Rana Azhar Shaheen

DESIGN ASPECTS OF MILLIMETER WAVE MULTIBAND FRONT-ENDS
RANA AZHAR SHAHEEN

DESIGN ASPECTS OF MILLIMETER WAVE MULTIBAND FRONT-ENDS

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Information Technology and Electrical Engineering of the University of Oulu for public defence in Auditorium F202 of the Faculty of Medicine (Aapistie 5 B), on 17 November 2021, at 12 noon

UNIVERSITY OF OULU, OULU 2021
Abstract

Multiple millimeter-wave (mmWave) frequency bands allocated for fifth generation (5G) mobile communication systems by the 3rd Generation Partnership Project (3GPP) standard have created the needs to develop multi-band mmWave receiver front-end systems. The required multiband operation can be achieved by one of the following ways: by utilizing a very broadband receiver, utilizing band specific receivers, which are selected using an external switch, or by utilizing a receiver where the center frequency can be changed by a programmable configuration.

The design aspects presented in this thesis are targeted primarily for integrated solutions of narrow-band and multi-band phased array receiver front-end designs for 5G mmWave systems. However, each of these are studied from a different perspective. To understand the implementation complexity, a fully integrated four elements receiver phased array front-end employing analog beamforming technique is designed. The design of an off-chip input matching circuit based on non-uniform transmission line (NUTL) segments is demonstrated, which transforms the reactive input impedance of the low noise amplifier (LNA) to real 50 Ohms with a sufficient bandwidth. The implemented receiver front-end supports two simultaneous data streams to facilitate multiple input multiple output (MIMO) processing. A comparison study is presented on the design of two different layout styles for cascade-connected active devices to optimize the device performance for mmWave frequency systems. The efficient design techniques of silicon-based passive devices and their accurate electromagnetic modelling methodology are presented. Moreover, design challenges and architecture limitations in the design of mmWave narrowband and frequency reconfigurable multi-band LNAs are also discussed in this thesis. Further details of these contributions are articulated in the author’s original publications I-VIII.

Keywords: 5G, LNA, mmWave, phased array
Shaheen, Rana Azhar, Monikanavaisten millimetriaaltoetupäiden suunnittelunäkökulmia
Oulun yliopiston tutkijakoulu; Tieto- ja sähköteknikan tiedekunta
Acta Univ. Oul. C 810, 2021
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä
3GPP (3rd Generation Partnership Project) on alokoinut useita millimetriaallon (mm-alueen) taajuuskaistoja viidennen sukupolven (5G) matkaviestinkäyttöön, ja tämä on luonut tarpeen monikaististen mm-aaltoalueen radioiden kehittämiseksi. Monikaistainen radiovastaanotin voidaan toteuttaa vaihtoehtoisesti laajakaistaisella vastaanottimella, ulkoisesti valittavilla kapeakaistaisilla vastaanottimilla, tai vastaanotinrakenteella, jonka keskitaajuutta voidaan muuttaa ohjelmanmallisella konfiguroinnilla.


Asiasanat: 5G, mm-aaltoalue, pienikohinainen vahvistin, vaiheistettu antenni
For their love and deep affections with me, this thesis is dedicated to my late Khala and Mamon who lost their life due to COVID-19
Preface

The research work presented in this thesis was carried out at the Center for Wireless Communication, Radio Technologies (CWC-RT) unit at the University of Oulu, Finland during 2015-2019. I would like to present my special thanks to Professor Aarno Pärssinen for providing me with an opportunity to pursue my PhD degree in his research group. I sincerely appreciate his guidance throughout this thesis work.

I would like to thank the reviewers of this thesis, Professor Mark Ingels from Delft University, The Netherlands, and Professor Andrea Mazzanti from the University of Pavia, Italy, who helped me to improve the readability and quality of this thesis. I also thank the members of my follow-up group, Dr. Tarmo Ruotsalainen, Dr. Markus Berg and Dr. Jussi Nissinen, who guided me through the right process of the PhD degree at the University of Oulu.

I want to express my sincere gratitude to Professor Timo Rahkonen for offering his support throughout my PhD degree. He guided me in each and every step of my PhD and taught me many things in the world of RFIC design. I would like to thank my colleagues in CWC-RT especially Rehman Akbar with whom I worked in multiple projects. Our coffee discussions were always mesmerizing which I miss very much. I would also like to thank my other colleagues, Mr. Alok Sethi, Dr. Janne Aikio, Mr. Nuutti Tervo, Dr. Olli Kursu and Mr. Mikko Hietanen, for all the discussions, coordination and technical support during my PhD degree. I would also like to thank Dr. Marko Leinonen, Mr. Risto Vuoltoniemi and Mr. Matti Polojarvi for their technical support in the Lab measurements during my research work.

I am thankful to all my friends in Oulu with whom I spent memorable moments during daily coffee meetups and various gathering events including sports sessions. Haresh Kumar and Zeeshan Asghar have always been very supportive in all the matters and they encourage me during difficult times of my PhD. For Haresh, I would like to quote, "Not all brothers are siblings". In addition, I would like to thank Ijaz Ahmad, Asadullah Javed, Salman Qayyum, Muzzamil, Irfan Khan, Mudassir Chaudhary, Nauman Bashir, Asif and Irtiza for all the good times, fruitful discussions, sports sessions and delicious food gatherings. I am also thankful to Mr. and Mrs. Saad Saud for all the pleasant moments we have spent together.

I would like to acknowledge the NOKIA Foundation and the Tauno Tönning Foundation for the financial support for this thesis.

In addition, I would like to extend my deepest gratitude towards my family back in Pakistan. Thank you my dearest Ammi ji, Abu ji (late), Paji Tariq, Paji Zulfiqar, Zahida
and Shahida for all of your support and guidance throughout my life. Paji Tariq has a special place in my life who has always supported me unprecedently, anytime and everywhere. Thank you Paji for this. All what I have been able to achieve would have never been possible without your support and prayers.

Finally, I would like to thank my wife Maryam for the continuous support you have provided me during these years, and especially for taking care of our children (Dua and Rehan) single-handedly during the long working hours due to various tape-outs. Without your support, it would have been very difficult to achieve this goal.
# List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>5G</td>
<td>Fifth generation</td>
</tr>
<tr>
<td>mmWave</td>
<td>Millimeter Wave</td>
</tr>
<tr>
<td>RF</td>
<td>Radio frequency</td>
</tr>
<tr>
<td>GHz</td>
<td>Giga hertz</td>
</tr>
<tr>
<td>IC</td>
<td>Integrated circuit</td>
</tr>
<tr>
<td>Si</td>
<td>Silicon</td>
</tr>
<tr>
<td>GaAs</td>
<td>Gallium arsenide</td>
</tr>
<tr>
<td>SiGe</td>
<td>Silicon germanium</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary metal oxide semiconductor</td>
</tr>
<tr>
<td>BiCMOS</td>
<td>Bipolar complementary metal oxide semiconductor</td>
</tr>
<tr>
<td>SOI</td>
<td>Silicon on insulator</td>
</tr>
<tr>
<td>5G-NR</td>
<td>5G New Radio</td>
</tr>
<tr>
<td>WoW</td>
<td>Way of working</td>
</tr>
<tr>
<td>RFIC</td>
<td>Radio frequency integrated circuit</td>
</tr>
<tr>
<td>LTE</td>
<td>Long term evolution</td>
</tr>
<tr>
<td>LTE-A</td>
<td>LTE-Advanced</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal Frequency Division Multiple Access</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature amplitude modulation</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>BW</td>
<td>Bandwidth</td>
</tr>
<tr>
<td>Gbps</td>
<td>Giga bits per second</td>
</tr>
<tr>
<td>FSPL</td>
<td>Free space path loss</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>Wavelength</td>
</tr>
<tr>
<td>( \phi_{in} )</td>
<td>Incident angle</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Time delay</td>
</tr>
<tr>
<td>$c$</td>
<td>Speed of light</td>
</tr>
<tr>
<td>$\omega$</td>
<td>Angular frequency</td>
</tr>
<tr>
<td>$f$</td>
<td>Frequency</td>
</tr>
<tr>
<td>$AF$</td>
<td>Array factor</td>
</tr>
<tr>
<td>$F$</td>
<td>Noise factor</td>
</tr>
<tr>
<td>$NF$</td>
<td>Noise figure</td>
</tr>
<tr>
<td>$BB$</td>
<td>Baseband</td>
</tr>
<tr>
<td>$IF$</td>
<td>Intermediate frequency</td>
</tr>
<tr>
<td>$DBF$</td>
<td>Digital beamforming</td>
</tr>
<tr>
<td>$HBF$</td>
<td>Hybrid beamforming</td>
</tr>
<tr>
<td>$PCB$</td>
<td>Printed circuit board</td>
</tr>
<tr>
<td>$VM$</td>
<td>Vector modulator</td>
</tr>
<tr>
<td>$\mu m$</td>
<td>Micro meter</td>
</tr>
<tr>
<td>$NUTL$</td>
<td>Non-uniform transmission line</td>
</tr>
<tr>
<td>$VGA$</td>
<td>Variable gain amplifier</td>
</tr>
<tr>
<td>$AC$</td>
<td>Alternate current</td>
</tr>
<tr>
<td>$PA$</td>
<td>Power amplifier</td>
</tr>
<tr>
<td>$PAE$</td>
<td>Power added efficiency</td>
</tr>
<tr>
<td>$MOSFET$</td>
<td>Metal oxide semiconductor field effect transistor</td>
</tr>
<tr>
<td>$f_t$</td>
<td>Cutoff frequency of a transistor</td>
</tr>
<tr>
<td>$f_{max}$</td>
<td>Maximum oscillation frequency of a transistor</td>
</tr>
<tr>
<td>$g_m$</td>
<td>Transconductance</td>
</tr>
<tr>
<td>$C_{gg}$</td>
<td>Gate capacitance of a transistor</td>
</tr>
<tr>
<td>$C_{gs}$</td>
<td>Gate to source capacitance</td>
</tr>
<tr>
<td>$C_{gd}$</td>
<td>Gate to drain capacitance</td>
</tr>
<tr>
<td>$r_g$</td>
<td>Gate resistance</td>
</tr>
<tr>
<td>$BOX$</td>
<td>Buried oxide</td>
</tr>
<tr>
<td>$HR$</td>
<td>High resistivity</td>
</tr>
<tr>
<td>$SiO_2$</td>
<td>Silicon oxide</td>
</tr>
<tr>
<td>$PCS$</td>
<td>Parasitic surface conduction</td>
</tr>
<tr>
<td>$PDK$</td>
<td>Process design kit</td>
</tr>
<tr>
<td>$SFR$</td>
<td>Self-resonant frequency</td>
</tr>
<tr>
<td>$L$</td>
<td>Inductance</td>
</tr>
<tr>
<td>$Q$</td>
<td>Quality factor</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Skin effect</td>
</tr>
<tr>
<td>$EM$</td>
<td>Electromagnetic</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>DBF</td>
<td>Digital beamforming</td>
</tr>
<tr>
<td>DBF</td>
<td>Hybrid beamforming</td>
</tr>
<tr>
<td>MoM</td>
<td>Moment of methods</td>
</tr>
<tr>
<td>FEM</td>
<td>Finite element method</td>
</tr>
<tr>
<td>2.5D</td>
<td>Two and half dimensions</td>
</tr>
<tr>
<td>3D</td>
<td>Three dimensions</td>
</tr>
<tr>
<td>DUT</td>
<td>Device-under-test</td>
</tr>
<tr>
<td>$y_{11}$</td>
<td>Input admittance</td>
</tr>
<tr>
<td>MGR</td>
<td>Metal ground ring</td>
</tr>
<tr>
<td>MIM</td>
<td>Metal-insulator-metal</td>
</tr>
<tr>
<td>$L_p$</td>
<td>Primary inductance of a transformer</td>
</tr>
<tr>
<td>$L_s$</td>
<td>Secondary inductance of a transformer</td>
</tr>
<tr>
<td>$k$</td>
<td>Mutual coupling between primary and secondary coils of a transformer</td>
</tr>
<tr>
<td>$f_L$</td>
<td>Lower resonance frequency of a transformer-based resonator</td>
</tr>
<tr>
<td>$f_H$</td>
<td>Higher resonance frequency of a transformer-based resonator</td>
</tr>
<tr>
<td>RTPS</td>
<td>Reflection type phase shifters</td>
</tr>
<tr>
<td>$\lambda/4$</td>
<td>Quarter wavelength</td>
</tr>
<tr>
<td>FR-LNA</td>
<td>Frequency reconfigurable LNA</td>
</tr>
</tbody>
</table>
List of original publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals I–VIII.


