The generated models are fully compatible with HSPICE and SPECTRE-RF simulators.
CHAPTER 4

DESIGN MM-WAVE LNAs

The LNA (Low Noise Amplifier) is the first gain stage in most wireless receiver path. It is supposed to provide signal amplification without degrading the signal-to-noise ratio (SNR). Its noise figure determines the lower bond of the entire system and it directly adds to the system noise figure. Since the signal power level at the antenna is very low and it degrades further due to physical obstruction. Hence, good and enough amplification is a critical factor to determine a robust performance in obstructed environment. Therefore, the LNA needs to be designed to meet certain tight specifications in order to achieve high performance for the whole receiver. Many parameters are used to measure the performance of the LNA: gain, input/output matching or return loss, noise figure, reverse isolation, linearity, and stability. CMOS FETs were considered slow, noisy devices up to about a decade ago. The scaling in CMOS technology has improved their performance dramatically, but the excess thermal noise exhibited by sub-micron CMOS devices constructs one possible threat to low noise operation.

Different architectures have been used to build LNA at low frequencies such common-source, cascaded common-gate and common-source, folded cascode common-source, and inductive source degeneration. The most famous topology used in designing LNA at low frequency is inductive cascode common source degeneration. More details
about these structures advantages and drawbacks will be discussed in details in the following sections. However, the most critical question is which of these architectures are suitable to design 60 GHz LNAs.

To provide some background, section 4.1 presents a review of LNA fundamentals and key parameter need to be considered in its design. A review of the most common LNA architectures used for low frequency applications is discussed in section 4.2. Section 4.3 discusses the most suitable mm-wave LNAs and presents a through mathematical treatment of the LNA architectures chosen in this work. The simulation results of the selected mm-wave LNAs are reported in section 4.4. At the end, section 4.5 summarizes the performance of the designed mm-wave LNAs and compared them with the present related work available in literature.

4.1 LNA Fundamentals

As we discussed before, LNA is the first amplification block in the receiver path, so its noise figure directly adds to that of the system. Many parameters are used to measure the performance of the LNA. The following subsections discuss the critical parameters in designing the LNA.

4.1.1 Gain

The minimum gain of the LNA in a heterodyne architecture is determined by three parameters: the loss of the image-reject filter, the noise figure, and IP3 of the mixer. LNA must provide its gain while driving 50Ω. While in the homodyne architectures, the gain and drive requirements of the LNA are relaxed because of the absence of the image-
reject filter. However, it is desirable to maximize the gain in the RF range to overcome the flicker noise problem in the baseband circuits. Gain of the LNA plays an important role in minimizing the noise effect of the following stages for the overall system as explained in Friis equation which indicates that the noise contributed by each stage in cascaded blocks decreases as the gain of the preceding stages increases. This implies the first stage (LNA) in a cascade is the most critical block since the noise figures of the following blocks will be scaled down by the gain of that LNA.

4.1.2 Input/Output Matching

A good input match is more critical when a preselect filter precedes the LNA since these filter are so sensitive to the quality of their termination impedance. The bandpass filter following the antenna is usually designed to be in various transceiver systems and must therefore operate with standard termination impedance, typically 50Ω. Therefore, LNA is designed to have 50Ω resistive input impedance. The passband and stopband characteristics of the filter may exhibit considerable loss and ripples if the load and source impedances seen by the filter deviate from 50Ω significantly. The quality of the input matching is expressed by the input "return loss" which is defined as $20\log |\Gamma|$. In heterodyne architectures, the LNA output impedance must equal to 50Ω too, so as to drive the image-reject filter with minimum loss and ripple.

Matching the LNA input impedance to 50Ω is a difficult task especially in common-source configurations where the input impedance is determined by the gate-drain capacitance that connects the output to the input. At very high frequency, this capacitor behaves as a short circuit that connects the gate and the drain terminals together and
creates a high input impedance roughly equal to $R_i/(1/g_m)$. In principle, this configuration can provide an input resistance equal to 50Ω, but the drawback of this solution is the relatively low voltage gain at high frequency due to the bandwidth limitation at the output node.

Different techniques can be used to achieve this goal in literature, but the most important key here is to exhibit a 50Ω input resistance without the thermal noise of a 50Ω resistor. The first technique uses resistive termination of the input port by 50Ω impedance, but this solution has a severe effect on the amplifier’s noise figure which makes this architecture unattractive for good desired terminations. The second technique uses common-gate or common-base as an input stage. Common-gate or common-base stage has an input impedance of $1/(g_{m} + g_{mb})$ which can be designed to give 50Ω termination and the input capacitance can be nulled by means of an external inductor. The noise contribution of this technique depends on the used devices where the transconductance of the input transistors cannot be arbitrary high and imposing a lower limit on the achieved noise figure. Also, for short-channel MOS devices, the hot electron phenomena increase the MOS thermal noise. Therefore, the minimum theoretically achievable noise figures tend to be a little bit higher than the common-source architecture noise figures. The third technique uses resistive shunt and resistive feedback to set the input and output impedance of the LNA. This technique has two major difficulties. The first one is the substantial increment in the noise due to the feedback. The second difficulty is the instability problem due to the total phase shift around the loop under certain source and load impedances. In addition to these difficulties, amplifiers use this
technique often have significantly higher power dissipation due to the fact that these amplifiers are naturally broadband and no L-C tuning circuits are used to reduce the power consumption. Shunt-series architecture usually requires reasonable on-chip resistors which are generally not available in CMOS technology. The last technique uses inductive source or emitter degeneration. The proper choice of $g_m$ and the gate-source capacitance of the input device and the degeneration inductor can set the real part of the input impedance to 50Ω. In order to maintain the resonance frequency at input port to be the same as the desired frequency, an additional inductor can be added at the gate terminal. This gate inductor can be used with the degeneration inductor and the gate-source capacitor to tune the LNA to the desired operating frequency while maintain the input matching to be 50Ω at this frequency. At very high frequency, the degeneration and gate inductor values become comparable in size with the inductance of the ground bond wire. To optimize the LNA performance in this case, multiple bonds are used and accurate modeling of the wire inductance is required. This technique is considered the most common used technique to design a narrowband LNA. It offers the possibility to achieve the required matching with the best noise performance.

### 4.1.3 Noise

The noise of the LNA is considered as an important and critical factor in determining its overall performance since the LNA is the first block in the receiver path and its noise adds to the system. The noise is measured using input referred noise or noise figure parameters. The latter one is the most commonly used parameter for the LNA. Noise figure is defined as the ratio of the signal-to-noise ratio at the input to the signal-to-noise
ratio at the output. Noise figure is a measure of how much the SNR degrades as the signal passes through a certain circuit. The dominant noise source in CMOS devices is the channel thermal noise which is usually modeled as a shunt current source at the output circuit of the device. Channel thermal noise is a white noise with a power spectral density of $4kT\gamma/g_m$ where $\gamma$ for short-channel devices is higher than 1. This noise increases in the presence of hot electrons in the channel.

The additional source of noise in CMOS devices is the gate resistance noise. This noise is modeled as a series resistance in the gate circuit. The distributed gate resistance can be represented by an effective gate resistance of $R_g = [R\Box W]/[3n^2L]$ where $R\Box$ is the sheet resistance of the polysilicon, $W$ and $L$ are the gate length and width respectively, $n$ is the number of gate fingers, and $1/3$ comes from a distributed analysis of the gate. If the gate is contacted at both sides, the factor $1/3$ reduces to $1/12$. The gate resistance effect is minimized in CMOS by using interdigitation technique and by using silicided CMOS processes which reduce the sheet resistance $R\Box$. The third noise source in CMOS is the back-gate epitaxial resistance which increases the value of $\gamma$ and the drain noise. The epitaxial layer can be modeled as a resistance in series with the bulk terminal of the device. Its noise contribution can be combined with the drain current noise and produces a total current noise of $4kT\gamma_{eff}/g_m$ where $\gamma_{eff} = \gamma + (g_{mb}^2R_{epi})g_m$. The $(g_{mb}^2R_{epi})g_m$ varies between $0.09 - 0.2$. With closely-spaced substrate contacts, this number will get reduced. Therefore, the epitaxial resistance is of secondary importance.
4.1.4 Linearity

Nonlinear system can create many problems at its output. Nonlinear system exhibits higher-order frequency components or harmonics of its input fundamental frequency. The amplitude of these harmonic components grows much faster than the fundamental one. The even-order harmonics can be minimized or reduced by having fully differential circuits. The small-signal gain of the LNA or any amplifier is obtained with the assumption that harmonics are negligible. However, the gain begins to vary when the signal amplitude increases. This causes a variation in the small-signal gain with the input level and causes output saturation or zero gain for sufficiently high input levels. Linearity in any RF blocks is measured by two parameters: 1-dB compression point (P$_{1\text{-}dB}$) and third intercept point IP3 point. The P$_{1\text{-}dB}$ measures the compression or saturation performance of the circuit as a function of its input and is defined as the input signal level that causes the small-signal gain to drop by 1 dB. Compressive circuits show severe characteristics in the presence of strong interferers. The desired weak signal might experience desensitization effect or blocking. The IP$_3$ point is used to measure the intermodulation distortion performance of the circuit in the presence of two-tone input signal. Intermodulation affects the performance of RF systems if the systems experience a weak desired signal accompanied by two strong interferers. These interferers might experience third-order intermodulation products fall in the band of interest and corrupt the desired component. The magnitude of the third-order intermodulation products grows at faster rate (three times the main component rate). If the LNA is designed to be fully differential, then the third-order intermodulation product plays an essential role in
determining the gain and the linearity of this block since the second-order products are vanished. For cascaded blocks, the linearity of the latter stages is more critical because the $IP_3$ of each stage is effectively scaled down by the total gain of the preceding stages. Therefore, the LNA linearity is not as critical as the mixer linearity, but its linearity is considered as an advantage to the overall system.