# Contents

**Abstract** ................................. 9  
**Tiivistelmä** ................................. 9  
**Preface** .................................. 11  
**List of abbreviations** ................. 15  
**List of original publications** ........... 17  
**Contents** ................................. 19

## 1 Introduction .......................... 19

1.1 Background ................................ 19  
1.2 Objectives and research targets .......... 20  
1.3 Contributions in original publications .... 22  
1.4 Organization of the thesis ................ 23

## 2 Integrated circuit design techniques for millimeter-wave applications .......... 25

2.1 Overview .................................. 25  
2.2 Phased arrays ............................. 26  
2.3 Phased array receiver ..................... 27  
2.3.1 Bandwidth limitation of a phased array receiver system ........ 29  
2.3.2 SNR improvement in a receiver phased array system .......... 30  
2.3.3 Linearity in phased array receiver .......... 31  
2.4 Beamforming architectures ............... 32  
2.4.1 Digital beamforming .................... 32  
2.4.2 RF beamforming architecture .......... 33  
2.4.3 LO beamforming architecture .......... 33  
2.4.4 IF beamforming architecture .......... 33  
2.4.5 Hybrid beamforming architecture ....... 33  
2.5 A 15 GHz phased array receiver front-end .... 38

## 3 Layout techniques for active and passive devices for mmWave circuits .......... 49

3.1 Overview .................................. 49  
3.2 Silicon-based technologies ............... 49  
3.3 Low-loss substrate processing techniques .... 50  
3.3.1 RF CMOS SOI ........................... 51  
3.3.2 Technology aspects adapted in this thesis ... 52  
3.4 Active devices ............................ 52  
3.4.1 Intrinsic and extrinsic models .......... 53  
3.5 Design and characterization of cascode devices ................ 55
3.6 Passive devices .......................................................... 58
  3.6.1 Magnetic field losses ............................................. 59
  3.6.2 Electric field losses ............................................. 60
  3.6.3 Skin effect ....................................................... 60
3.7 Electromagnetic modelling ........................................... 61
  3.7.1 Definition of RF ground ......................................... 62
  3.7.2 Local metal layer as RF ground ................................. 62
  3.7.3 PCB as RF ground ............................................... 63
  3.7.4 Effect of the ground ring on inductor performance .......... 64
  3.7.5 Design methodology for mmWave circuits ..................... 65
3.8 Transformers as resonators .......................................... 69
3.9 Design and characterization of a differential reflection-type phase
  shifter ................................................................. 70
3.10 Active and passive signal combiners for mmWave phased array
  systems ............................................................... 74
3.11 Conclusion ........................................................... 81

4 Low noise amplifiers for mmWave applications 83
  4.1 40 GHz LNA .......................................................... 83
  4.2 Millimeter wave frequency reconfigurable LNAs .................. 86
    4.2.1 Frequency reconfigurable LNA1 ............................... 87
    4.2.2 Frequency reconfigurable LNA2 ............................... 90
    4.2.3 Comparison of different topologies ......................... 93

5 Conclusions ............................................................. 95
References ............................................................... 99
Original publications .................................................. 109
1 Introduction

1.1 Background

Industrial evolution during the past two decades has paved the way for an enormous amount of multimedia applications, such as video-on-demand, live streaming, internet-of-things, etc, in the consumer market. Furthermore, with the ever increasing trend in the number of mobile devices, there is a demand for increased data rates in wireless mobile networks. Existing wireless standards (sub-6 GHz frequency bands) are not enough to cope with the demand from the increasing number of multimedia mobile users. The data rate is proportional to the frequency bandwidth, which is a limited resource in existing mobile communication technologies. However, an abundance of bandwidth is available around multiple frequencies in a centimeter wave (cm-wave) and a milli-meter wave (mmWave) frequency spectrum [1]. Use of mmWave frequency bands for mobile communications has evolved into the fifth generation for mobile communication (5G). Previously, mmWave frequencies (30 GHz to 300 GHz) have been used for radio astronomy, military and automotive radar applications. The smaller wavelengths of a carrier frequency at mmWave frequencies introduces a higher free space path loss. This can only partially be compensated by increasing the antenna aperture size and using phased arrays [2], for example. However, this is mostly not sufficient and beamsteering in a transmitter or receiver are therefore inevitable in 5G mmWave communication systems.

Due to the massive amount of antenna feeds, 5G mmWave phased array systems require a higher integration of the system that helps to reduce the size, cost, complexity and faster productization. Developments in the integrated circuit (IC) design technologies during the past 40 years have created opportunities for the seamless integration of complete radio systems in a single chip. Recent trends towards the miniaturization of devices have demanded the integration of radio front-end and baseband digital circuitry on a single die. This requires a process technology which offers faster devices for digital circuits and also provides low loss passive devices for RF/mmWave front-end circuit blocks. Compound semiconductor technologies such as gallium arsenide (GaAs) offer very good performance in terms of output power and noise at very high frequencies. However, due to low yield, higher cost and lack of dense digital circuitry, they are not suitable for high volume commercial production. Silicon germanium (SiGe) based bipolar complementary metal oxide semiconductor (BiCMOS) technology is becoming a popular choice for mmWave systems due to comparatively low cost, high yield and
reasonable performance of high frequency circuits. Developments in silicon (Si) based Complementary Metal Oxide Semiconductor (CMOS) technologies have significantly shrunk the transistor’s gate length to realize faster devices for digital circuitry. However, inadequate RF modelling and substrate related losses limit their performance at mmWave frequency operations. Contrarily, CMOS Silicon on Insulator (SOI) technology has advantages of high resistive low loss substrate for high performing integrated passive structures. Additionally, the availability of faster devices for digital circuitry and accurate RF models have made it a very promising choice for integration of mmWave front-end and baseband blocks of a wireless system on a single die. Multiple mmWave frequency bands for 5G systems ranging from 24 GHz to 40 GHz, have been selected for network deployments in different regions across the globe [1]. With the emergence of multi-band paradigm in 5G mmWave systems, a serious research topic has evolved to investigate the high level integration of a frequency reconfigurable front-end design.

1.2 Objectives and research targets

Silicon-based mmWave phased arrays have been utilized for many applications, for example, car radars, motion sensors, etc. However, the introduction of a mmWave spectrum for commercial communication systems (5G new radio (5G-NR)) has lead to extensive research in this domain [3, 4, 5, 6, 7, 8, 9]. A silicon chip supporting multiple channels having gain and phase control in each channel increases the layout complexity by many folds due to the complex RF routing. One of the scopes of this thesis is to study such integration at centimeter-wave as well as mmWave frequencies.

The primary focus of this thesis was to study and develop methods and a way of working (WoW) for the design of complex circuits for mmWave ICs. The size and complexity of RF front-end ICs for 5G mmWave systems grow considerably higher than Radio frequency integrated circuits (RFICs) for previous generation communication systems, i.e, fourth generation long term evolution (4G LTE). This is because of the multiple channels which are integrated in a single IC to support the beamforming capabilities of the system. By studying the methods and processes involved in the development of such integration at mmWave frequencies by the accurate design and modelling of active and passive devices, we can understand the layout effects of these devices towards optimum performances, which was part of the major objectives of this thesis.

Another key challenge of the 5G mmWave systems is the existence of more than one frequency band for communication, which results in an increased amount of design effort to design the front-end IC modules each for a separate frequency band. Similar to
prior generation multi-band communication systems [10, 11], there would be a clear need for a single FE hardware solution covering all the possible 5G frequency bands. One solution is to develop a wideband FE system covering the mmWave frequency bands of the 5G-NR standard [1]. Nevertheless, this also results in the receiver FE receiving out of band high power unwanted blocker signals which may compress the FE blocks upon reception and results in the distortion of the desired communication. To avoid this problem, a frequency reconfigurable FE module can be useful to efficiently filter the out of band blocker signals. One of the primary objectives of this thesis was to study the possible solutions for frequency reconfigurable receiver FE modules for 5G mmWave systems.

Some of the principal research targets are listed as follows:

1. Analyse the characteristics of various beamforming architectures and their key building blocks, such as, LNAs, phase shifters, signal combiners, etc.

   This research target was set at the start of the research to understand different architectures supporting the beamforming capabilities of a receiver system and their analysis in terms of power dissipation, area, and implementation complexity. Moreover, understanding the key building blocks and their target specifications in a phased array system was part of this target.

2. Design a fully integrated solution of a receiver phased array FE system including the packaging.

   At the early stage of the research, it was decided to design and implement a prototype of a phased array FE receiver system at a rather lower center frequency compared with the 5G frequency bands (i.e., 15 GHz). One of the main reasons for this was to gain a sufficient understanding about the complete IC design flow for implementation before moving up to mmWave frequency designs to solve more complex problems.

3. Understand the main limitations to achieving the optimum performance from the silicon-based active and passive devices.

   It is always beneficial to grasp a necessary knowledge of performance optimization techniques in terms of layout implementation for active and passive devices. This target helped to understand the key elements affecting the performance of a silicon-based device (active and passive) at high frequencies.

4. Accurate electromagnetic modelling for the passive devices together with their associated ground reference, $rc$-extraction for active devices and their interconnections in order to reduce the discrepancy between simulated and measured data.

   With the increase in operating frequency, modelling of the silicon-based passive devices with and without ground plane (underneath and around the device) is
inevitable. Another main objective of this thesis was to develop an appropriate methodology for the suitable modelling techniques for different sections of a mmWave circuit.

5. Analyse and design of the mmWave receiver FE modules for frequency reconfigurable system blocks, such as LNAs.

1.3 Contributions in original publications

This dissertation is based on three journal articles and five conference papers. The author of the thesis had the main responsibility of developing the ideas, their implementation, evaluating the performance results, and writing the papers [I, II, III, IV, V, VI, VII and VIII], some of which have been a joint effort of a large group of researchers. Alok Sethi designed the bias circuit and Olli Kursu designed and provided the digital circuitry for all the fabricated chips. The co-authors have provided constructive criticism on ideas, guided in their implementation, and provided important comments on the writing. In papers [I] and [II], the thesis author was mainly responsible for the receiver front-end design, including the design of the LNA, vector modulator phase shifter, combining network, mixer and baseband amplifiers. Alok Sethi contributed to the brainstorming of the system concept, while Rehman Akbar designed the 30 GHz local oscillator (LO) distribution network for the downconversion mixers and performed the top level integration of the complete chip. In addition, Risto Vuohtoniemi offered his support in the characterization and measurements of the test circuits.

In order to accomplish the key objectives of this thesis, several IC prototypes were developed based on the design methodologies for LNAs, phase shifters, variable inductors, a coupled line quadrature coupler and a phased array front-end system, etc. Deep submicron (45 nm) CMOS SOI technology was used to design and fabricate the above mentioned circuit blocks. The comprehensive outcome of all these design achievements is published in I to VIII. A brief description of the most significant contributions of this thesis to the existing research is given as follows:

1. A 15 GHz four-channel phased array receiver front-end was designed supporting two MIMO channels. The test prototype was fabricated and measured to validate the proposed approach. The single chip offers eight antenna feeds supporting two sub-arrays with four antenna ports in each sub-array. Each sub-array downconverts the combined signal from 15 GHz to baseband using external LO. This system offers the opportunity to further process two baseband signals externally in order to implement hybrid beamforming. The implemented circuit showed reasonable
gain and noise performance at center frequency when measured in single channel configuration. Details can be found in publications I and II.

2. A multi-gate (MG) style layout for cascode-connected NMOS devices is proposed towards optimum device performance and compared with conventional style layout at mmWave frequencies. It was observed that the gate resistance of the MG device is smaller than that of a conventional style device. However, additional wiring used in the MG device also increases the gate capacitance that reduces its performance at high frequencies. A prototype was fabricated together with a conventional style layout and key parameter results were compared. A 40 GHz mmWave LNA was designed using this multigate layout technique. Details can be found in publications VI and VII.

3. A co-planar waveguide (CPW) based differential quadrature coupler was designed and fabricated using vertical coupling between two lateral metal lines, which offers insights into the implementation of differential coupler designs with a reduced area footprint. Details can be found in publication IV.

4. Three types of signal combiners for phased array applications were analyzed and simulated. A comparison of power combiners is provided on the basis of the drawn area, insertion loss, isolation between two channels and complexity is presented. A compact power combiner based on transformers is presented and compared with a conventional Wilkinson power combiner. A transformer-based power combiner exhibits lower IL with reduced isolation when compared with Wilkinson combiner in the same technology. Details can be found in publication V.

5. Two mmWave frequency reconfigurable LNAs were designed based on a magnetically tuned variable inductor. The proposed circuit blocks were analysed and fabricated for complete validation, which provided new insights on the design of tunable structures and their impacts on the signal and noise performance of the circuits in mmWave frequencies. Details can be found in publication VIII.

1.4 Organization of the thesis

This thesis is divided into two parts. In the first part, a brief introduction of the research area is presented. An outline of the main contributions by the author is also summarised in this part. In the second part of the thesis, a summary of the original publications I-VIII is compiled.

Chapter 2 consists of a concise introduction of the research field, which includes the fundamental concepts of the key architectures of the phased array systems. Moreover, this chapter presents the design and characterization of an 8 antenna phased array
receiver front-end IC supporting two MIMO channels with four antenna ports in each sub-array. This IC is implemented using 45 nm CMOS SOI technology.

Chapter 3 summarizes the layout of silicon-based active and passive devices for mmWave applications using 45 nm CMOS SOI technology. These devices include: 1) cascode connected active devices; 2) a differential reflection-type phase shifter based on a vertically-coupled line coupler; 3) a comparison study of the different types of active and passive signal combiners for phased array applications.

Chapter 4 provides insights into the design and validated results of mmWave LNAs. These LNAs include the design of: 1) a 40 GHz LNA implemented in 45 nm CMOS SOI technology, and 2) two frequency reconfigurable LNAs covering 5G NR standards, also implemented in 45 nm CMOS SOI technology.
Integrated circuit design techniques for millimeter-wave applications

This chapter provides an introduction to integrated circuit design techniques for mm-wave front-end designs. The basis to understanding the research topics in this thesis is presented in this chapter. An overview of the phased array receiver systems is given in the start of the chapter followed by the types of phased array architectures and their use cases. Finally, the design of a fully integrated 15 GHz four-element phased array supporting two baseband channels (articles I, III and II) is presented.

2.1 Overview

Existing mobile communication systems in fourth generation (4G) cover frequency bands from 450 MHz to 3600 MHz using a scalable bandwidth from 1.4 MHz to 20 MHz [12]. Preferably, an Orthogonal Frequency Division Multiple Access (OFDMA) technique is used to distribute the resources based on the demand [13]. Introduction of the advanced version of 4G (LTE-Advanced or LTE-A), which allows aggregated bandwidth of 100 MHz and 128 quadrature amplitude modulation (QAM), enables download data rates of more than 3 Gbps in ideal conditions [14]. Advances in multimedia applications during the past decade, such as, Internet of Things (IoT), Video on Demand (VoD), live surveillance and streaming, etc, have increased the demand of data rates many times higher than what current technologies can offer with existing infrastructure. Unfortunately, the situation gets worse in denser environments where multiple users are streaming videos simultaneously. Therefore, a new communication system is required in the near future to fulfill the growing demands of the wireless network.

Advanced digital modulation techniques such as 64, 128 or 256 QAM are the key enablers to achieve higher data rates for a given bandwidth. In order to realize these modulation techniques, higher values of signal-to-noise ratio (SNR) are required as discussed in [15, 16, 17]. A higher channel BW allows a higher bitrate for a lower order modulation and reduces the SNR requirements. Furthermore, not enough channel BW is available in currently used frequency bands for mobile communication. Inevitably, new frequency bands are required with more bandwidth (few 100s of Megahertz) available. In mmWave frequency spectrum from 24 GHz to 100 GHz, several frequency bands are available with BW in the order of 1 GHz. Therefore, mm-wave communication is one of the core focus areas towards the implementation of future mobile communication.
systems, including 5G for mobile communication. Three frequency bands between 24 GHz to 40 GHz are already standardized by the 5G-NR standard, as shown in Table 1.

### Table 1. Allocated millimeter-wave frequency bands in 5G NR FR2 standard [1].

<table>
<thead>
<tr>
<th>5G NR FR2 Bands</th>
<th>Center Frequency</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>n257</td>
<td>28 GHz</td>
<td>3 GHz</td>
</tr>
<tr>
<td>n258</td>
<td>26 GHz</td>
<td>3.250 GHz</td>
</tr>
<tr>
<td>n260</td>
<td>39 GHz</td>
<td>3 GHz</td>
</tr>
</tbody>
</table>

#### 2.2 Phased arrays

One of the many barriers for mmWave frequency communications are losses due to atmospheric absorption and free space path loss (FSPL) because of decreasing wavelengths [18] and their radiator area. FSPL is a function of wavelength ($\lambda$) and distance ($d$) between receiving and transmitted antennas given as follows [19]:

$$FSPL = \frac{4\pi d}{\lambda}.$$  

(1)

Equation 1 shows that the path loss per antenna element increases by moving higher in frequencies (mmWave). However, this equation holds true in the condition when the antenna gain is kept constant over the frequency range [2, 18]. The path loss can be compensated by keeping the physical area of antenna, i.e, the aperture, constant over the frequency range. Since the size of a single antenna element is proportional to the wavelength ($\lambda$) of the frequency, a multiple element system such as phased array can be used to keep the effective antenna area ($A_{eff}$) constant over frequency. For example, an $N$-element array has an effective area given as

$$A_{eff} = N(A_{element}),$$  

(2)

where $A_{element}$ is the area of a single element. Assuming an array consisting of planar patch antenna elements with the spacing between the elements of $\lambda/2$, the area of each element is then $(\lambda/2)^2$, with the array gain is given as [13]:

$$G_{array}(dB) = 10log_{10}(N) = 10log_{10} \frac{A_{eff}}{A_{element}} = 10log_{10} \frac{A_{eff}}{(\lambda/2)^2}.$$  

(3)

When FSPL is added to the array gain, we will have:

$$G_{array}(dB) - FSPL(dB) = 10log_{10} \frac{A_{eff}}{(\lambda/2)^2} - 20log_{10} \frac{4\pi d}{\lambda},$$

$$= 10log_{10}(A_{eff}) - 20log_{10}(2\pi d).$$  

(4)
This means that the propagation loss becomes independent of frequency and the array gain of a phased array with its effective area compensates the path loss at any distance.

A large antenna aperture area results in the increase in antenna gain, i.e., directivity. This loss of omnidirectivity (an even or almost even distribution of the signal to all the specified directions) requires antenna beam steering capabilities to find and maintain the radio link and to avoid interference from dynamic objects in the channel [20].

A phased array is an electronically controllable antenna system which steers the direction of the incident/transmitted beam to improve the directivity of the communication system. The beam steering property of phased arrays also enables the spatial or directional filtering from the interferers in a direction other than the incident signal by using proper control. The key principle for a beamforming phased array system is shown in Figure 1. The beam front of a directional transmit beam with an angle \( \phi_{in} \) is received by a set of receiver antennas. For example, if each antenna receives a far field signal with an incident angle \( \phi_{in} \), the signal received by the individual elements will have a delay from the previous element. This delay depends on the distance between the two elements. This difference in delay in the received signals can be compensated by introducing an additional delay to the received signal, resulting in the similar in phase signal from each element before combining. As a result, the signals combine in a constructive manner and the combined signal is increased.

![Figure 1. Key principle of a phased array system (Modified from Arun N. [21]).](image)

### 2.3 Phased array receiver

Figure 2 demonstrates an \( n \) antenna element phased array and an incident wave with angle \( \phi \). The spacing between two successive elements is \( d \). The incident wave covers
an additional distance of $ds\sin\phi$ between two successive elements which introduce a time
delay of $\tau = \frac{ds\sin\phi}{c}$, where $c$ is the speed of light. Signals received at the first element
and a delayed signal at the $m$th element are given by,

$$S_m(t) = A(t)\cos[\omega_c t + \phi(t)],$$
$$S_m(t - m\tau) = A(t - m\tau)\cos[\omega_c t - m\omega_c\tau + \phi(t - m\tau)],$$

where $S_m(t)$ and $A(t)$ represent the signal and its amplitude, respectively, and $\omega_c$ is the
angular frequency of the transmitted signal, i.e., $2\pi f$. This progressive delay between
elements can be compensated by "steering" the antennas towards the direction of the
incident beam, i.e., $\theta$. This is achieved by introducing the equivalent time delay ($\tau_m'$) in
each element, as shown in Figure 2. After combining the delay compensated signals, the
output signal is given as:

$$S_{out}(t) = \sum_{m=0}^{n-1} S_m(t - \tau_m').$$

(6)

When the amount of delay is completely compensated by adjustable delay element,
i.e., $\tau_m' = -m\tau$, the total output signal is given by:

$$S_{out}(t) = nA(t)\cos[\omega_c t + \phi(t)].$$

(7)
In other words, the captured output signal power of a phased array is increased by a factor of \( n \) when the signal delay in each element is compensated and the compensated signals are added together. This property of a phased array system is characterized by the array factor (AF), which is a measure of the additional power gain of the phased array system compared to that of a single channel and is given by [23]:

\[
AF = \left( \frac{\sin(N(\omega \Delta \tau - \frac{2\pi f}{c} \sin \phi_{in}))}{\sin(\omega \Delta \tau - \frac{2\pi f}{c} \sin \phi_{in})} \right)^2.
\]  

(8)

AF in equation 8 provides the array gain depending on the value of the difference between the various paths in the antenna array, i.e., \( \Delta \tau \). A maximum array gain of \( N^2 \) is achieved when \( \Delta \tau = \frac{dsin \phi}{c} \), which means that the array gain is smaller for other values of \( \Delta \tau \) or, in other words, signals from other directions are spatially filtered by a phased array [23]. Phased array is a reciprocal system, i.e., a similar concept is valid for both receiver and transmitter.

2.3.1 Bandwidth limitation of a phased array receiver system

The amount of time delay or phase shift required to steer the beam of a phased array system to a desired angle can be implemented by either a true time-delay unit or a phase shifter. A phase shift can be formulated from a desired time delay as:

\[
\Delta \phi = \omega \Delta \tau = 2\pi f \Delta \tau = 2\pi f \frac{dsin \theta}{c} = \frac{2\pi sin \theta}{\lambda}.
\]

(9)

Equation 9 shows that a desired phase shift \( \Delta \phi \) to steer the beam to angle \( \theta \) has a dependency on the wavelength of the signal (\( \lambda \)) or, in other words, a phase shifter is a frequency dependent component which has a limited operational bandwidth. This frequency dependency of a phase shifter can result in an error in the scan angle of the beam, which is also known as a beam squinting error [24, 25]. This is a problem for wide bandwidth receivers and transmitters, for which there is a wide difference between the lowest and highest frequency of interest. If a phased array antenna system has a beam steered to angle \( \theta_o \) at frequency of \( f_o \), at an offset frequency of \( f_o + \Delta f \), the beam angle is also changed to \( \theta_o + \Delta \theta \). This error increases with the increase in change in frequency. On the other hand, the operation of a true time-delay is frequency-independent and can be used for controlling the time delay for a wide range of frequencies. However, its practical implementation is often more complex.

29
2.3.2  **SNR improvement in a receiver phased array system**

A two-port wireless receiver system is mainly characterized by its output signal-to-noise ratio \((SNR_{out})\) which is always lower than its input signal-to-noise ratio \((SNR_{in})\). The ratio between input and output SNR of a system is defined as the noise factor \((F)\) of the system and is given as:

\[
F = \frac{SNR_{in}}{SNR_{out}},
\]

**(10)**

However, equation 10 is not valid for multi-port system such as phased arrays. An \(n\)-element phased array receiver system is shown in Figure 3, where one can note that element receives part of a signal with progressive phase difference depending on the time delay. At the output of a single channel, signal \((S_{o,\text{path}})\) and noise \((N_{o,\text{path}})\) are defined as follows:

\[
S_{o,\text{path}} = S_{in,\text{path}}G_{\text{amp}}, \quad N_{o,\text{path}} = N_{\text{amp}} + N_{in,\text{path}}G_{\text{amp}}, \tag{11}
\]

where \(G_{\text{amp}}\) and \(N_{\text{amp}}\) are the gain and noise of the amplifier in the channel, respectively. From equation 11, output SNR can be derived as,

\[
SNR_{o,\text{path}} = \frac{S_{o,\text{path}}}{N_{o,\text{path}}} = \frac{S_{in,\text{path}}G}{N + N_{in,\text{path}}G} \tag{12}
\]
According to the definition of AF in equation 8, if input signals are added in the phase, (or coherently) then the output signal is given as [22],

\[ S_{o,\text{array}} = n^2 S_{\text{in,path}} G, \]  

(13)

where \( n \) is the number of parallel signal paths in a phased array. Assuming that there is no correlation between each antenna terminal, the noise contribution by each element is added out of the phase or non-coherently, and is given as:

\[ N_{o,\text{array}} = n(N + N_{\text{in,path}} G), \]  

(14)

where \( N \) is the internal noise of each parallel branch. Using equation 13 and 14, the array output SNR (\( \text{SNR}_{o,\text{array}} \)) is as follows:

\[
\text{SNR}_{o,\text{array}} = \frac{n^2 S_{\text{in}} G}{n(N + N_{\text{in,path}} G)},
\]

(15)

Using equations 11 and 15, following conclusion can be made:

\[
\text{SNR}_{o,\text{array}} = n\text{SNR}_{o,\text{ch}}.
\]  

(16)

Equation 16 implies that the output SNR of a receiver array system (\( \text{SNR}_{o,\text{array}} \)) improves by a factor of \( n \) from the output SNR of a single channel (\( \text{SNR}_{o,\text{path}} \)) in the array when the phases in the various paths are set correctly such that the recombination is optimal. In other words, the sensitivity of a phased array system is improved by \( 10\log(n) \) when the input signals at each antenna element are combined in phase.

### 2.3.3 Linearity in phased array receiver

The linearity of a conventional two-port receiver front-end system is characterized by two main methods, i.e., 1) the compression point and 2) inter-modulation products from the non-linear behaviour of transistors in the system. Several techniques have been proposed in the literature to address the non-linear behaviour of the system, for example, mixer first or LNA-less receiver [26] and blocker-tolerant RF FEAs [27, 28]. The linearity of a single channel in a multi-port FE system such as phased array also has similar characteristics and is addressed accordingly. Due to a free space path loss at mmWave frequencies, the typical power of a received signal at a single antenna element is at such a small level that it does not harm the linearity of the system. Furthermore, the ability of...
spatial and spectral filtering in mmWave phased arrays reduces the power of potential blockers and thus limits the linearity requirements[8].