### 4.1.5 Stability

The stability of the LNA is an important factor especially in the presence of feedback paths from the output to the input. In such cases, the LNA may become unstable for certain combinations of source and load impedances. Therefore, a nominally stable LNA circuit may oscillate under extreme manufacturing variation or at unexpectedly high or low frequencies. Stern stability factor is the most common parameter used to characterize the stability and is defines as \[25\]

\[
K = \frac{1 + |\Delta|^2 - |S_{11}|^2 - |S_{22}|^2}{2 |S_{21}| |S_{12}|}
\]

\[
\Delta = S_{11} S_{22} - S_{21} S_{12}
\]

For unconditionally stable circuit (i.e., it does not oscillate under any combination of source and load impedance), $K$ needs to be greater than 1, while $\Delta$ needs to be less than 1. The above equation indicates that stability improves as the reverse isolation of the circuit increases (or $S_{12}$ decreases). This condition can be accomplished by neutralizing the input-output capacitive path using an inductor to resonate this capacitance at the frequency of interest, but the parasitic capacitance of the inductor in addition with the coupling capacitance load the input and output nodes. The best solution to achieve this
goal uses the cascode configuration. The cascode transistor isolates the input node from the output node at the cost of some extra noise figure due to the additional cascode transistor. If the LNA load impedance is well defined, then the stability of the LNA can be checked and testified by simply ensuring that the real part of the input impedance is positive at all frequencies.

LNA may become unstable for different reasons. The first reason as discussed above is due the capacitive feedback path between the input and output nodes. The second reason is due to the ac ground and supply loops resulting from bond wire inductance. At higher frequency of operation, the few nanohenries of inductance from bond wires may cause a coupling between stages through the ground node. Careful design and layout of LNA circuit and using an accurate package modeling play an important role in controlling the stability due to this factor.

### 4.1.6 Reverse Isolation

LNA reverse isolation is very essential in improving the performance of the following stage; Mixer. LNA reverse isolation define the amount of the LO signal that leaks from the mixer to the antenna and mixed again in the mixer. Different reasons cause this leakage such as the capacitive paths, the substrate coupling, and the bond wire coupling. The leakage in homodyne receivers is attenuated by the LNA reverse characteristic only. Therefore, LNA must exhibit tuff requirement on its reverse isolation. However, the leakage in heterodyne receivers is attenuated by three more blocks besides the LNA. The leakage in such receivers is significantly suppressed by a high first IF, the image-reject filter, and the front-end duplexer. This relaxes the reverse isolation requirement on the
first stage amplifications stage; LNA.

4.1.7 Power Consumption

All portable wireless devices are operating from batteries. The weight and the size of portable devices is determined by the amount of batteries which is directly impacted by the power dissipation by the electronic circuits. The cost of providing power and the associated cooling have resulted in significant interest in power reduction. Recently, many researches present different techniques to reduce power dissipations at the four levels of CMOS system design: technology, circuits, architectures, and algorithms.

4.2 Low-Frequency LNA Architectures

Up to a decade ago, CMOS FETs were considered slow, noisy devices. The scaling in CMOS technologies shows a dramatic improvement in their performance especially in their noise figure numbers. The two major sources of the noise in CMOS are the generated noise in the channel and the thermal noise of the gate resistance. The channel noise is reduced by the scaling in CMOS technologies. However, the gate resistance is minimized by layout out the transistor as a parallel combination of many narrower devices (multi-fingers). In addition, MOS devices are more linear than their counterpart bipolar transistors.

Each basic LNA configurations have advantages and limitations. This section discusses these basic configurations to design LNAs at low frequency of operation, while the next section discusses the challenges in using these configurations at mm-wave frequency band and suggests some modifications on these circuits to overcome the
present limitations at these high frequencies. The first topology is the common-source (CS) LNA as shown in Figure 4.1. This circuit has relatively low gain since MOS devices has typically low transconductance. This causes a significant influence of its load resistance noise and the noise of following stages on the overall system noise performance. The small-signal gain of common-source stage depends on both the parasitics of the input device and the load impedance. Increasing the CS gain can be achieved by either increase the input device transconductance or use large load impedance. For larger transconductance, either the input device size or the bias current needs to be increased. Increasing the input device size leads to a higher input capacitance which attenuates the input signal and magnifies the noise of the load impedance and the input device. Larger biasing current degrades the noise and power consumption performance. Increasing the load impedance can be achieved by replacing the passive load with an active current source. Replacing the resistive load with a current source can increase the common-source LNA gain, but this current source itself contributes noise. In summary, to achieve higher gain, minimize the input referred noise, and drive a 50Ω load, more stages may be required. Thus, the first stage must have a preceding matching network for 50Ω input matching, must drive the second stage input capacitance, and must have a sufficient gain to minimize the noise contribution of the following stages. The next stages need to be designed to increase the overall LNA gain while driving either a 50Ω load or the next stage input load.
Figure 4.1 Common-Source LNA

The second basic LNA configuration is the common-gate (CG) topology shown in Figure 4.2. Common-gate LNA has four advantages: simpler input matching, higher linearity, better stability, and greater reverse isolation. Firstly, the input impedance of CG is approximately $1/(g_m + g_{mb})$ which is different from the capacitive input impedance of the CS LNA. To match CG circuit to 50Ω, simply equalize its effective transconductance $1/(g_m + g_{mb})$ to 50Ω. This input impedance is frequency independent which makes CG configuration suitable for broadband applications. Secondly, Common-gate LNA exhibits better linearity performance since the source resistance linearizes the input-output characteristic by softening the drain current excursions. Thirdly, CG also improves the reverse isolation of the LNA by lowering the LO leakage produced by the following mixer. Lastly, CG improves the stability of the circuit by minimizing the feedback from the output to the input. The primary drawbacks of common-gate stage are the relatively high noise figure and the difficulty to increase the amplifier gain while maintaining a 50Ω input matching.
The most famous configuration used in designing LNA at low frequency is the inductive degeneration cascode common source shown in Figure 4.3. The degeneration inductor plays two important roles: conjugate matching of the CS input capacitance and linearizing tool of the circuit. The input impedance of the degenerated CS LNA is expressed as:

\[ Z_{in} = \frac{g_m}{C_{gs}} L_s + L_s s + \frac{1}{C_{gs} s} \]

The second and third terms can be equalized to obtain a real input impedance equal to the first term only. The degeneration inductance \( L_s \) can be set to generate 50Ω input impedance from the first term of \( Z_{in} \) expression, but setting \( L_s \) to meet the input matching requirement might contradict with the input resonant frequency generated by equalizing the second and third terms in \( Z_{in} \). Therefore, another element needs to be added to the circuit to control the resonance frequency without effecting or degrading the input matching performance. This goal can be achieved by adding an additional inductor \( L_g \) at
the gate terminal. In this case the input impedance is modified into:

\[ Z_{in} = \frac{g_m L_s}{C_{gs}} + (L_s + L_g) s + \frac{1}{C_{gs} s} \]

where \( g_m \) is used to control the gain of the LNA, \( L_s \) is used to control the input matching, and \( L_g \) is used to set the resonance frequency independently.

![Inductive Degeneration Cascode Common-Source LNA](image)

Figure 4.3 Inductive Degeneration Cascode Common-Source LNA [21]

The two transistors in this architecture share the same bias current and save power through the use of the same bias current. The cascade transistor plays two roles in this architecture: Firstly, it increases the reverse isolation of the LNA by lowering the LO leakage. Secondly, it improves the stability of the circuit. The cascode stage helps the LNA to drive a 50 \( \Omega \) load. Cascode topologies provide low noise figure, good input matching, and high reverse isolation.