A careful link-budgeting of the system (as in [21]) is required in order to specify the gain, noise, and linearity of a single block in the signal chain to keep the the overall performance of the system within an acceptable range. For example, if the cascaded voltage gain from a single chain is high, it may result in a large signal (due to an array gain) at the output of the signal combiner which may saturate the input stage of the mixer. Alternatively, if there is not enough gain at any particular stage in the signal chain, it may fall below the noise floor which can be difficult to recover.

2.4 Beamforming architectures

Various beamforming architectures are used depending on the position of phase shifting within the signal chain, combining or splitting of the signals, mixing of the signal, and the number of mixer stages. The complexity, area and power consumption of an RF system increases with parallel baseband streams. The gain partitioning of the building blocks in each signal chain is important to optimize the total noise Figure, linearity, area, and power consumption of the system. For example, passive phase shifters exhibit more losses (noise Figure) in the signal path with additional penalty of area; however, they are linear with zero power consumption [29, 30]. Alternatively, active-based phase shifters are used because of their smaller area footprints and phase shift with fine steps at the cost of power consumption and linearity [31, 32, 33, 34, 35, 36]. Similarly, passive and active techniques are used in signal combining in phased arrays. Passive combining is usually based on distributed Wilkinson power combiners due to their better matching and isolation characteristics [37] [38, 39]. However, they appear to be the largest structure in the IC amongst other building blocks. To reduce the area of Wilkinson combiners, an alternative approach based on lumped-elements is adopted in [40]. The aforementioned passive combiners add extra loss to the system, mostly due to the ohmic loss of the metal layers. Contrarily, a current-based combining approach used in [41, 42, 43] offers smaller footprints without adding additional losses to the system.

2.4.1 Digital beamforming

Beamforming operations such as phase shifting and signal combining, in digital beamforming architecture (Figure 4) are performed in digital processing blocks. However, this architecture requires the same amount of baseband data streams as the number of antennas which leads to increased complexity in converting an RF signal.
to baseband. The increased number of analog-to-digital converters burn a huge amount of energy to convert each analog signal into the digital bits. Additionally, the digital processing of these bits requires more power.

2.4.2 RF beamforming architecture

In this architecture, phase shifting and combining is performed in the RF/mmWave domain, as shown in Figure 5. This RF beamforming architecture uses only one baseband (BB) data stream for a given set of antenna elements resulting in a power and area efficient architecture.

2.4.3 LO beamforming architecture

In this architecture, the phase shifting to the received signal is achieved through the local oscillator signal in the downconversion stage [5, 6], as shown in Figure 6. The downconverted phase shifted signals are then combined in the intermediate frequency (IF) domain. This technique reduces the number of mmWave blocks in the signal chain, but at the same time increases the power consumption and complexity of the LO routing to all the elements.

2.4.4 IF beamforming architecture

Phase shifting and combining is performed after the down-conversion of the received signals, as shown in Figure 7. Similar to LO beamforming, IF beamforming also requires as many mixers as the antenna elements [44, 45]. As all phases of the mixers are equal, LO routing is less complex in IF beamforming than in LO beamforming.

2.4.5 Hybrid beamforming architecture

An efficient distribution of the signal processing between an RF domain and digital domain can benefit from reduced complexity, power and cost of a beamforming system. For example, a single array can be split into multiple sub-arrays in the RF domain. Phase shifting and combining of each sub-array is first performed in the RF domain. These multiple sub-array signals are then carried to the baseband circuitry for further processing in the digital domain. This type of architecture is known as hybrid beamforming architecture (Figure 8).
In a multi-user environment, spatial multiplexing can be used to separate multiple data streams (RF to BB) from the same receiving antenna aperture (phased array antennas). This is performed by either digital beamforming (DBF) or using hybrid beamforming (HBF) architectures. Digital beamforming is not considered very efficient in terms of power consumption because of parallel RF to BB chains, including the first parts of digital processing block. On the other hand, hybrid beamforming offers benefits of spatial filtering with less complexity and power consumption.

Hybrid beamforming is further characterized in three different architectures, i.e, partially-connected HBF, fully-connected HBF, and hybrid-connected HBF [42]. In a partially-connected HBF architecture, an antenna sub-array is connected to a single baseband stream. This architecture limits the array gain for each stream to the associated sub-array. On the other hand, in a fully-connected HBF architecture, each antenna is connected to all the data streams and results in an improved spectral efficiency for the same number of antennas at the cost of complexity. This complexity increases with the increasing number of BB streams. In [42] and [41], a hybridly-connected beamforming is proposed to reduce the complexity of fully-connected HBF architecture by using sub-arrays with limited BB streams of fully-connected beamformers, as shown in Figure 8.

5G mmWave communication systems for base stations have stringent requirements to achieve high data rates (at least 1 Gbps) for broad range of users. Different types of beamforming architectures are combined in order to address power, heat dissipation and area specifications of these large antenna panels. This requires a high level of integration of phased array ICs which can be reused towards a larger system implementation controlling hundreds of antenna modules. A single wafer-scale system supporting a large number of antennas as in [50], improves the integration. However, this type of architecture only supports a single baseband stream (down-conversion or up-conversion). Multiple simultaneous baseband streams, i.e, Multi-User Massive MIMO, are required in a complex and dense user environment having multiple scatterers. Additionally, an RF beamforming architecture supporting multiple MIMO streams can also benefit in enhancing the link quality and relaxing the higher SNR requirements compared to a single user case [15, 16, 17]. Furthermore, an antenna with two polarization elements can be used to feed two orthogonal channels from two separate MIMO streams [47] which helps in reducing the footprints of an antenna panel to half.
Table 2. Performance comparison of state of the art beamformers.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Architecture</th>
<th>Freq. (GHz)</th>
<th>Gain (dB)</th>
<th>Elements</th>
<th>Tech.</th>
<th>Area* (mm²)</th>
<th>DC Power* (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JSSC’17 [39]</td>
<td>RF</td>
<td>28-32</td>
<td>10.5</td>
<td>4</td>
<td>130 nm SiGe</td>
<td>0.3</td>
<td>136</td>
</tr>
<tr>
<td>RFIC’16 [46]</td>
<td>RF</td>
<td>26-28</td>
<td>12.2</td>
<td>4</td>
<td>45 nm CMOS SOI</td>
<td>1.24</td>
<td>42</td>
</tr>
<tr>
<td>EuMW’19 [4]</td>
<td>RF</td>
<td>28</td>
<td>15</td>
<td>8</td>
<td>55 nm CMOS</td>
<td>43</td>
<td>-</td>
</tr>
<tr>
<td>ISSCC’20 [8]</td>
<td>IF</td>
<td>28</td>
<td>&gt;16</td>
<td>4</td>
<td>65 nm CMOS</td>
<td>0.66</td>
<td>28.1</td>
</tr>
<tr>
<td>JSSC’17 [47]</td>
<td>Sliding IF</td>
<td>28</td>
<td>28-34</td>
<td>16×2</td>
<td>180 nm SiGe</td>
<td>-</td>
<td>206</td>
</tr>
<tr>
<td>JSSC’11 [48]</td>
<td>Sliding IF</td>
<td>60</td>
<td>70</td>
<td>16</td>
<td>120 nm SiGe</td>
<td>-</td>
<td>56</td>
</tr>
<tr>
<td>JSSC’04 [5]</td>
<td>LO</td>
<td>24</td>
<td>43</td>
<td>8</td>
<td>180 nm SiGe HBT</td>
<td>11.5</td>
<td>910</td>
</tr>
<tr>
<td>RFIC’17 [49]</td>
<td>Hybrid</td>
<td>28</td>
<td>25</td>
<td>4×2</td>
<td>28 nm CMOS</td>
<td>7.28</td>
<td>400</td>
</tr>
<tr>
<td>JSSC’18 [42]</td>
<td>Hybrid</td>
<td>25-30</td>
<td>34</td>
<td>8</td>
<td>65 nm CMOS</td>
<td>6.16</td>
<td>27</td>
</tr>
<tr>
<td>ISSCC’20 [9]</td>
<td>Hybrid</td>
<td>28/37</td>
<td>&gt;16</td>
<td>4</td>
<td>65 nm CMOS</td>
<td>0.66</td>
<td>28.1</td>
</tr>
</tbody>
</table>

* Per element, **complete array, ^ single TRX
Fig. 4. Digital Beamforming architecture (Modified from Xiang G. 2004 [22]).

Fig. 5. RF Beamforming architecture (Modified from Donghyup S. 2011 [38]).

Fig. 6. LO path Beamforming architecture (Modified from Xiang G. 2004 [22]).
Fig. 7. IF Beamforming architecture (Modified from S. Kishimoto 2009 [44]).

Fig. 8. Hybridly-connected hybrid beamforming architecture (Modified from S. Mondal 2018 [42]).
2.5 A 15 GHz phased array receiver front-end

At the starting phase of this thesis (2015), many frequency bands above 10 GHz were being studied and considered to become the potential candidates of a frequency spectrum for 5G systems above 6 GHz. Frequency bands with an available bandwidth of around or more than 500 MHz were vacant at 15 GHz, 24 GHz, 26 GHz and 28 GHz, 40 GHz, etc. Therefore, it was decided to develop the first phased array prototype at 15 GHz because of the following two main reasons:

1. a signal wavelength in the range of centimeters leads to less complexity in circuit designs than the mmWave frequencies that would offer hands-on experience for full chip design flow,
2. almost 1 GHz bandwidth is available around a centre frequency of 15 GHz that could be selected as one of the potential future 5G bands.

After the World Radio Communication conference at the end of 2015, it is now obvious now that the minimum 5G frequency band beyond 6 GHz is at 24 GHz and there is essentially no 5G system application at 15 GHz. However, many lessons learned in developing a phased array system and its associated blocks at 15 GHz are applicable to the systems at rather higher frequencies.

Classical research flow in RFIC designs typically starts with the development of basic building blocks towards the design of a more complex system and this can be termed as the bottom-to-top approach. However, this thesis follows the top-to-bottom approach, where efforts were made first to find the solution of a more complex problem and then the conclusions were drawn based on the lessons learnt. This approach is expensive in terms of chip fabrication costs against the success ratio of the manufactured design. However, it does help those researchers with fresh or no-experience to build their knowledge by implementing big systems on their own. It also helps to understand the key problems and find solutions towards the final implementation of a more sophisticated system.

This section presents the design of a fully integrated four-element phased array receiver front end supporting simultaneously two MIMO channels. Publications I, II and III present a silicon-based integrated solution for a phased array receiver supporting two MIMO channels in a single die. The front-end is designed using 45 nm CMOS SOI technology by Global Foundries. The received signal at 15 GHz is down-converted to baseband balanced IQ signals using 30 GHz local oscillator with the help of on-chip divide-by-two dividers. The block diagram of the system is shown in Figure 9. A common-source inductively degenerated cascode LNA circuit topology was selected as the first stage of the two-stage single-ended to differential LNA, as shown in Figure 10 a.
A by-pass antenna switch was implemented to bypass the first stage of the LNA to implement a lower gain mode for the receiver front-end. The input transistors M1 and M2 (Figure 10a) are optimized and biased for the lowest noise figure. A resonance tuning circuit composed of PMOS switches and unit capacitors is used to tune the resonance of the inductive load. The second stage of the LNA is composed of an active single-ended to differential (S2D) balun based on current-reuse LNA topology (Figure 10b). The by-pass switch of the first stage has a significant impact on the output impedance of the first stage in both operational modes which affects the inter-stage matching between the two stages. The common-gate topology of the S2D block to match its input impedance for the wideband of input frequencies helps to mitigate the effect of this impedance mismatch due to the functionality of the by-pass switch. The two-stage LNA is followed by a vector modulator (VM) phase shifter. This has many advantages over the passive phase shifters, such as low loss, smaller area footprints and it offers fine steps for gain and phase tuning. However, power consumption and linearity constraints are some limitations related to VM based phase shifters. The linearity of a VM phase shifter in receiver phased arrays is very important because it has an amplified signal at its input from the preceding amplifying stages, such as in Figure 9. If the VM is not able to withstand to this amplified signal, it can degrade the overall linearity performance of the receiver. Due to its inherent property of achieving fine phase and gain steps across all four quadrants, a VM is used for two operations to perform various beamforming operations (e.g., beam steering, amplitude tapering, etc.), i.e., phase tuning and amplitude tuning. A VM should have similar linearity performance for all the possible phase and gain settings. Contrarily, any error in phase or gain leads to malfunctioning of the beamforming operations. Figure 11 shows the block diagram of the VM and the circuit design of its key building blocks. The virtual natural capacitors are used together with poly resistors to realize smaller and symmetrical design. The layout of the polyphase filter is drawn using dummies of the capacitors and resistors to minimize any process variations during the manufacturing. A current-combining based signal combiner is utilized to combine the signals of all four channels. The block diagram of the signal combiner is shown in Figure 12, which is followed by a current-bleeding Gilbert-cell mixer topology used in this design to improve the third-order linearity of the block (Figure 13). The library models of the process design kit (PDK) for the inductors are used in this design. The signal routes and power supply lines are modeled using rc-extraction.