### 4.3 Proposed Architectures for mm-wave LNAs

In general, the design of the LNA is very critical in any receiver since it needs to meet tight specifications in order not to degrade the performance of the whole receiver.
Furthermore, at mm-wave and sub-THz frequencies, the LNA design faces other unique challenges. One of these challenges is the active device models, where a number of intrinsic or extrinsic device components that are not well captured at low frequencies have considerable effect on the performance of the device at higher frequencies. Small device interconnect inductance, resistance, and capacitance introduce additional losses and dispersion effects that are not captured by low frequency models. Therefore, a detailed full-wave electromagnetic simulation is required for an accurate prediction of these losses at the cost of lengthy and potentially non-converging simulation.

The second challenge in designing CMOS LNAs at mm-wave frequencies is in how to accurately model the passive devices. Compact and low loss passive components in silicon processes are required to counteract the induced losses of interconnects and conductive silicon substrate at high frequency. At sub-THz frequencies, the size of inductors for matching networks and resonators become increasingly small (on order of 30 – 300 pH). Typically, nanometer CMOS processes (e.g. 90nm, 65nm, or 45nm) can provide such small inductances using lumped spiral inductors as part of the process design kit. Alternatively, transmission lines (microstrip or coplanar waveguide) can be used to precisely implement these small inductors. Efficient and accurate modeling of these spiral inductors or transmission lines is one of the most challenging tasks in designing mm-wave radios. Two methods can be used to perform the model extraction. The first one is using electromagnetic simulation based on finite size physical layout (e.g. Ansoft HFSS). This simulator provides the most accurate results at the expense of long simulation time. The second method is using electrical models provided by Aglient ADS.
Electrical models in ADS have acceptable and scalable models for passives, especially if Momentum is used to extract the parasitic parameters. Momentum also has other advantages such as easy method of extraction, easy integration in commonly-used simulators and optimizers, less simulation time, and well defined parameters to model passive elements. In this work, all inductors are designed, simulated and modeled using Method of Moments (MoM) in ADS. The output model files are imported under cadence environment to simulate all passive components in the LNAs.

The third challenge in designing LNAs, is the choice of the active device sizes and the LNA architecture. The device $f_T$ and $f_{\text{max}}$ are highly affected by the size and the biasing of the active devices. For maximum achievable $f_T$ and $f_{\text{max}}$, different simulations are performed to select the optimum input transistor device size and biasing voltages.

At frequencies well below the $f_T$ of transistor, cascode topologies provide a low noise figure, good input matching, and high reverse isolation. On the other hand, at high frequency, the pole at the cascade node shunts a considerable portion of the RF current to ground. This causes lowering the gain and raising the noise contributed by the cascade device. Furthermore, the circuit is so sensitive to package parasitics due to the small required degeneration and gate series inductances. Therefore, either other topologies are required to build 60 GHz LNA or some modifications need to be introduced to the inductive source degeneration LNA to overcome the above limitations.

One proposed solution to design a 60 GHz LNA uses a single transistor as a common-gate stage before voltage amplification [22] – [24] as shown in Figure 4.4. The common-gate stage degrades the input matching and increases the noise figure unless large devices
are used. To solve this problem, a resonator is built at the source to cancel the generated capacitance at this node. Similarly, the output node must resonate to cancel the capacitance seen at the transistor drain and introduced by the next stage. The inductor \( L_1 \) is used to resonate with the gate-source capacitance and the source-bulk capacitance of \( M_1 \) and the pad capacitance. While, the inductor \( L_2 \) is used to resonate with the total capacitance at node \( X \). Transistor \( M_2 \) is used to provide additional gain stage and to drive the subsequent mixer stage.

![Diagram of a two-stage CG-CS LNA](image)

Figure 4.4 Two stage CG – CS LNA [22]

The reported simulated results of this LNA show a noise figure of 4.5 dB, voltage gain of 12 dB, and power consumption of 4 mA with 1.2 V power supply. However the measured results show a higher noise figure by about 56 % and a lower gain by about 25 % than expected. The sources of these discrepancies between simulated and measured results will be discussed at the end of this chapter.

Another way to build a 60 GHz LNA uses multi cascade common-source stages with input, output, and interstage reactive matching [16] as shown in Figure 4.5. Common-source configuration shows the best noise performance especially if the operating frequency is close to \( f_T \). The major concern of common-source 60 GHz LNA is the
stability of the circuit. Therefore, the source/drain impedance for each stage needs to be selected with a safe distance from the instability region while maintaining good gain/noise performance. Coplanar waveguide (CPW) transmission lines have been used for matching impedances, for interconnects, and for biasing networks. The length of these transmission lines is kept less than $\lambda/4$ to reduce the loss and to minimize the contributed noise at the amplifier input.

![Figure 4.5 Three-Stage Common-Source LNA [16]](image)

The reported results for this LNA show a peak power gain of 12.2 dB at 63 GHz, a noise figure of 6.5 dB, power consumption of 10.5 mW, and input/output return losses of -13 dB and -25 dB respectively. The measurement shows a good agreement with the simulation, but with a 1 GHz of frequency mismatch. Also, the output match of the circuit shows a big difference from the simulated results which pushes the circuit to the edge of instability.

A third way to build a 60 GHz LNA uses an inductive cascode source degeneration topology, but some modifications or considerations need to be taken into account to make this configuration suitable for this high frequency of operation. Razavi in [21] uses a modified single-stage cascode topology to build his LNA as shown in Figure 4.6. The capacitance introduced at the cascode node (node between the input and cascode devices)
severely limits the bandwidth and raise the contribution of the cascode transistor to the output noise. The effect of this capacitor can be cancelled or reduced by using a resonator inductor connected at this node. The metal sandwich capacitors C1 and C2; which tied to a wide ground plane, are used to minimize the effect of the parasitic inductances of L3, L4, and L5 which degrades the quality factor or raise their values. Introducing an inductor at this node might cause a stability problem in designing this LNA.

Figure 4.6 RF-Front End with Modified Inductive Cascode Source Degeneration LNA [21]

Doan in [12] – [15] uses a multi-stage common-source cascode amplifier to build their 60 GHz LNA as shown in Figure 4.7. Doan uses gain stages of NMOS common source cascode amplifiers to reduce miller capacitance and improve the stability to be unconditionally stable above 27 GHz. Inductive degeneration is avoided at this frequency in order to maximize the gain. CPW transmission lines were used to realize small inductances in the matching networks and to implement interconnects wiring. Proper design of ac coupling and bypass capacitors is required at these frequencies to avoid self-resonance and unwanted oscillation problems.
The reported LNA achieves a peak power gain of 12 dB, input and output return loss greater than 15 dB, 3 dB bandwidth of 51 – 65 GHz, output $P_{1dB}$ of +2.0 dBm, noise figure of 8.8 dB, and reverse isolation better than 45 dB up to 65 GHz. The simulations predict an output $P_{1dB}$ of +1.0 dBm and a noise figure of 6.9 dB.

### 4.4 Selected mm-Wave LNAs and their Simulation Results

As discussed in section 4.3, one of the big challenges in designing mm-wave LNAs is the choice of the active device sizes and the LNA architecture. For maximum achievable $f_T$ and $f_{max}$, different simulations are performed to select the optimum input transistor device size and biasing voltages as shown in Figure 4.8 and 4.9. For 90nm RF CMOS process, the optimum transistor size is found to be 16μm, while the best gate-to-source voltage bias, $V_{gs}$, is 1V at a supply voltage of 2.2V. Assuming these biasing conditions, the maximum achievable $f_T/f_{max}$ is 195/235 GHz, respectively. Two topologies have been used to design mm-wave CMOS LNAs in IBM 90nm CMOS technology. The first one uses three stages of cascode LNA, while the second one uses four-Stage Common-Gate/Common-Source Amplifiers as discussed below.
Figure 4.8 Effect of Transistor Width on $f_T$ and $f_{\text{MAX}}$

Figure 4.9 Effect of Transistor Biasing on $f_T$ and $f_{\text{MAX}}$
Topology I – Three Stages of Cascode LNA

As discussed in section 4.3, the cascode architecture provides good input matching, high reverse isolation, low noise figure, and low power consumption since the two transistors share the same bias current, but it has some limitations and challenges need to be considered in designing mm-wave LNAs. In this work, a cascode topology (with no inductive degeneration) has been used to realize the mm-wave and sub-THz LNAs. For optimum performance, \( f_T \) needs to be much higher than the operating frequency. At high frequencies, the pole of the cascode node shunts a large portion of the RF current to ground. This results in lowering the achieved gain and raising the contributed noise by the cascode device. This problem can be solved by proper selection of the input transistor size and biasing voltage to maximize \( f_T \).