The fabricated IC is soldered on a test PCB using flip chip solder bumps. During the design process, the input of the LNA was mistakenly left to be matched or designed by using off-chip components or transmission lines which would also compensate
Fig. 9. Block diagram of the receiver front-end (Reprinted, with permission from Publication II © 2018 Springer).
the parasitics of solder bump. However, it was learned that transforming the highly capacitive or reactive input impedance of the input transistor into real impedance should always be done on-chip, because the input pad and solder bump parasitic capacitance is also added to the input resulting in a more capacitive impedance. The transformation of the impedance from a high quality factor or high reactance impedance plane on the smith-chart to the real impedance plane (center horizontal line) is challenging to design and sensitive to the frequency shift. Another challenge was the small pitch between the two adjacent channel input pads, i.e, 190 µm. Considering all the aforementioned challenges, an input matching network was designed with the help of an off-chip non-uniform transmission line (NUTL) structure on the PCB (paper III). The detailed design procedure of this matching structure is shown in Figure 14. The target of this design was to match the highly capacitive impedance (lower half on the smith-chart) for a broad range of frequencies around the center frequency, i.e, 15 GHz. With the help of different impedance transformers between points A to E in Figure 14, a wide-band impedance matching (1 GHz) was achieved at 15 GHz. The design of the complicated 2 × 4 input matching network for the two receiver inputs was the most challenging problem.

The chip micrograph of the phased array front-end is shown in Figure 15. The measurement setup and matching network of a single channel are shown in Figure 16. The measured peak conversion gain from one channel is 23 dB at 14.9 GHz with the input return loss (S11) better than -10 dB at 14.9 GHz, as shown in Figure 17a. Vector modulator phase constellation points are shown in Figure 17b. It was noticed that
Fig. 11. Circuit design of a vector modulator, (a) block diagram, (b) variable gain amplifier, (c) polyphase filter, and (d) quadrant selector switch (Reprinted, with permission from Publication © 2017 IEEE).
these measured points from sweeping VM controls show compression in one quarter. After investigations, it was found that the output impedance of the VM changes with the change in the active control devices in it’s VGAs. This could be mitigated by using more sophisticated design approaches. The linearity measurements show quite low value of IP1dB point (-37 dBm). The main reason behind this limitation is that the cascaded amplifying blocks in the channel were placed without proper gain and linearity partitioning. Bypass mode for amplifiers was designed for the receiver but there were some problems to capture it’s performances properly. Furthermore, during the measurements an unwanted coupling was noticed from channel-to-channel. After investigations, it was found that this coupling was due to the weak isolation between different passive blocks within a channel and also between adjacent channels. This coupling can be avoided by using low ohmic ground shields or ground walls between adjacent blocks to isolate their magnetic coupling between each other. A single channel consumes 86mA of dc current from a 1.1V supply. The single channel measured results
are compared with the state of the art beamformers in Table 3. The performance of this work is compared to the beamformers designed for other frequency bands, due to the lack of available beamformer circuits designed for centimeter-wave frequency range. This work has demonstrated the minimum area when compared with the state-of-the-art designs in Table 3.

Since two separate phased array receiver FEs were integrated in the same die, the fabricated IC can support four antennas for two orthogonal polarizations, i.e, vertical and horizontal, using the same PCB footprints. Additionally, the output signals from each array could be further processed for different beamforming operations such as Inter-beam Interference Reduction, as described in [51] in the analog or digital domain.
Fig. 14. Design procedure of the off-chip input matching network for on-chip LNA (Reprinted, with permission from Publication II ©2018 Springer).
Fig. 15. Micrograph of the chip (Reprinted, with permission from Publication I ©2017 IEEE).

Fig. 16. (a) Measurement setup (Reprinted, with permission from Publication II ©2018 Springer), and (b) input matching feed network for the LNAs (Adapted, with permission from Publication II ©2018 Springer).
Fig. 17. Measurement results: (a) conversion gain (fixed IF=100MHz), (b) constellation points for vector modulator phase sweep, (c) compression points (Reprinted, with permission from Publication I ©2017 IEEE).
Table 3. Performance comparison table of with SOA phased arrays (Reprinted, with permission from Publication ©2017 IEEE).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>This work</th>
<th>[21]</th>
<th>[47]</th>
<th>[31]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freq (Ghz)</td>
<td>15</td>
<td>60</td>
<td>10</td>
<td>28</td>
</tr>
<tr>
<td>Array size</td>
<td>4×2</td>
<td>16</td>
<td>4</td>
<td>16×2</td>
</tr>
<tr>
<td>Phase Shifter</td>
<td>Active</td>
<td>Active</td>
<td>Passive</td>
<td>Passive</td>
</tr>
<tr>
<td>Gain (dB)</td>
<td>23</td>
<td>58</td>
<td>10.1</td>
<td>34</td>
</tr>
<tr>
<td>NF (dB)</td>
<td>5.4a</td>
<td>7.4</td>
<td>3.4</td>
<td>3.7</td>
</tr>
<tr>
<td>1dB Comp. (dBm)</td>
<td>-37</td>
<td>16</td>
<td>-12.5</td>
<td>-22.5</td>
</tr>
<tr>
<td>IIP3 (dBm)</td>
<td>-28</td>
<td>N/A</td>
<td>-4</td>
<td>N/A</td>
</tr>
<tr>
<td>DC Power Consm. (mW)</td>
<td>463b</td>
<td>3300</td>
<td>144</td>
<td>1800c</td>
</tr>
<tr>
<td>Area (mm²)</td>
<td>1.807d</td>
<td>165.3</td>
<td>7.25</td>
<td>37.7c</td>
</tr>
<tr>
<td>Technology</td>
<td>45 nm CMOS</td>
<td>SiGe</td>
<td>130 nm SiGe</td>
<td>BiCMOS</td>
</tr>
<tr>
<td></td>
<td>SOI</td>
<td>BiCMOS</td>
<td>CMOS</td>
<td>BiCMOS</td>
</tr>
</tbody>
</table>

3 Layout techniques for active and passive devices for mmWave circuits

3.1 Overview

The biggest limiting factor in mmWave circuits are the parasitic components associated with the device layout. Therefore, in addition to the device models, optimum layout design towards reduced parasitic components is pivotal for high performance circuits. In this chapter, the main focus is on the layout optimization techniques of key circuit blocks in mmWave systems. Two layout techniques for active cascode devices are discussed by the author. The design of a transmission line based differential quadrature hybrid is presented in this chapter. As signal combiners are a fundamental building block of a phased array system, a comparison of different mmWave signal combiners is also discussed in this chapter. The presented building blocks in this chapter are primarily based on the research outcome from articles IV, V, VI and VII.

3.2 Silicon-based technologies

With the exploding number of wireless applications and connected devices (Internet of Things), silicon-based technologies are trying to offer integrated solutions (application specific integrated circuits [ASIC]) where faster digital circuitry and high performance radio frequency front-end circuits are fabricated in a single chip. Compound semiconductor technologies have been well known for outperforming Si at RF and mmWave frequency applications [52]. However, due to the lack of high-speed digital devices, they are only suitable for RF/mmWave front-end designs. CMOS transistors have been traditionally used for digital circuits and their research for RF applications started when the first AC models of the MOSFETs were introduced in 1987 [53]. Advancements in circuit design techniques [54, 55] and the availability of better RF/mmWave circuit models [56] allow the circuit designers to meet the specified RF performance using CMOS based technologies with almost similar performance compared to SiGe BiCMOS technology. For example, a 28 nm CMOS power amplifier (PA) presented in [54] achieves a peak power added efficiency (PAE) of 26.6% and an output power of 26 dBm at 39 GHz. Similarly, in [56] a 24 to 28 GHz low noise amplifier (LNA) implemented in a 45 nm CMOS SOI process has shown a noise figure (NF) of 1.4 dB with a performance very close to the performance of the latest GaAs LNAs. Due to the low cost, better
digital and RF/mmWave performances and high-level of integration, CMOS based technologies are considered the biggest players to drive fifth generation communication technology.

Initially, the MOSFET models provided by the foundries were based on AC models with only parasitic components associated from the structures in the silicon substrate [57], the layout dependent metal wiring parasitics, especially gate resistance were missing. However, it was investigated in [58] that the RF performance of a transistor predominantly relies on the parasitic resistance associated with the gate wiring of the device, i.e, \( r_g \). Similarly, the gate parasitic capacitance of silicon devices establish a limit to the cut-off frequencies of the device (\( f_t \) and \( f_{\text{max}} \)). The cutoff frequency \( f_t \) is defined as the maximum frequency of the device at which the short-circuit current gain becomes unity and is given as

\[
f_t = \frac{g_m}{2\pi(C_{gs} + C_{gd})},
\]

where \( g_m \) is the transconductance of the device and depends on the current density of the technology, and \( C_{gs} \) and \( C_{gd} \) are the intrinsic parasitic capacitances at the gate terminal of the device. Similarly, the frequency of a device at which the maximum achievable power gain of the device becomes unity is called the maximum oscillation frequency (\( f_{\text{max}} \)) and is given as,

\[
f_{\text{max}} = \frac{1}{4\pi \sqrt{r_g C_{gd}(C_{gs} + C_{gd})}},
\]

where \( r_g \) is the overall gate resistance of the device and mainly contributed by the resistance of the poly silicon area and its associated metal contacts. The key parameters \( f_t \) and \( f_{\text{max}} \), characterize RF performances of an active device for a given technology. The presence of both \( r_g \) and gate capacitance (\( C_{gg} \) or \( C_g \)) in the equation 18 makes \( f_{\text{max}} \) a more layout dependent parameter which can be optimized for a high performing device. Another key performance parameter for RF/mmWave frequencies is the quality factor (Q) of on-chip passive devices which is mainly dependent on the resistivity of the substrate and the properties of available metal layers in the technology metal stack.

### 3.3 Low-loss substrate processing techniques

The silicon substrate is the biggest constraint for the performance of an integrated passive device at mmWave frequencies. The behaviour of the substrate depends on its material properties, doping profile and eventually the material properties used for the back plate of the substrate. Using the substrate as an RF ground is typically avoided due to multiple reasons, i.e, 1) the conductive nature of substrate adds loss to the return
currents from the circuit blocks, 2) expensive and/or the unavailability of through silicon vias to connect the RF ground with backplate metal sheet, 3) cross talk between different sub-blocks through the substrate. A silicon based substrate can be optimized for the best performance at RF/mmWave frequency operations by using various processing techniques, e.g, silicon on very low loss materials [59, 60, 61], selective etching [62], deep trench isolation (DTI) [63], and silicon on insulator (SOI) [64, 65].

![Silicon on Insulator wafers cross section](image)

**Fig. 18.** Silicon on Insulator wafers cross section, (a) high resistivity silicon SOI, (b) HR silicon SOI with trap rich layer (Modified from Esfeh 2016 [64]).

### 3.3.1 RF CMOS SOI

Weak doping profile silicon substrates possess low conductivity or high resistivity (HR) properties. Due to latch up problems between MOSFET devices, low doping HR substrates cannot be implemented in bulk silicon technology [66]. However, an insulator layer, for example, buried oxide (BOX), can be introduced to isolate the active devices from the base HR silicon, as shown in Figure 18a. Employing such an HR substrate in the CMOS SOI process helps to achieve an improved RF performance for active and passive devices and leads to RF CMOS SOI technology. One limitation of SOI technology is the poor thermal conductivity of the BOX material (SiO$_2$). Thermal conductivity is inversely proportional to the thickness of the BOX layer that implies that a thin layer of the BOX layer degrades the thermal properties of the SOI device less compared to a thick one. The use of an HR-SOI substrate and a thin layer of BOX
facilitates high performance passive devices, increased isolation MOS switches, reduced substrate coupling, etc.

Unfortunately, a low ohmic conduction channel is induced at the junction of the BOX and silicon substrate due to the fixed charges of oxide that attract free carriers at the interface of Si and $SiO_2$. The emergence of this channel degrades the resistivity of the substrate significantly. This effect is known as parasitic surface conduction (PCS). In 2005, researchers investigated the PCS effect in SOI devices and proposed a modification in the silicon substrate surface under the BOX [66]. This modification consists of introducing the trap layer at the Si/BOX junction [64][66][67], as shown in Figure 18b.

The evolution of SOI technology started with the introduction of stacked field effect transistors (FET) in 1989 [68]. J. P Raskin et al. demonstrated in [69] a significant reduction in crosstalk using HR-SOI technology. RF CMOS SOI technology has become a very attractive choice for low loss front-end designs, for example, transceiver switches [70], stacked-FET PAs [71] and LNAs [72]. Moreover, RF CMOS SOI technologies have also shown exceptional performances at mmWave frequencies, offering measured $f_{\text{max}}$ of 200 GHz and 283 GHz in [73] and [74], respectively.

3.3.2 Technology aspects adapted in this thesis

RF CMOS SOI technology is a good candidate for the design of complex 5G mmWave systems mainly because of its ability to offer a high-resistive substrate for low-loss passive devices and high isolation and low-loss switch design [75]. As SOI technology offers significant advantages to achieving a good receiver Noise Figure in mmWave designs, this technology was chosen for the design of the circuits presented in publications I to VIII.

3.4 Active devices

Active devices determine the high frequency gain and noise performance of an amplifier. Process design kits (PDKs) of RF CMOS technologies include appropriate models of active devices that contain sufficient information about the behaviour of devices for high frequency up to some predefined frequency. These models are based on predefined equivalent and scalable models of a transistor that is typically based on processed and measured devices.
3.4.1 Intrinsic and extrinsic models

The model of a Si-based active device is divided into two parts, i.e., the intrinsic model and the extrinsic model. The intrinsic model of a device consists of the key parasitic components of the device provided by the foundry. This device model includes the silicon area, its surrounding sub-circuit and wiring, and possible floating-body effects. For example, the parasitic resistance and capacitance due to the metal connections of the poly-silicon gate fingers to a certain metal layer is included in this model. In addition to the intrinsic parasitic components, the performance of a CMOS transistor at mmWave frequencies is also limited significantly by extrinsic parasitic components, primarily from transitions of the device terminals to or from other devices, i.e., interconnects. These interconnects are composed of metal tracks and vias from lower metal layers up to higher metal layers. This wiring adds additional resistance and capacitance at the transistor terminals. The effects of both intrinsic and extrinsic parasitics can be reduced by using good layout practices to maximize the device performances. As an example, Figure 20 shows different optimization techniques available in the intrinsic model of the device. A 5 μm wide NMOS device can be laid out in different forms, e.g., the layout in Figure 20a has only one finger with 5 μm having \( r_g \) contributed by the poly-si area which is 5 μm × L. However, the same device is split into five fingers and the total \( r_g \) is also decreased by five times, as shown in Figure 20b. The multiple finger device also increases the gate capacitance \( (C_g) \) and therefore, a trade off exists between \( r_g \) and \( c_g \). In addition to the gate split into multiple fingers, \( r_g \) is further reduced by increasing the number of contacts from poly to metal1 (doubly contacted [73]), as shown in Figure 20c.

\( f_{max} \) is used as the fundamental parameter to optimize the layout related device performance at mmWave frequencies. Higher values of \( f_{max} \) provide optimum device performance in terms of transconductance \( (g_m) \) and noise Figure \( (r_g) \) with an optimum current density. The \( g_m \) of the device increases with the current density (mA/μm)

Fig. 19. An NMOS transistor model and layout with its parasitic components (Modified from Harish K. 2016 [76]).
and saturates at a peak current density due to the velocity saturation of the charge carriers. Therefore, $f_{\text{max}}$ also reaches saturation at the maximum current density of the technology. It has been noticed in [78] that current densities for peak $f_t$ and $f_{\text{max}}$ are constant across multiple CMOS technologies, i.e., ~200 $\mu$A/$\mu$m. This feature helps to conveniently migrate the designs from one technology to others. Modern process technologies offer dedicated PDKs for mmWave circuit designs, in which optimized layout effects are included in the intrinsic model of the active devices. This feature helps designers to estimate key specification parameters already at the concept phase of the design. An additional reduction of $r_g$ is possible by using custom-made layouts at the unit cell level, e.g., the multi-gate style layout. A comparison study between two different styles of the layout for optimum $f_{\text{max}}$ was performed in publication VI and VII.

Various layout techniques have been proposed in literature to minimize extrinsic parasitic components in terms of $r_g$ and $c_g$ for mmWave frequency applications, for example, a gate above device [73] (Figure 20d), a multi-gate layout design for stacked transistors [79, 80], a round-table device [81], a zipper style device [82], a drain-above device [83], etc. These layout techniques are designed as unit cells and are mostly used for the PA design to realize the bigger device size. A multi-gate design is useful for a better substrate connection for thermal flow in large devices, as the round-table layout design benefits from the reduced resistance at its terminals but at the cost of a higher
parasitic capacitance at the drain connection. On the other hand a zipper-style layout has a good symmetric design for drain and source connections, while a drain-above layout technique helps in optimizing the resistance of source and drain terminals. In addition, the modern state of the art RF CMOS PDKs offers layout optimization techniques already included in the intrinsic model of the device, for example, a number of gate connections, an extension of gate, and source and/or drain terminals at metal3 or metal4. This type of modelling feature will help in evaluating layout trade offs already at an early phase of the designs.

3.5 Design and characterization of cascode devices

The presence of Miller’s capacitance in a common-source \((c_{gd})\) transistor reduces isolation between input and output and can cause instability at the frequency of operation and this can become a serious concern at high frequencies, i.e, especially the mmWave region. This capacitance can be optimized to a certain value by using good layout practices however, it cannot be fully eliminated. Common-source cascode design topology is the leading option for mmWave low noise amplifiers (LNAs) among other topologies due to its inherent properties of optimum noise match and better stability [84, 85, 86]. Publication VI and VII present two layout design techniques for cascode connected devices, and their key performance parameters are compared, for example, \(r_g\), \(c_{gg}\), \(g_m\), \(f_{max}\). The presented layouts are designed and fabricated using 45 nm CMOS SOI technology.

![Schematic of cascode devices](image)

Fig. 21. Conventional-style layout (Reprinted, with permission from Publication VII ©2020 Springer).
In this stage of the research, two layout structures of cascode-connected NMOS devices each having a 20 $\mu$m device width were designed and characterized. The first layout (Figure 21) was designed using the gate-over-device technique [73]. The second structure (Figure 22 and Figure 23) was based on the multi-gate style layout technique also presented in [79] and [80]. The schematics and the micrographs of the test structures are shown in Figure 24. Figure 25 shows the comparison of different extracted parameters between the two layout styles of similar device width (i.e., w=20$\mu$m). A brief comparison of the extracted parameters at 40 GHz frequency is also summarized in Table 4.
From the results in Table 4, it can be seen that $r_g$ of the multi-gate style layout is lower than the conventional style layout ($\sim$10%). This is noticeable from the layout as the gate wiring is implemented in a tree configuration, thus reducing the overall wiring resistance. However, due to the complex wiring for the source and drain terminals, the capacitance is also increased significantly in the multi-gate layout compared to the conventional style layout. Transconductance of the devices shows a similar performance across the gate bias voltage. $F_{\text{max}}$ is calculated by using the extracted values of the gate resistance and capacitance using equation 18. Furthermore, $F_{\text{max}}$ curves are compared in Figure 25 (d) and shows that multi-gate style layout has a lower $F_{\text{max}}$ due to its increased gate capacitance.

From this work, it was noticed that the multi-gate layout style benefits from a reduced gate resistance which can be useful to relieve the inter-stage matching conditions as required in [87, 88, 89]. The increased number of vias in this layout technique at different terminal connections offer good reliability aspects, especially in the power amplifier designs. However, these benefits come at the cost of increased complexity and wiring capacitance which results in a reduced device performance ($F_{\text{max}}$). The mmWave LNA circuits are designed based on both of the above mentioned layout techniques and are further discussed in chapter 4.
Fig. 25. Extracted data from measured s-parameters (a) $r_g$, (b) $c_{gs}$ and $c_{gd}$, (c) $g_m$, (d) $f_{\text{max}}$ (Reprinted, with permission from Publication VII © 2020 Springer).

Table 4. Comparison table of extracted parameter values at 40 GHz (VDD=1 V, $V_{\text{bias}}$=580 mV) (Reprinted, with permission from Publication VII © 2020 Springer).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Conventional</th>
<th>Multigate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_g$ (Ohms)</td>
<td>14.3</td>
<td>12.9</td>
</tr>
<tr>
<td>$C_{gs}$ (fF)</td>
<td>60</td>
<td>62</td>
</tr>
<tr>
<td>$C_{gd}$ (fF)</td>
<td>2.77</td>
<td>4.33</td>
</tr>
<tr>
<td>$g_m$ (mS)</td>
<td>22.5</td>
<td>22.5</td>
</tr>
<tr>
<td>$f_{\text{max}}$ (GHz)</td>
<td>243</td>
<td>215</td>
</tr>
</tbody>
</table>

### 3.6 Passive devices

Passive devices are the fundamental building block in mmWave frequency systems. They are mainly used as matching networks, resonator loads, baluns, etc. The performance of Si-based passives at mmWave frequencies is characterized by the component’s maximum frequency of operation, i.e., self-resonant frequency (SFR) and the quality
factor (Q). A loss of energy due to unwanted electromagnetic coupling occurs between any conductor and its surrounding structures, as shown in Figure 26. Furthermore, all the interconnects (for example, vias and transitions) exhibit finite inductance, capacitance and resistance, and at the mmWave frequency these parasitic effects can result in a significant shift in the performance of the devices from their estimated values. Therefore, careful modelling of the passive elements, their interconnects and their surrounding is essential in order to precisely model all the associated parasitic components. In publication VII, some of the key characteristics of passive devices and their applications in mmWave IC designs are discussed.

![Electromagnetic coupling phenomenon between two conductors on a silicon substrate (Modified from F. Gacim, 2017 [63]).](image)

### 3.6.1 Magnetic field losses

When a high frequency alternating current (AC) is flowing through a metal conductor, a magnetic field is induced around it. This induced magnetic field generates eddy currents on the surface of the surrounding structures, including the substrate, as shown in Figure 26. These eddy currents produce a magnetic field which tries to oppose the incident magnetic field (Lenz’s law). Typical CMOS silicon substrates (highly doped) are inherently conductive (low resistive) in nature, thus eddy currents find a conductive paths to flow in the low ohmic substrates and this results in a loss of energy and appears as the loss in quality factor of the device. These magnetic field losses cannot be avoided.
unless the magnitude of the induced parasitic currents can be decreased by increasing the resistance of the substrate material using different processing techniques, e.g. silicon on insulator (SOI), micromachining or selective etching, as discussed in section 3.3, etc.

### 3.6.2 Electric field losses

An electric field is also generated when a high frequency AC current flows through a conductor, which results in capacitive coupling with the interaction of surrounding materials, for example, metal structures and silicon substrate. This capacitive coupling generates displacement currents in the substrate which appear as a loss in inductance and the quality factor of the device. A metal plate stops or shields a significant portion of the electric field lines from flowing into the substrate and hence, reduces the associated losses in the substrate, also shown in Figure 27. Due to the capacitive coupling, eddy currents are also generated in this metal plate and results in additional losses. Various types of metal shield patterns made of lower metal layers are used [90, 91, 92]. However, the use of a metal shield also increases the capacitance between the main conductor and the shield, which results in the reduction of the SRF of the device.

![Electric field lines](image)

**Fig. 27. An inductor drawn on a silicon substrate (Modified from F. Gacim, 2017 [63]).**

### 3.6.3 Skin effect

At high frequency operations, in addition to the DC resistance of a conductor, a frequency dependent resistance also adds up in the total resistance of the conductor. Due to RF current flowing inside the conductor, electric fields generate eddy currents on the surface of the conductor. The development of eddy currents tends to cancel the flow of the current in the middle of the conductor. As a result, the maximum current density occurs on the surface of the the conductor, reducing the effective area for the current to flow. In other words, this increases the resistance of the conductor which is a strong function of the operating frequency. This phenomenon is known as skin effect ($\delta$) and it is expressed
as
\[
\delta = \frac{1}{\sqrt{\pi \mu \sigma_{\text{cond}}}},
\]
(19)

where \( \delta \) is the skin depth, \( \mu \) is permeability and \( \sigma_{\text{cond}} \) is the conductivity of the material. Equation 19 shows that with an increase in frequency, the skin depth of a material is decreased and the effective resistance is higher.

Various techniques have been proposed in the literature to mitigate the aforementioned non-idealities for passive devices at high frequency operations. For example, a magnetic substrate coupling can be avoided by changing the substrate properties used under passive devices, such as high resistive substrate [64, 65], micromachining techniques [93, 62], formation of trenches in the substrate [63], etc. To reduce the capacitive coupling, ground shielding techniques [90, 91, 92] are utilized under and around the passive devices.

3.7 Electromagnetic modelling

Close correlation between simulated and measured results of an integrated circuit depends on:

1. the availability of accurate models of the active devices and their RC extraction methodology,
2. precise electromagnetic (EM) modelling of passive devices,
3. accurate realization of interconnects between active and passive stages, for example, metal vias, transmission lines, etc.