Figure 4.10 shows a simplified circuit diagram of the cascode stage. Three cascode stages are then cascaded to obtain the full amplifier. Inductor \( L_{in} \) and \( L_g \) are used to control the input matching quality (i.e. the center frequency and the frequency band where the amplifier achieve a 50\( \Omega \) input matching), while the cascode device size with \( L_{out1} \) and \( L_{out2} \) are used to control the output matching quality, the peak gain, and the bandwidth of the amplifier. This is critical for the LNA design since the output impedance of each stage needs to match the input impedance of the next stage, while for the last stage needs to match 50\( \Omega \) output impedance.
Another approach to overcome the challenges in designing mm-wave LNAs is to use a (CG) stage to improve the amplifier bandwidth as shown in Figure 4.11. However, the relatively large input device in the CG stage can result in large degradation of input matching and noise figure. To overcome this problem, a resonator is built at the source terminal to cancel the large capacitance at this node. Similarly, another resonator is built at the output node to cancel the capacitance seen at the transistor drain and introduced by the next stage. The input transistor size and the value of $L_{in1}$ are used to control the input matching impedance.

To further increase the amplifier gain, a second CS stage is used after the CG input stage. The CS stage also controls the output matching impedance of the overall amplifier. In this work, two stages of the circuit shown in Figure 4.11 are cascaded to achieve higher gain without introducing significant noise to the overall LNA. The
transistor size of $M_2$ and the values of $L_{\text{load}}$, $R_{\text{load}}$ and $L_{\text{in}2}$ are used to control the peak gain, while $L_{\text{out}1}$ and $L_{\text{out}2}$ control the output matching quality. All the inductor sizes in this design are in the range of 30 – 200 pH.

![One Stage of CG-CS LNA](image)

**Figure 4.11 One Stage of CG-CS LNA**

The three-stage cascade LNA (Topology I) and four-stage CG – CS LNA (Topology II) were designed in IBM 90nm RF CMOS technology to design 60 GHz and sub-THz LNAs. Figure 4.12 - 4.15 depicts the simulated gain and noise figure for both topologies in the 60 GHz and Sub-THz LNAs. As shown in the figures, Topology I in the 60 GHz LNA shows wider bandwidth, higher peak gain of 37.1 dB at 60 GHz, and lower noise figure of 3.24 dB. However, Topology II in the 60 GHz LNA shows lower bandwidth, lower peak gain of 24.79 dB, and higher noise figure of 5.3 dB. Topology I in the Sub-THz LNA has a very wide 3-dB bandwidth of 25 GHz extending from 80 GHz to 105 GHz with a peak gain of 23.5 dB at 92.1 GHz. On the other hand, Topology II has a narrower 3-dB bandwidth of 15 GHz extending from 95 GHz to 110 GHz with a peak gain of 23.2 dB at 105 GHz. Figures 4.15 displays the simulated noise figure (NF) for the
two LNA topologies. As shown in the figures, topology I achieves better noise performance over a wide frequency range than topology II, which is directly impacted by the lower-gain-bandwidth product for Topology II. Nevertheless, both LNA topologies have a noise figure less than 6dB over the entire sub-THz frequency band.

Figure 4.12 Simulated gain for Topology (I) and Topology (II) 60 GHz LNAs

Figure 4.13 Simulated gain for Topology (I) and Topology (II) Sub-THz LNAs
Figure 4.14 Simulated Noise Figure (NF) for Topology (I) and Topology (II) 60 GHz LNAs

Figure 4.15 Simulated Noise Figure (NF) for Topology (I) and Topology (II) Sub-THz LNAs

Figures 4.16 – 4.19 depict the simulated S-parameters for topology I and topology II for both 60 GHz and Sub-THz LNAs. The main challenge in optimizing the output
matching is its affect on the stability of the overall amplifier. In particular, a tradeoff between the output matching and the stability of the LNA limit the degree by which output matching can be achieved. As a result, the size of the cascode transistor (in Topology I) or the common source transistor (in Topology II) is selected carefully to guarantee the stability of the LNA. The stability of the circuit has been verified using the two parameters, $B_{if}$ and $K_f$, provided by the Cadence circuit simulator. Inductor values are selected carefully to achieve good input/output matching and excellent reverse isolation. Figure 4.18 and 4.19 show that both topologies has good input match (> 10 dB) over the band 57 – 64 GHz, and good output match over the band 59 – 63 GHz. As shown in Figure 4.17 and 4.18, both topologies in the sub-THz LNA achieve good input match over the band 90 – 115 GHz, good output match over the band 85 – 115 GHz, and excellent reverse isolation over the entire frequency band.

![Figure 4.16 Simulated S-Parameters for Topology (I) 60 GHz LNA](image)
Figure 4.17 Simulated S-Parameters for Topology (II) 60 GHz LNA

Figure 4.18 Simulated S-Parameters for Topology (I) Sub-THz GHz LNA
This chapter presents the design of wideband LNAs in mm-wave frequencies utilizing IBM’s 90nm RF CMOS technology. The designs are based on a three-stage cascode CS amplifier topology and a four-stage CG and CS amplifier topology. The results indicate that standard 90nm CMOS processes are well suitable to design mm-wave blocks and demonstrate the potential of using CMOS for mm-wave applications. Table 4.1 and 4.2 summarize the performance of the proposed LNAs and recently reported 60 GHz and Sub-THz CMOS LNAs. The simulated performance of the proposed LNAs is shown to be better than the reported CMOS LNAs in their maximum peak gain over a very wide 3-dB bandwidth with low power consumption. Our measured values are expected to be slightly worse than the simulated ones since better models are used for passive devices using ADS MoM models and for active devices by introducing additional components to
the low frequency model.

Table 4.1 Reported performance of CMOS amplifiers in 60 GHz frequency band

<table>
<thead>
<tr>
<th>Reference</th>
<th>This work Topology I</th>
<th>This work Topology II</th>
<th>[16]</th>
<th>[20]</th>
<th>[21]</th>
<th>[23]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>90nm CMOS</td>
<td>90nm CMOS</td>
<td>130nm CMOS</td>
<td>130nm CMOS</td>
<td>130nm CMOS</td>
<td>130nm CMOS</td>
</tr>
<tr>
<td>Circuit Topology</td>
<td>3-stage Cascode</td>
<td>2-stage CG-CS</td>
<td>3-stage CS</td>
<td>6-stage Cascode CS</td>
<td>Cascode CS - CS</td>
<td>CG</td>
</tr>
<tr>
<td>3 dB BW (GHz)</td>
<td>7 (57 – 64)</td>
<td>7 (57 – 64)</td>
<td>-</td>
<td>4 (60 – 64)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Peak Gain (dB)</td>
<td>37.1</td>
<td>24.79</td>
<td>12</td>
<td>20</td>
<td>Rx: 26 – 31.5</td>
<td>28</td>
</tr>
<tr>
<td>NF (dB)</td>
<td>3.24</td>
<td>5.3</td>
<td>8.8</td>
<td>8</td>
<td>6.9 – 8.3</td>
<td>12.5</td>
</tr>
</tbody>
</table>

Table 4.2 Reported performance of CMOS amplifiers in Sub-THz frequency band

<table>
<thead>
<tr>
<th>Reference</th>
<th>This work Topology I</th>
<th>This work Topology II</th>
<th>[49]</th>
<th>[50]</th>
<th>[51]</th>
<th>[52]</th>
<th>[53]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>90nm CMOS</td>
<td>90nm CMOS</td>
<td>65nm CMOS</td>
<td>65nm CMOS</td>
<td>90nm CMOS</td>
<td>65nm CMOS</td>
<td>90nm CMOS</td>
</tr>
<tr>
<td>Circuit Topology</td>
<td>3-stage Cascode</td>
<td>2-stage CG-CS</td>
<td>3-stage CS</td>
<td>3-stage CS-CG</td>
<td>3-stage CS</td>
<td>3-stage Cascode</td>
<td>3-stage Cascode</td>
</tr>
<tr>
<td>3 dB BW (GHz)</td>
<td>25 (80 – 105)</td>
<td>15 (95 – 110)</td>
<td>27 (142 – 170)</td>
<td>10 (103 – 105)</td>
<td>-</td>
<td>-</td>
<td>22 (86-108)</td>
</tr>
<tr>
<td>Peak Gain (dB)</td>
<td>23.52 @ 92.13 GHz</td>
<td>23.2 @ 104 GHz</td>
<td>8.3 @ 150GHz</td>
<td>8 @ 140GHz</td>
<td>9.3 @ 104GHz</td>
<td>13 @ 95 GHz</td>
<td>17.4 @ 91GHz</td>
</tr>
<tr>
<td>NF (dB)</td>
<td>5.0 – 5.5</td>
<td>5.8 – 6.11</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Power Consumption (mW)</td>
<td>27</td>
<td>40</td>
<td>25.5</td>
<td>22</td>
<td>6-7</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The above simulation results for both LNAs are run under nominal conditions (typical process corner and Temperature = 27°C). The performance of the two LNAs after accounting for adding parasitic elements, process corners, and voltage and temperature variations is degraded as discussed in the next chapter. Therefore, the next step in this work is to adopt digital self-calibration techniques to further improve the LNA performance over variations in process, voltage and temperature conditions.
CHAPTER 5

MANUFACTURABLE SELF-CALIBRATED MM-WAVE LNA

The integration of analog mixed signal circuits in a System-On-Chip (SOC) platform is considered as a key solution for the new generation of complex radio transceivers. CMOS technology scaling and innovation improve the performance of the digital part tremendously, but the radio part remains a major bottle-neck. The integration of analog and digital circuits faces several challenges such as coupling digital noise through substrate and power lines. A fully integrated CMOS radio requires several design cycles to meet the product specifications with relatively low yield. Analog circuits are affected strongly by random variations in process, temperature, operating conditions, and power supply variations. These variations do not scale with process. This causes an increase in the non-recurring engineering (NRE) cost and missing market windows. There are two possibilities to solve these problems: either by designing at worst-case corner simulation; but this leads to over-design and increases power consumption, or by building self-calibration circuits inside the RF blocks resulting in radio circuits that meet all the specifications.

The above challenges in designing RF blocks are more critical at mm-wave frequency band than at low radio frequencies. This chapter proposes a methodology suited to self-calibrate mm-wave LNAs so that they achieve high yield at different corners under
random process, temperature and supply variations. The proposed solution takes advantage of the technology scaling and used digital calibration to control the current mirror that bias several points of the LNA. This chapter is organized as follows: section 5.1 gives an introduction about the concept of digital built-in self-calibration (BIST), and then section 5.2 discusses the general methodology of BIST. Section 5.3 presents few circuits used to design sensors required to perform the calibration. The proposed Gain/Input Matching self-calibration for mm-wave LNAs is presented and discussed in section 5.4. Also, section 5.4 shows some results of using these techniques to enhance the yield of the mm-wave LNAs and meet the desired performance at all corners. At the end, section 5.5 summarizes the efficiency of the proposed techniques in improving the performance of mm-wave LNAs under process, supply, and temperature variations.

5.1 Introduction

As discussed in chapter 1, first pass silicon is so important to reduce the cost of an integrated solution of analog mixed signal circuit in SoC or SoN platforms. Generally, chip failure can be due to any of five overlapping reasons: complexity, errors at the system level, errors at the block levels, SoC/Integration issues, and yield. The first reason, complexity, is an unavoidable result of integration. Complexity increases the number of blocks in the system, which increases the chance that at least one of those blocks will not work properly. The second reason for chip failure is the errors at the system level. In general, the block level specifications are derived under stable system level specifications. Therefore, any system level problem causes a chip failure even if the
blocks meet all their performance requirements.

The third reason for chip failure is the errors at the block level. The four most common errors in the block design are: process variations, parasitic elements, bad or insufficient models, or incorrect simulation test benches. Process variations can not be seen in advance, but design needs to be simulated across a wide range of process changes for robust designs. RF devices have corresponding models derived based on a series of measurements on test structures. The generated layout in the design kits closely matches these models. At the block layout level, extra routings above and around the RF device are added. However, currently parasitic extraction tools can not distinguish between the present parasitic elements in the model and the additional elements that need to be included in the extraction. This causes an over estimation of parasitic elements because of double counting of model parasitic elements. The accuracy of the device models needs to be verified before starting the block design cycle by fabricating test chips and measuring each of the used devices. Test chips need to be tolerated in the same way that actual chips are. Incorrect test bench errors occur among new designers. The most common error in a test bench is the incompletion. A proper test bench must account for the input loading, the output loading, parasitic elements on the supplies and ground networks, bond wires from the package, and coupling between aggressor and victim nodes. Including all these factors increases the test bench size, increases the run time, and might cause a converge problem. Therefore, designers need to determine the dominant factors in the performance in order to select the test bench. Errors in block design can have a large impact on the testability of a chip.
The fourth chip failure reason is SoC/Integration issues. Integrating RFIC into a SoC introduces new types of problems such as noise, cross talk, and large digital and analog circuitry verification. Moreover, chip testing becomes more difficult as the level of integration increases. Special test modes need to be implemented to allow for testing of blocks on-site. The last chip failure reason is the yield. Digital circuits, which occupy much more silicon area, have a yield above 99%, while analog and RF have much lower yields. The yield issue is critical in SoC solutions because the chip yield can be approximated as the lower of the analog yield and the digital yield.

As discussed in chapter one, different problems and issues need to be considered with the high levels of integration. The lack of observation is considered as one of the biggest problems with high levels of integration. Previously, transceivers were built using discrete components which allows a maximum amount of observability in the signal chains. This allows tracking the signal at the boundary of each discrete component and tracing back any problem to the block causing that problem. This solution is also applied to the early integrated transceivers where broad level image rejection filters are used for the off-chip IF stages. For fully integrated chips, these observation points are no longer available and complex test modes are required inside the chip where the signal is buffered with test buffers and driven off chip through dedicated test pins. These test buffers cause many problems: loading the signal path, limiting the performance of the block, and limiting the testing since only one test buffer can drive a test pin at a time. Moreover, designing a test buffer to drive an RF signal to a 50Ω load is not trivial, and consumes a considerable amount of current and a large area.
The noise coupling is the second issue with high level of integration. Filtering the bias network and filtering current mirrors approaches inside each block helps to mitigate noise coupling. However, the large biasing network filter causes slow startup and shutdown in the block. This problem can be avoided by using the dual time constant filter, in which a small time constant is used at startup for fast startup and a large time constant is used during active operation for better noise isolation. Using multiple bandgaps or voltage regulators approach provides another way to mitigate noise coupling.

Using separate power supply domain for each major section of the chip is used to minimize supply coupling. These power supplies are split between receiver and transmitter, RF and baseband, PLL, input/output, and digital giving seven different supply domains: Receiver RF supply, Receiver BaseBand supply, Transmitter RF supply, Transmitter BB supply, PLL supply, Input/Output supply, and Digital supply.

A capacitor filler cell is used to fill the extra space after assembling the chip top level. This filler cell maximizes the unit capacitance per \( \mu \text{m}^2 \) and maximizes the metal density. The simplest and most effective approach to get the largest unit capacitance in a given area is to use the stacked cell of a moscap, metal overlap capacitance, and then a mim cap on top. This structure requires to be dampened using a series of parallel resistors connected in series with the decoupling capacitance in order to prevent ringing on the RLC network formed from the supply lines and decoupling. This solution can provide a very high on-chip capacitance which might be not enough for very noisy circuits. In these cases, low-voltage dropout regulators (LDOs) are used to provide a regulated supply voltage to each independent block.
The two main issues that need to be addressed for the SoC are the performance issues and the functional issues. The performance issues can be classified into noise and crosstalk; specifically substrate isolation, package isolation, signal coupling, and performance verification of large mixed-signal blocks. The functional issues are the mixed signal verification and pre-silicon functional validations.

The substrate is shared among the RF, the analog and the digital circuits. Therefore, substrate isolation is so critical. Digital noise couples to both analog and RF blocks, while analog noise couples to the RF. Different techniques can be used for substrate isolation. High resistivity substrate on the order of 10-50 Ω–sq provides high substrate isolation through physical separation and provides an excellent RF medium by minimizing eddy currents generated from inductors in the substrate. This helps in fabricating high quality inductors on silicon and gives excellent block to block isolation. From the digital point view, a high resistivity substrate means the substrate is near floating. This requires more substrate contacts in order to ground the bulk which means a larger standard cell. Also, a high resistivity substrate makes blocks more sensitive to latch up, which requires larger separation among the digital cells and grows the digital size. Source and bulk separation of the digital nmos transistors is another technique to provide substrate isolation. This technique reduces the switching noise injected into the substrate by biasing the substrate using the quieter ground than the digital ground. Special design of the digital blocks is the last approach to provide substrate isolation. The magnitude of the digital noise can be minimized if the digital blocks are designed so that the large amount of transistors does not all switch at the same time. Using delay
elements to distribute the clock switching time on the non-critical signals can be used to accomplish this task.