Numerous EM simulation software such as Momentum (Keysight), Sonnet, HFSS and EMX (Cadence) are utilized to extract the EM model of a metal structure (for example, inductors, transmission lines, transformers, etc). There are different methods available for the model extraction of the structures, for example, moment of methods (MoM), finite element method (FEM) and 3-dimensional extraction. The right choice to use these methods depends on the structure, layout and its application. For example, the MoM method also called a two and a half dimensional (2.5D) extraction used for planar structures (inductors, transformers and transmission lines) inside the IC, is based on the surface meshing method. However, 3D extraction methods, e.g. FEM, use the volume meshing approach and are used for extracting 3D structures, for example, the metal connections of a transistor or interconnects for IC IOs, e.g. bond wires or solder bumps, etc. MoM is preferably used because it is faster than 3D methods with sufficient accuracy. The outcome of these extractions is a set of scattering parameters [94]. These s-parameter models are then used in a simulation test bench together with
other circuit models, i.e, RC-extracted models of active devices. EM software calculates electromagnetic fields in and around the metal structures based on the port setup, the surrounding environment, ground reference plane for ports, substrate properties, material of the structure, etc. It is essential to include all the necessary information of the surrounding environment in the simulation setup in order to compute the electric and magnetic coupling and their associated losses.

### 3.7.1 Definition of RF ground

A low ohmic RF ground is required to be used as a reference for the port definition for its return currents. Based on this ground reference, electric and magnetic fields and the associated parasitic currents are calculated in the conductor which define the inductive and capacitive characteristics of the structure. This RF ground should be very close to the actual implementation. A layer definition file in an EM tool includes the necessary information about material properties (i.e, permeability, permittivity, resistivity, thickness) of the substrate, dielectric material, available metal layers, vias and global ground plane reference. In these setup files, the bottom of the substrate is defined as the global ground plane as a port reference for EM simulations. One way of realising this RF ground is to use the metal plate at the bottom of the substrate. However, in reality its connection to the top metal layers of the IC requires through-hole silicon vias. These through-hole vias require additional processing steps that are unavailable in most of the standard IC manufacturing processes. Moreover, eddy currents can be generated on this solid back metal plate that results in losses to the passive structures.

Another way of defining the port, is to use silicon substrate as the RF ground reference. When a port is defined with the reference at the bottom of the substrate as the RF ground reference (Figure 28a), the height of the port becomes equal to the total height of the substrate. This implies in [94] that the height of the port should be electrically small, i.e, $\lambda/10$ at most, where $\lambda$ is the wavelength of maximum frequency of simulation. A typical substrate height is a few hundreds of microns which translates into port dimensions equal to the substrate height and eventually limits the maximum frequency of simulation for mmWave frequency operations.

### 3.7.2 Local metal layer as RF ground

The aforementioned limitation of port dimensions can be relaxed by defining a local ground reference for the port excitation, as shown in Figure 28b. In this type of port setup, a robust metal ground plane is drawn using metal layers. The local ground
reference of the top most metal layer is also more realistic in implementation, since it is composed of a metal layer which has the smallest distance between the chip and the solid ground on PCB. Additionally, the local ground on the top metal layers benefits from a low metal resistivity as well providing electrical isolation between the neighbouring passive structures. Impedance of the metal ground [95] around a device-under-test (DUT) affects the characteristics of the DUT. Therefore, this ground plane should have a very low impedance in order to minimize its impact on the quality factor and inductance of the device’s performance.

3.7.3  PCB as RF ground

In order to obtain an accurate and realistic model, an RF ground can be customized to other layers, e.g. metal1, or even outside the IC domain, i.e, PCB layer, as shown in Figure 29. This can be done by defining two additional layers (one for the solder bump or copper pillar and a second for the PCB layer) inside the substrate definition and
define the PCB layer as the RF ground reference for the EM setup. A port definition using the aforementioned setup is shown in Figure 29b.

Fig. 29. Simulation layer stack for EM simulations using a PCB metal layer as an RF ground plane for port definitions (a) silicon die connections with PCB using solder bumps in flip-chip technology, (b) defining EM ports using RF ground of a PCB layer (Reprinted, with permission from Publication VII ©2020 Springer).

3.7.4 Effect of the ground ring on inductor performance

The magnitude of the inductance of an inductor depends on the magnetic field strength induced around its conductor and the shape of the conductor [96]. For example, a piece of metal wire of length \( l \), possesses a smaller inductance when it is implemented as a straight line compared to a loop shaped, as shown in Figure 30. A parasitic electric field is also present which limits the frequency of operation of the inductor.

The performance of silicon based integrated inductors is mainly characterised by the properties of the silicon substrate and the availability of low loss thick copper metal layers. Generally, the metal resistivity can be further reduced by stacking multiple metal layers [97]. Inductor parameters (\( L \) and \( Q \)) can be calculated from an EM extracted s-parameter model as given in [98] (single-port excitation),

\[
L = \frac{\text{Imag}(y_{11})}{2\pi f},
\]

\[
Q = -\frac{\text{Imag}(y_{11})}{\text{Re}(y_{11})},
\]

where \( y_{11} \) is the input admittance of the inductor. Generally, models of passive structures are extracted alone with a metal ground plane (MGP) around them. This MGP is connected with the global ground plane on the chip to provide electrical isolation of the structures from neighbouring structures [99]. Return currents induced from electric and magnetic fields flow back in the ground plane and the inductance and resistance of the
MGP affects the performance of the inductor. Therefore, a low ohmic and low inductive ground plane is essential for high performing mmWave circuit designs. The effect of a finite impedance ground plane on the total $L$ and $Q$ of a single turn coil is shown in Figure 31. It can be seen that this effect depends on the ratio between the distance ($d$) of the coil from the ground plane and the inner radius ($r$) of the coil. The effect of the ground plane is minimized when $d/r$ is equal or greater than 1, which means that when the $d$ becomes greater than $r$, the effect of parasitic currents due to the ground plane becomes negligible.

### 3.7.5 Design methodology for mmWave circuits

A brief flow chart of a design methodology for mmWave circuit designs from specifications to the final result is presented in Figure 32. Starting from the specifications, the active and passive components are simulated using their simplest schematic models in order to do faster simulation iterations. Once the component values are known, more detailed modelling is added, e.g., re-extraction for active devices and losses in passive components. This modelling would require another tuning step in passive component values. EM modeling of passive components is first performed using cascaded views of individual passive components together with big metal plates of MIM capacitors (mostly DC-caps), power supply nets, and the local ground of each passive component. The cascaded approach is helpful in its faster iterations with sufficient accuracy. After the design is
completed with the individual modelling of each passive component in the circuit, a global view is generated with all the circuit elements, except the active devices, i.e., transistors. This includes the inductors, transmission lines, global ground plane, metal plates of MiM capacitors and the power supply network (VDD) of the whole design. This single layout is extracted using an EM software to generate an EM model. This EM model is simulated together with rc-extracted models of the active devices to obtain the most accurate and realistic results of the design which includes all the parasitic effects, i.e., wanted or unwanted coupling between different parts of the design. There are many iterations required to optimize the design flow. Amongst these, three iterations are marked in the flow diagram. Iteration1 comes first in the flow as the optimization of the layout of active devices, which helps in improving the rc-extracted model of the devices. Iteration2 appears when the first layout of the passive structures is drawn using the local ground planes. This iteration is used for the optimization of the passive devices, which is rather fast to perform as modelling of the individual structure is always

Fig. 31. (a) Inductance and (b) quality factor of an inductor with a ground plane for different \( d/r \) ratios (Reprinted, with permission from Publication VII ©2020 Springer).
faster. The final iteration (Iteration3) is the global EM optimization of the circuit which includes all the passive structures and their respective ground planes in a single design. This iteration includes the complete circuit and therefore it consumes more time and processing resources for detailed modelling.
Fig. 32. EM methodology design steps for mmWave circuit designs.
3.8 Transformers as resonators

Transformers are used to transform impedance, for example, the low output impedance of the power amplifier (i.e., 5 Ohms) to the 50 Ohm impedance of an antenna. Moreover, transformers are also used as frequency resonators in RF/mmWave systems due to their characteristics of generating multiple resonances [42, 100]. A circuit symbol and lumped-model of a transformer is shown in Figure 33. The performance of a transformer is characterized by the self inductance of the primary winding ($L_p$), the self inductance of the secondary winding ($L_s$), and the mutual coupling ($k$) between the primary and secondary inductors.

Fig. 33. Transformer schematic (a) and model with parasitics (b) (Reprinted, with permission from Publication VII ©2020 Springer).

Fig. 34. Example of a transformer layout (Reprinted, with permission from Publication VII ©2020 Springer).

The multiple-resonance feature of transformer-based resonators has enabled narrowband and wideband circuit designs for RF/mmWave applications in a compact area. A simple transformer-based resonator in Figure 33b resonates at two frequencies, i.e., $f_L$ and $f_H$. 
which can be computed by using following equations as,

\[ f_L = \frac{1}{2\pi \sqrt{LC(1+k)}} \],
\[ f_H = \frac{1}{2\pi \sqrt{LC(1-k)}} \]. 

(21)

It can be observed from equation 21 that lower values of \( k \) move the two resonances towards each other, while a higher value of \( k \) moves \( f_L \) and \( f_H \) apart from each other, as shown in Figure 35. This feature of a transformer provides a flexible design approach in designing narrowband [83] or wideband circuits [100, 42]. Various transformer layout techniques have been utilized in the literature to optimize different \( k \) values with quality factors of \( L_P \) and \( L_S \) [101, 102, 103, 104]. A 3D layout diagram of a 1:1 transformer with lateral coupling is shown in Figure 34. Extraction of the primary and secondary coil parameters is done by using the equation 20, while coupling between the coils is calculated as (two-ports excitation)

\[ k = \frac{\text{Imag}(Z_{12})}{\sqrt{\text{Imag}(Z_{11}) \times \text{Imag}(Z_{22})}} \]. 

(22)

3.9 Design and characterization of a differential reflection-type phase shifter

A quadrature coupler is a four-port device, as shown in Figure 36. When an input signal is fed at input (IN) port 1, this signal is split into through (TH) port 2 and coupled (CPL)
port 3 with equal magnitude but with a 90° phase difference between each other. The isolated port 4 (ISO), in ideal conditions has zero power transferred from the input port. These couplers are commonly used in many applications for mmWave systems, for example, the quadrature generation, signal combiners, reflection type phase shifters (RTPS), etc. An RTPS, shown in Figure 36b, is composed of a quadrature coupler in which ports 2 and 3 are terminated with resonant load circuits. An input signal incident at port 1 is reflected from port 2 and port 3 with an additional phase depending on the value of reflection due to the resonant load at the port 2 and port 3. This reflected signal is taken out from the ISO port, which is the output port. Different values of phase settings can be achieved with the help of tunability of the resonant load. An RTPS is a bi-directional passive device and can be shared in the transmitter and receiver in a transceiver system. The lengths of the silicon-based quarter wave transmission lines at 30 GHz are in the order of ~1mm and, therefore, the area footprint of a TL-based quadrature coupler is a limiting factor to be extensively used in the phased array systems. The lumped-element based quadrature hybrid coupler has shown promising performance features and is being adopted in many applications [105, 106].

Publication IV proposed a differential RTPS based on co-planar waveguide (CPW) based coupled lines using vertical coupling between the differential lines. Block diagrams of coupler and RTPS are shown in Figure 36. The variable $LC$-based reflection load of the phase shifter is composed of a digitally tunable capacitor bank (Figure 37 a) and a fixed value inductor (Figure 37 b). The fabricated chip micrograph of the test structure is shown in Figure 38. The on-wafer measurement was performed using ground-signal-ground-signal-ground (GSGSG) differential probes. The measured insertion loss and insertion phase of the RTPS for different settings of the switching capacitor bank are shown in Figure 39. The achievable insertion phase is 115° with a 4 dBs variation in insertion loss. The performance table and the comparison with the state-of-art designs are presented in Table 5. The insertion loss variation of the RTPS is higher than the compared designs, which is mainly because of the parasitics of the switched capacitor. This is a symmetrical and bi-directional design and therefore can be used in a transceiver as a shared block between the receiver and the transmitter.

71
Fig. 36. Block diagrams of a 3 dB coupler and an RTPS (Reprinted, with permission from Publication IV ©2019 IEEE).

Fig. 37. Reflection load elements (a) switchable capacitor bank (b) inductor (Reprinted, with permission from Publication IV ©2019 IEEE).
Fig. 38. Micrograph of the RTPS (Reprinted, with permission from Publication IV ©2019 IEEE).

Fig. 39. Measured insertion loss from input to coupled port (Reprinted, with permission from Publication IV ©2019 IEEE).
Table 5. Comparison of RTPS with state-of-art designs (Reprinted, with permission from Publication IV ©2019 IEEE).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>This work</th>
<th>[107]</th>
<th>[108]</th>
<th>[109]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>RTPS</td>
<td>RTPS</td>
<td>RTPS</td>
<td>RTPS</td>
</tr>
<tr>
<td>Process</td>
<td>40 nm CMOS SOI</td>
<td>SiGe 0.13 µm</td>
<td>SiGe 0.13 µm</td>
<td>BiCMOS 55 nm</td>
</tr>
<tr>
<td>Topology</td>
<td>Differential</td>
<td>Differential</td>
<td>Differential</td>
<td>Single-ended</td>
</tr>
<tr>
<td>Freq (GHZ)</td>
<td>39-45</td>
<td>26.5-32.8</td>
<td>60</td>
<td>30-50</td>
</tr>
<tr>
<td>Phase (°)</td>
<td>115</td>
<td>180</td>
<td>156</td>
<td>60</td>
</tr>
<tr>
<td>IL (dB)</td>
<td>9</td>
<td>6</td>
<td>6.2</td>
<td>5</td>
</tr>
<tr>
<td>Variation (dB)</td>
<td>4</td>
<td>2</td>
<td>2.2</td>
<td>2</td>
</tr>
<tr>
<td>Area (mm²)</td>
<td>0.214</td>
<td>0.64</td>
<td>0.33</td>
<td>0.18</td>
</tr>
</tbody>
</table>

3.10 Active and passive signal combiners for mmWave phased array systems

Publication V briefly discussed different types of combiners, i.e, active combiner [42], Wilkinson combiner (distributed [37] and lumped-element based [110, 111, 112]) and a transformer-based combiner [113, 114]. Two differential 4-channel passive combiners, i.e, a transmission line (TL) based Wilkinson combiner and a transformer-based combiner, are designed at 28 GHz using 45 nm CMOS SOI technology and their simulated performances are compared.