Mixed signal simulations are necessary for block level verification. However, simulating large digital blocks with analog and RF takes a long time and suffers from the lack of different levels of abstraction in existing tools. Therefore, a mixture of behavioral and transistor level circuits and high speed simulators can be used to speed up the simulation time. The block accuracy in these simulators can be set based on the targeted performance, but these simulators only perform transient analysis in the time domain. The analog and RF blocks are verified for connectivity and model functionality using a developed behavioral model.

After finishing the layout of all blocks and integrating them together at the top level, several additional steps are needed for the tapeout. These steps are: adding dummy blocking layers, adding p-implant blocking layers, and adding the die seal. Minimum metal density is required by the fabrication company to insure uniform metallization for the greatest accuracy in the fabrication steps. Minimum metal density is around 30%, depending on the layer. Minimum metal density is a challenging task in RF chips because of the inductors. Inductors require no other metal near them to function properly and occupy the large portion of these chips. Therefore, it can be very difficult to meet this minimum metal density condition. To meet the metal density requirement, a dummification program needs to be run by the fab or the customer to fill in the empty spaces with a dummy pattern. The program checks the metal density by looking at a small window and inserts metal in a specific pattern if the window does not meet the
minimum metal density. For matched and RF circuits, the dummy pattern can create non-matched conditions and disturbance to the fields generated by the high frequency operation. Therefore, it is necessary to mark areas where dummy metal should not be placed using special layers called dummy blocking layer. Dummy blocking layers surround all active devices, inductors, mim capacitors, precision resistors and matched elements.

The P-implant blocking layer is added to a certain process similar to the dummy blocking layer to prevent the native p-substrate from being doped to a low-resistivity substrate. This high resistivity substrate cannot be used to make active devices. Therefore, it is only used for special RF devices, such as inductors, RF mim capacitors and RF signal routings. The last additional step in the tapout is adding the die seal. The die seal is important to prevent the substrate from cracking during the sawing of the dice. It consists of all stacked metal layers with vias included. The special die seal called a staggered die seal is used in RF design to minimize the coupling noise from all parts the chip around to the other parts through the ring of metal (die seal). The staggered dies seal consists of two die seals with breaks overlapping in each seal.

The final DRC check is performed on the entire chip after adding all the extra layers, placing the die seal and adding the chip ID. The chip is sent for fabrication in gds2 format once its DRC is clean. Then, the fab runs its own DRC check. The reticle layout which generates the layout of a series of chips next to each other to give the stepper an image large enough to use and starts after verifying a clean DRC. The reticle is used to produce more than one version of a chip in a single full mask tapout. Test structures are
placed between the dice on the reticle to monitor the process. The last DRC check is run on the whole reticle. Then, the reticle will be sent to the photomask house if it passes the last DRC check. At the photomask house, the data is converted to the used format (mebes) and produce the final mask pattern from the data using their own programs. The long checklist needs to be passed. The key points for this checklist are: mapping every layer correctly, placing dummy patterns correctly and on the correct layer, including every layer and checking each die seal in the reticle. This manual checking takes from several hours to several days depending on the number of individual designs. The data is then sent down to the photomask fab to create the mask after the designers are satisfied with it and sign the mask signoff form. At the end, the completed wafers are sent to a packaging house to cut the individual dice out, package them and then ship them to the customer.

5.2 General Methodology

As mentioned in chapter 1, self awareness in RFIC means that each block and system performance is evaluated and quantified by measuring its input and output characteristics. This is performed using special RF detector to measure the key parameters for each block. This RF detector will be discussed in more details in section 5.3. This section discusses the general methodology in developing this self awareness and self calibration in RF blocks in order to meet their specifications under worst case scenarios and with process, temperature, and power supply variations.

The use of the RF detector at the input and output of the Device Under Test (DUT)
initiate the self awareness in the blocks where each block has a mechanism to track its own performance. But in order to implement calibration, few steps need to be done. Firstly, the cause of the impairments needs to be recognized. Usually, similar blocks share the same root cause of an error. This means that there is no need to use different circuit techniques to address the same issues in different blocks. For example, reduced gain in RF blocks with tuned tanks is caused by parasitic capacitance and inductance in the tank. To overcome this parasitic problem, tunable elements are usually used in this tank. Secondly, certain circuit techniques need to be used to turn on the block performance in order to complete the calibration loop as shown in Figure 5.1.

![Figure 5.1 Self Awareness and Calibration Methodology](image)

Using self aware and calibration techniques have several advantages. First, with these techniques, blocks are able to compensate for corner variations without the need to be over-designed. So the block runs over or below the nominal situation only if the self awareness shows that it is below or above the specification respectively. Second, with self awareness and self calibration, the blocks are locally optimized based on global parameters. So this relaxes some parameters’ performance at the cost of the degradation of the other’s performance and improves the overall performance of the block. Third, these tuned blocks can be used for multi-standards and multi-center frequency with no
need to have separate block for each standards. So this extends the coverage of these blocks and uses these shared blocks without duplicating the used hardware.

The last step to complete this loop is to build a self configuration technique where the system is able to adjust the goals of self calibration routines in order to change the block configuration. This gives the block the flexibility to adapt to different set of requirements. Two approaches are used to achieve block parameters changes in order to maximize a system parameter. The first approach is the system analysis approach where the performance is derived from an analytic analysis of the system blocks and the used modulation. This approach provides insight into the most critical parameters, but it is time consuming for a multi-standard SoC. The second approach is the algorithmic approach where optimization engine are used to implement optimization routines. This can be done without knowing the used modulation schemes in advance.

An excellent self-calibration technique has been proposed in [51] to calibrate inductive source denegation LNAs operating at one-digit radio frequencies. Calibration cycle requires tracking the input and the output of the device under test (DUT) using suitable and sensitive sensor, converting the dc output voltage of the sensor into its equivalent digital signal, analyzing the results, generating a correction signal using digital calibration algorithm, and then feeding back the correction signal to the DUT as shown in Figure 5.1 [52]. The calibration circuit should have fast calibration time, low additional power consumption during calibration and during operation.
5.3 Sensor Design

An RF sensor shown in the above figures must satisfy two main conditions: it should minimally load the input and the output of the DUT, and it must translate the important information from the RF signal into a low-frequency or DC signal which is suitable for further processing by the calibration algorithm. Two ways can be used to test and measure the performance of single blocks or amass of blocks. The first way is performed by using multiple detectors that are inserted at different points to extract signal amplitudes. The other way is done by using a single detector with switches that selects various measurement points. For the first way, the detector is designed for area efficiency. While for the second way, the detector is designed for very wide dynamic range. For accurate and fine detection of small amplitude changes, detector also must satisfy other requirements including low power, wideband operation; high input
impedance in order not to load any circuit connected to it, and high high-frequency-to-dc conversion gain. A high conversion gain is very desirable for self-test and self-calibration as it eases the requirements on the digitizing ADC. However, increasing the conversion gain will limit the dynamic range as the power supply voltage decreases.

An example of a first RF detector used for low frequency detection is given in [53]. The schematic of that detector is shown in Figure 5.3. The device is in the active region when the input amplitude is low and the output voltage is calculated by the drain current equation. While the device is in the nonlinear region when the input is large and $C_2$ is discharged because $I_d$ becomes larger than $I_{bias}$. The input/output relationship is shown in Figure 5.4. This detector was designed and tested using 3.3 V supply voltage. In fact, all previous detectors described in the literature [53] – [58] do not offer high resolution and mostly suffer from a low RF-to-dc conversion gain of around -1V/V as shown in Figure 5.5. Therefore, a modified detector is required to fulfill the requirements mentioned earlier and to operate properly at mm-Wave frequencies.

![Figure 5.3 Schematic of an RF detector [53]](image)
The modified detector is proposed in [59] and is shown in Figure 5.6. It is similar to one proposed in [53] and shown in Figure 5.3. However it depends on a separate gate biasing circuit and it does not use a feedback resistor for biasing. Detector input transistor
is biased in the sub-threshold region and its saturated drain current can be expressed as [49]

\[ I_n = I_{D0} \left( \frac{W}{L} \right) \exp \left( \frac{V_{GS}}{nV_T} \right) \]  

where \( I_{D0} \) is a current constant independent of gate-to-source voltage, \( (W/L)_n \) is the aspect ratio of the NMOS transistor, \( n \) is a process dependent term related to depletion region characteristics, and \( V_T \) is the thermal voltage.

If a small sinusoidal input voltage is superimposed, \( V_a \cos(\omega t) \), on the gate bias, \( V_{bias} \), the following power series approximation of the current equation holds [59]:

\[
I_s = I_{D0} \left( \frac{W}{L} \right) \exp \left( \frac{V_{bias} + V_s \cos(\omega t)}{nV_T} \right) = I_{D0} \exp \left( \frac{V_s \cos(\omega t)}{nV_T} \right)
\]

\[
\approx I_{D0} \left[ 1 + \frac{V_s}{nV_T} \cos(\omega t) + \frac{1}{2} \left( \frac{V_s}{nV_T} \right)^2 \cos^2(\omega t) \right] \]

\[
= I_{D0} \left[ 1 + \left( \frac{V_s}{2nV_T} \right)^2 + \frac{V_s}{nV_T} \cos(\omega t) + \left( \frac{V_s}{2nV_T} \right)^2 \cos(2\omega t) \right]
\]

Where \( I_{BO} = I_{D0}(W/L)_n \exp(V_{bias}/nV_T) \) is the dc-bias current of the transistor.

If the input capacitor is initially charged to \( V_{DD} \) and an input signal is applied at the input terminal, then DC\(_{out}\) is discharged by the drain current \( I_D \) through the NMOS transistor. This creates a negative I-V relationship with respect to an increase in the mm-Wave signal amplitude. Because of low-pass RC filter at the load, only the dc component of \( I_D \) is shown at the output and is given the first two terms in the last equation 5.2
To achieve higher conversion gain, input device sizing and biasing should be selected properly. However, for wide dynamic (i.e. amplitude) range, the detector should employ several modes of operation each covering an overlapping sub-range of amplitudes. Changing the gate biasing of the input device is used to achieve this goal. If the gate bias decreases, higher amplitude signals is required to turn the input transistor on. A 3-bit programmable voltage bias circuit is designed as shown in Figure 5.7. This biasing circuit has eight modes of operation that covers a wide signal range from 0 to 0.5V\textsubscript{amp} with an overlap of 50mV\textsubscript{amp} (\Delta V\textsubscript{bias}=35mV). The DC detector output is linear from 0.2V to 1V (for a 1.2V supply) and has a slope of around -9V/V with high impedance above 800 Ω across the 60 GHz band as shown in Figure 5.8. The detector has been built and tested in IBM’s 90nm CMOS technology.

Fig. 5.6 Modified mm-wave detector [59]
Fig. 5.7 Programmable Digital Mode Select Circuit [59]

Fig. 5.8 HF/DC conversion response [59]
5.4 Gain/Input Matching Calibration for mm-wave LNAs

This section proposes design techniques to deal with the problem of parasitic elements and process shift that affect the RF circuit tank. It also presents a tunable biasing technique to control the LNA gain. Using these techniques, the performance of the LNA block is quantified and calibrated to achieve the desired performance. These techniques have been demonstrated through simulations of four LNAs: two 60 GHz LNAs using topology I and II discussed in chapter 4, and two sub-THz LNAs using topology I and II. The implemented calibration circuits show fast calibration time, low additional power consumption during calibration, and negligible power consumption during operation.

As shown in Figure 5.2, the calibration cycle starts by processing the input and output signals using the sensor discussed in section 5.3. After that, the sensed DC output is converted into digital word, and then processed by the digital calibration algorithm. The input testing signal can be designed using an external unique signal, or it can be generated using the integrated transmitter in the complete transceiver.

The proposed technique in [57] for gain and input matching calibration can be exactly applied for mm-wave LNAs discussed in chapter 4. The two major differences between the proposed techniques in [57] and the ones presented in this work are: First, the gain calibration for mm-wave LNA is performed using a digitally controlled biasing voltage. Second, the additional two capacitors $C_d$ and $C_s$ in [57] cannot be used for the mm-wave LNA, because this will further decrease the value of the inductors making the design to be more sensitive to packaging parasitics. Therefore, other calibration techniques need to
be used to improve the input matching performance of the LNA under worst case conditions and under process and temperature variations.

The gain of the LNA depends strongly on the transconductance of the NMOS input transistor, $g_m$, and $g_m$ is related and controlled by the biasing current flow through the input transistors. Therefore, the best way to calibrate the gain is by controlling the biasing current that affects the input transconductance. This can be performed by changing the biasing point $V_{b1}$ and $V_{b2}$ for each stage of the cascode CS topology and the basing points $V_{b1}$, $V_{b2}$ and $I_{bias}$ for each stage of CG/CS topology. To control the input matching and the center frequency, a switchable load and input devices are used. The digitally tunable elements used in the two mm-wave LNA topologies are shown in Figures 5.10 and 5.11.

![Figure 5.10 Tunable One Stage Cascode LNA](image)

Figure 5.10 Tunable One Stage Cascode LNA
Figure 5.11 tunable One Stage of CG-CS LNA

The biasing of the LNA is achieved by ratioed mirroring currents derived from a reference voltage supply generated using a bandgap circuit. $V_{bias}$ is changed by switching 4 PMOS transistors sized with binary weights for each biasing voltage separately. These transistors are switched digitally using a 4-bit control word. The output biasing voltage varies by $\pm 100$ mV around its nominal value. Figure 5.12 shows the used current mirror biasing and calibration circuit for both $V_b$ and $I_{bias}$. Figure 5.13 shows the gain variation for different biasing calibration words. This variation covers the expected degradation in the LNA gain under random process, temperature and supply variations. The reference voltage for the biasing circuit was designed using a bandgap circuit. Parametric simulations have been run on the biasing voltage of the input transistors to demonstrate its effect on the overall gain of the LNA. The calibration of the PMOS current mirrors ($V_b$ and $I_{bias}$) are designed to have a tuning range of $\pm 25\%$. 

117
To calibrate the input/output matching quality and the center frequency for the LNA, variable passive and active components are used instead of fixed ones. There are two possibilities to achieve this task: either by controlling $C_{gs}$ of the input device which depends on the input device size, or by using switchable passives for the input and output.
matching circuits. Changing the input device size might change the achievable gain in a reverse direction. Therefore it is not recommended to be careful in using this technique for matching quality calibration. In this work, only tunable input and load passive devices are used for this purpose as shown in Figure 5.14.

Figure 5.14 Tunable Passive and Active Devices

The calibration algorithm can be stated as follows:

1. Set the input signal at the desired frequency.

2. Sweep the input passive device until maximum sensor output is achieved (optimum input matching, set $L_{in} = L_{in, opt}$).

3. Sweep the final load passive device until maximum output is achieved (optimum output matching, set $L_{out} = L_{out, opt}$).

4. Additional sweeps can be done on the interstages passives to maximize the achieved output.

5. Sweep the basing voltages to maximize the produced output (optimum gain is achieved set $V_b = V_{b, opt}$ and $I_{bias} = I_{bias, opt}$).
6. Calibration finished. Resume normal operation with $L_{in} = L_{in,\text{opt}}$, $L_{out} = L_{out,\text{opt}}$, $V_b = V_{b,\text{opt}}$ and $I_{bias} = I_{bias,\text{opt}}$.

As a demonstration of the proposed calibration techniques, full transistor level simulations of three stages of cascode CS and four-stage CG/CS for 60GHz and Sub-THz LNAs are performed. The four references LNAs are designed in IBM 90nm technology.

Table 5.1 shows the simulation results for both 60 GHz LNAs under typical case where the normal simulation corner is used and no extra parasitics are included at 60 GHz frequency ($V_{dd} = 1.2V$, Temperature = $25^\circ$C). When the parasitic elements, process corners, voltage or temperature change, the reference LNA performance is completely degraded as shown in Table 5.1. The corner conditions vary from $-10^\circ$C to $75^\circ$C and +/-10% power supply variations.

Table 5.1 Simulation Results for the Reference 60 GHz LNAs

<table>
<thead>
<tr>
<th>Parameter / Corner</th>
<th>Topology I</th>
<th>Topology II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av (dB) [TT]</td>
<td>37.1</td>
<td>24.79</td>
</tr>
<tr>
<td>NF (dB) [TT]</td>
<td>3.246</td>
<td>5.3</td>
</tr>
<tr>
<td>Av (dB) [SS]</td>
<td>27.98</td>
<td>15.87</td>
</tr>
<tr>
<td>NF (dB) [SS]</td>
<td>5.83</td>
<td>8.7</td>
</tr>
<tr>
<td>Av (dB) [FF]</td>
<td>40.06</td>
<td>27.43</td>
</tr>
<tr>
<td>NF (dB) [FF]</td>
<td>3.15</td>
<td>4.49</td>
</tr>
</tbody>
</table>

The calibrated LNAs for the two topologies have been implemented using the tunable circuits shown in Figure 5.12 and 5.14 to tune the gain of the LNA by controlling the biasing voltage for the used transistors and to tune the quality of the input/output matching by controlling input and output passives. The proposed calibration method has been executed for the same corner conditions (TT and SS) under different correction
words to show the effect of the calibration circuits on each LNA’s gain. Table 5.2 shows the simulation results of the calibrated LNAs under different corner conditions. As shown, the proposed calibration technique can bring the LNA back to meet its specifications under typical and worst case conditions. As noticed from the results, an additional benefit of the gain and input/output calibrations is the improvement of the noise figure of the LNA. The remarkably improved performance of the two LNA topologies under different corners using the proposed calibration techniques offers interesting possibilities in designing 60 GHz LNAs with high yield from the first design cycle.

Table 5.2 Simulation Results for the Calibrated 60 GHz LNAs.

<table>
<thead>
<tr>
<th>Parameter / Corner</th>
<th>Topology I</th>
<th>Topology II</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_v$ (dB) [TT]</td>
<td>36.4</td>
<td>24.13</td>
</tr>
<tr>
<td>$NF$ (dB) [TT]</td>
<td>3.62</td>
<td>5.87</td>
</tr>
<tr>
<td>$A_v$ (dB) [SS]</td>
<td>32.82</td>
<td>20.43</td>
</tr>
<tr>
<td>$NF$ (dB) [SS]</td>
<td>4.13</td>
<td>6.53</td>
</tr>
</tbody>
</table>

For Sub-THz LNAs, the simulation results displayed in chapter 4 for both LNAs are run under nominal conditions (typical process corner and Temperature = 27°C). The performance of the two LNAs after accounting for adding parasitic elements, process corners, voltage, and temperature variations is degraded as listed in Table 5.3. The fast-fast (FF) and slow-slow (SS) corners have temperature variation of $-10^\circ$C to $80^\circ$C and power supply variation of +/-10% respectively. Topology I LNA shows a gain variation of 19.3 – 27.4 dB and noise figure variation of 3.3 – 7.3 dB, while topology II LNA shows a gain variation of 10.5 – 29 dB and noise figure variation of 5.1 – 7.9 dB. This wide variation, in particular for the LNA gain, illustrates the importance of introducing
the above self-calibration mechanisms to meet the nominal LNA specifications under both typical and worst case conditions.

Table 5.3 Corner Simulation Results for the two Sub-THz LNAs

<table>
<thead>
<tr>
<th>Parameter / Corner</th>
<th>Topology I</th>
<th>Topology II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak Gain (dB) [TT]</td>
<td>23.5 @ 92.1 GHz</td>
<td>23.2 @ 105 GHz</td>
</tr>
<tr>
<td>Peak Gain (dB) [FF]</td>
<td>29 @ 89 GHz</td>
<td>27.5 @ 102 GHz</td>
</tr>
<tr>
<td>Peak Gain (dB) [SS]</td>
<td>16.3 @ 98 GHz</td>
<td>12.8 @ 105.9 GHz</td>
</tr>
<tr>
<td>NF (dB) [TT]</td>
<td>5.0</td>
<td>5.9</td>
</tr>
<tr>
<td>NF (dB) [FF]</td>
<td>3.3</td>
<td>4.75</td>
</tr>
<tr>
<td>NF (dB) [SS]</td>
<td>7.3</td>
<td>7.9</td>
</tr>
</tbody>
</table>

The calibrated Sub-THz LNAs for the above two topologies have been designed using the tunable circuits shown in Figure 5.12 and 5.14. Table 5.4 shows the simulation results of the calibrated LNAs under different corner conditions. Again using the proposed calibration technique, the two Sub-THz LNAs meet their gain and input/output matching specifications under typical and worst case conditions with an additional noise figure benefit.

Table 5.4 Simulation Results for the Calibrated Sub-THz LNAs

<table>
<thead>
<tr>
<th>Parameter / Corner</th>
<th>Topology I</th>
<th>Topology II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak Gain (dB) [TT]</td>
<td>23.16 @ 92.1 GHz</td>
<td>23.06 @ 105 GHz</td>
</tr>
<tr>
<td>Peak Gain (dB) [SS]</td>
<td>21.7 @ 93 GHz</td>
<td>19.7 @ 105.1 GHz</td>
</tr>
<tr>
<td>NF (dB) [TT]</td>
<td>5.15</td>
<td>6.19</td>
</tr>
<tr>
<td>NF (dB) [SS]</td>
<td>6.1</td>
<td>6.43</td>
</tr>
</tbody>
</table>

5.5 Conclusions

This chapter proposes digital self-calibration techniques which allow for the tuning of the input matching, gain, as well as the load tank for mm-wave LNAs’ to enhance their yield to at least 90%. It is proven that only measuring the output voltage of the LNA is enough in order to minimize the input/output mismatch and to maximize the achievable
gain. A circuit to perform this measurement is presented and calibration routines are introduced to perform the input/output matching and gain tuning. The proposed technique shows the effectiveness of the input/output matching and gain calibration algorithms to maintain typical specified performance at worst case corners and hence allowing for manufacturable 60GHz and Sub-THz RF CMOS design for high volume applications without leading to over-design or increasing power consumption. The LNA has been designed and tested in IBM 90nm technology.
CHAPTER 6

CONCLUSIONS AND FUTURE WORK

New generation of radio transceivers are required to deliver high data rate with minimum modulation complexity and to operate at more frequencies with each new generation. CMOS mm-wave radios can be used to meet this demand while keeping costs to minimum for short range wireless LAN systems. CMOS transceivers potentially provide highly integrated inexpensive mm-wave devices and can eventually be transformed into main stream high volume applications. Building mm-wave transceivers require careful selection of transceiver architecture and circuit topology for each RF block to overcome the challenges and limitations at this high frequency band.

The exponential rate of reduction in minimum feature size of CMOS transistors has accelerated the pace of technology adoption for millimeter-wave (mm-wave) System-on-Chip (SoC) solutions. In comparison, CMOS circuits have specific advantages over SiGe or BiCMOS counterparts covering low-power active devices with high transition frequencies \( f_T \), and low-cost high volume manufacturing capabilities for mixed-signal/RF systems. The major drawback of standard CMOS technology is the limited RF performance of its on-chip passive components and interconnects, which are deemed critical for mm-wave circuits. As such, the maximum achievable circuit speed and thus its yield are often constrained by the quality and accuracy of circuit’s interconnects.
Existing parasitic extraction tools supplied by process design kits (PDKs) provide inadequate and inaccurate modeling parameters at the post-layout simulation level. This is attributed to frequency dependent distributed effects of interconnects that are not considered in the model extraction flow. Therefore, an efficient and accurate model for transmission lines (passives or interconnects) is critically important to reduce the discrepancies between simulated and measured performance of mm-wave circuits.

As the number of integrated blocks increases in the transceiver, the opportunity for mistakes in the design grows. In addition to design mistakes, the process, temperature, and supply variations do not scale with newer technology generations, leading the design to deviate from its nominal specifications and leading to over design and excessive power consumption.

The work presented here addresses few solutions to improve the mm-wave CMOS transceivers performance – in particular the LNA block (first RF block in the receiver). The first part of this work presents two LNA topologies based on a three-stage cascode CS amplifier topology and a four-stage CG and CS amplifier topology. It was demonstrated that these circuits could achieve very high gain, wideband LNAs with low power consumption in mm-wave frequencies.

This research work has some contributions on designing the future mm-wave radios. Firstly, this work presents an efficient and accurate RLCG frequency dependent distributed electrical model for transmission lines in a lossy silicon substrate based on a full-wave simulation technique using Agilent’s Momentum simulator (MoM). It was demonstrated that the distributed model confirms the accuracy of the extracted model.
from 20-65 GHz. The proposed extraction method improves the accuracy of characterizing TLs compared to commonly used models based on standard CMOS technology files and reduces the required simulation time compared to full-wave electromagnetic simulators. The generated models are fully compatible with HSPICE and SPECTRE-RF simulators. The development of a segmentation algorithm to further improve the matching between measured and simulated S-parameters is currently under development.

Secondly, this work addresses the design issues for higher level of integration through introducing few techniques of self calibration. It was demonstrated that the gain and input/output matching of the LNA can be calibrated without the use of external test equipment and without leading to over design your circuit under all different corners and process, temperature and supply variations. These techniques for calibrating the LNA gain and input/output matching can be extended in the future to the mixer, PLL, and analog baseband blocks. The development of gain and linearity calibration of mm-wave mixers is currently under development.
LIST OF REFERENCES