Active combiners are consist of a transconductance ($g_m$) stage in each channel that converts the voltage signal into a current. These currents are then combined at one common point (symmetrical voltage supply) using a common load for all $g_m$ stages, as shown in Figure 40. MmWave active combiners are compact compared to the passive combiners; however, they are based on non-linear devices and hence burn extra power to linearize. Moreover, they are uni-directional devices and cannot be shared between the Tx and Rx in a transceiver system.

Based on $\lambda/4$ impedance transformers [37], a Wilkinson combiner outperforms all existing structures in terms of matching conditions at all ports and isolation between the combining or splitting ports. However, the bandwidth of this type of combiner is a limiting factor. A schematic diagram and its method of operation is shown in Figure 41, where one can note how an input signal (Vin1) at IN1 is combined with a signal at IN2 at the output terminal (OUT). However, its leakage signal (180° phase shifted) is travelled back to input IN2 and cancelled with the original signal passed through the
isolation resistor $R_{iso}$. A lumped-element based Wilkinson combiner (Figure 42) can also be realized such that the structure can be made area efficient.

A four-channel differential Wilkinson combiner based on $\lambda/4$ transmission lines is designed, as shown in Figure 43. The layout symmetry of the Wilkinson combiner is essential to make sure the equal phase and loss is present between each input and combined output. The complete structure is simulated in Momentum 2.5D EM software from Keysight. The simulation results of the insertion losses of channels and isolation between two channels are shown in Figure 44. An insertion loss of 3.7 dB and a good isolation down to -36 dB at the operating frequency between two channels is observed at the center frequency.

Similarly, a 4-way differential combiner connected in a parallel configuration is designed to support 5G mmWave frequency bands. A single turn center-tapped primary coil is used as the input for each channel. The input signal from each channel is magnetically coupled with the secondary coils which are connected together electrically to combine all the channels, as shown in Figure 45. The frequency response of these transformers is exploited from their maximum magnetic coupling which generate dual resonances and only one resonance is selected around the operating frequency of 28 GHz, as described in section 3.8. The transformers are designed using lateral coupling between primary and secondary conductors to achieve maximum magnetic coupling [96]. The primary and secondary coils are drawn using thick copper layers from the technology metal stack, i.e., M7 and M8, respectively. Connecting two adjacent secondary coils together, implementing a cross-connection, as shown in Figure 45 b, significantly improves the isolation between the two inputs. This is due to the cancellation of any magnetic coupling between two coils by changing the flow of current in the opposite direction. The simulated insertion loss of the individual channels is 1.5 to 3 dB across
the wideband frequency range (23 GHz to 43 GHz), covering most of the 5G NR bands. However, isolation between the two adjacent channels is only 14 dB.

Two 4-way passive combiners, i.e., Wilkinson based and transformer based combiners, were simulated and presented in this section. The comparison shows that there is a trade-off between two combiners in terms of IL, isolation and area. Although the Wilkinson combiner provides the best isolation between two channels at the cost of larger chip areas, transformer-based combiners are compact and have very low IL. However, in terms of isolation between two adjacent channels they are not the best choice.
Fig. 41. A 2-way Wilkinson combiner (Modified from Wilkinson 1960 [37]).

Fig. 42. A 2-way Wilkinson combiner based on lumped-element $\lambda/4$ transformers (Modified from Park 2016 [110]).
Fig. 43. A 2-way Wilkinson combiner based on lumped-element $\lambda/4$ transformers (Reprinted, with permission from Publication V ©2019 IEEE).
Fig. 44. Simulated results of a 4-way Wilkinson combiner (Reprinted, with permission from Publication V ©2019 IEEE).

Fig. 45. A 4-way transformer based passive combiner using parallel configuration (Reprinted, with permission from Publication V ©2019 IEEE).
Fig. 46. Simulated results for a transformer-based combiner, (a) IL and isolation between two channels, (b) amplitude difference between channels, (c) phase difference between channels (Reprinted, with permission from Publication V ©2019 IEEE).
3.11 Conclusion

This chapter mainly discussed the characteristics and layout design techniques for the silicon-based active and passive devices in order to achieve their optimum performances in mmWave phased array systems. For mmWave frequency systems, it is very important to understand the major contributors present in the intrinsic and extrinsic models of an active device which limits the maximum frequency operation of the device. These layout dependent contributors were briefly investigated and a comparison study of two different layout techniques of a cascode-connected device was also presented in this chapter. Similarly, layout dependent effects for passive devices, e.g., inductor, transformer, were also discussed, including the ground plane and port definitions. Moreover, a design methodology for the complete design flow of silicon-based mmWave circuit design was presented. This way of working would be helpful in designing mmWave circuits with a good co-relation between simulation and measurement results. In addition to aforementioned, a differential reflection-type phase shifter based on a vertically-coupled coupler and a switched-capacitor bank was presented. The layout of this structure was made compact and symmetrical which can be integrated for a bidirectional operation in a phased array channel. Finally, a brief comparison was presented between different signal combiners in a phased array system including their pros and cons in terms of drawn area, insertion loss, isolation between two combining channels, linearity and power consumption.
4 Low noise amplifiers for mmWave applications

Low noise amplifier is a key building blocks of the receiver. A typical phased array system (RF beamforming) consists of various active blocks in the signal chain, such as splitters, vector modulators, variable gain amplifiers, combiners, switches, etc. [38, 115, 48]. An LNA is the first building block and determines the minimum possible noise figure of the whole receiver system [19]. With sufficient gain and a low enough noise figure, it also limits the gain requirements of subsequent components in the signal chain and allows the use of passive structures without impacting the overall noise figure of the system. Although the phased array receiver systems benefit from array gain [48], still the noise figure of the LNA is typically dominant in the total SNR of the system. Various mmWave LNAs have been proposed in the literature, including narrowband [116], and wideband [117, 118] topologies. This chapter discusses the design techniques for mmWave low noise amplifiers consisting of narrowband, wideband and multi-band architectures. The main target applications for these LNAs are 5G NR mmWave receiver front-end systems. LNAs are designed and fabricated using 45 nm CMOS SOI process technology. The outcome of this chapter is based on the published work of VII and VIII.

4.1 40 GHz LNA

Research article VII discusses the design of a 40 GHz LNA using 45 nm CMOS SOI technology from Global Foundries. Figures 47 and 48 show the simplified schematic and the micrograph of the LNA. The LNA is composed of two cascade stages and provides single-ended to differential conversion. A common-source inductively degenerated cascode topology is used as the input stage due to its good input matching characteristics.

In the first stage, the layout of the active devices was drawn using the multi-gate technique discussed in the article VI. The size of the input devices was selected in accordance with the required current density for minimum noise figure. Noise matching was achieved by adjusting the source and gate inductances to match the optimum gate impedance of the device for noise matching. Due to its compact layout, a CPW line was used as the source degeneration inductor. The second stage was composed of differential CS amplifier. Transformer based resonators are utilized as a load for both stages. The transformers were designed to achieve maximum coupling characteristics between the primary and secondary coils to resonate the load capacitance at 40 GHz with its first
resonance as described in Section 3.8. The first stage transformer also performed a balun function to convert the single-ended signal into a differential signal. The total area of the LNA including GSG pads is 350 $\mu$m $\times$ 1061 $\mu$m. The LNA was measured using GSG probes and de-embedding was performed up to the probe tips using external calibration structures. The measured and simulated s-parameters and noise figure from single-ended output of the LNA were plotted in Figure 49. The LNA achieved a peak gain of 10.2 dB and a noise figure of 4.2 dB at 40 GHz. Input and output return loss (S11 and S22) showed a slight downward frequency shift compared with simulations, which resulted in a reduction of the gain. The measured noise figure of the LNA was closed to the simulation value. The LNA consumed 13 mW from a voltage supply of 1V. This design was an effort to understand and validate the design methods of the layout in order to achieve the optimum performance of the overall building block at 40 GHz frequency. To benchmark the performance fo the LNA, a figure of merit (FOM) was used and is defined as

$$FOM = \frac{Gain[abs] \times 1000}{(F - 1)[abs] \times P_{DC}[mW]} \left(\frac{\text{Watt}}{\text{Watt}}\right)^{-1},$$

(23)

where $F$ is the noise factor, i.e, $F = 10^{NF/10}$. Key measured results were compared with published LNAs at 40 GHz in Table 6. The proposed LNA demonstrated the lowest power consumption with one of the best FOM, except [119].
Fig. 48. Micrograph of 40 GHz LNA (Reprinted, with permission from Publication VII ©2020 Springer).

Fig. 49. Measured results of 40 GHz LNA (a) s-parameters, (b) noise figure (Reprinted, with permission from Publication VII ©2020 Springer).
Table 6. Performance comparison of 40 GHz LNA with existing SOA LNAs (Reprinted, with permission from Publication VII ©2020 Springer).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>This work</th>
<th>[120]</th>
<th>[119]</th>
<th>[121]</th>
<th>[122]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>45 nm SOI</td>
<td>130 nm CMOS</td>
<td>90 nm CMOS</td>
<td>65 nm CMOS</td>
<td>90nm SOI</td>
</tr>
<tr>
<td>Freq (GHz)</td>
<td>40</td>
<td>38</td>
<td>37</td>
<td>42</td>
<td>26-42</td>
</tr>
<tr>
<td>Max. Gain (dB)</td>
<td>10.2</td>
<td>12.6</td>
<td>13.8</td>
<td>14.3</td>
<td>11.9</td>
</tr>
<tr>
<td>NF (dB)</td>
<td>4.2</td>
<td>4</td>
<td>3.8</td>
<td>6</td>
<td>4.2</td>
</tr>
<tr>
<td>$P_{\text{diss}}$ (mW)</td>
<td>13</td>
<td>24</td>
<td>18</td>
<td>43.2</td>
<td>40.8</td>
</tr>
<tr>
<td>$P_{\text{1dB}}$ (dBm)</td>
<td>-11</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>FOM ((Watt)^{-1})</td>
<td>152</td>
<td>117</td>
<td>194</td>
<td>40</td>
<td>58</td>
</tr>
<tr>
<td>ChipArea (mm^2)</td>
<td>0.31</td>
<td>0.252</td>
<td>0.48</td>
<td>0.286</td>
<td>0.18</td>
</tr>
</tbody>
</table>

4.2 Millimeter wave frequency reconfigurable LNAs

Frequency reconfigurable LNAs (FR-LNAs) have been commonly used for supporting multiple communication standards using shared hardware at lower frequency ranges [123, 124, 125, 126]. MmWave multiband structures or resonators have been extensively used for voltage controlled oscillator designs [127, 128, 129]. A frequency reconfigurable RF/mmWave LNA requires that either (or both) of their input matching network and load have the capability to switch its resonance frequencies. Four possible design topologies of FR-LNAs are shown in Figure 50. Figure 50a and 50b are based on a single-input and a single-output topology. In Figure 50a, a single wideband matching network is utilized at the input covering all the possible frequency bands with a switchable load at the output. Although this technique benefits from smaller area footprint, however, band selectivity at the input is the worst amongst other options. A tunable matching network can replace the wideband network, as shown in Figure 50b. This technique requires the use of switches at the input of the LNA which leads to an additional loss or noise figure. A third approach of an input matching network is shown in Figure 50c, which utilizes a separate matching network for each frequency band, and the selection of the band (at input) is performed in its transconductance block. Though this approach offers a good selectivity option at the input, it requires more area for the implementation. Similarly, a dedicated load network for each band can replace the tunable load at the output but at the expense of larger chip area (Figure 50d).

This section describes the design of frequency reconfigurable LNAs for mmWave frequency bands including 24 GHz, 28 GHz and 39 GHz from the 5GNR standard. A
digitally controlled switchable resonator load based on the work presented in paper VIII is used for the resonance tuning of the LNAs. These FR-LNAs are designed and fabricated using 45 nm CMOS SOI technology from Global Foundries. Sub-sections 4.2.1 and 4.2.2 describe the design of FR-LNAs based on the architecture in Figure 50a and 50b, respectively.

4.2.1 Frequency reconfigurable LNA1

A simplified schematic of FR-LNA1 is shown in Figure 51a. The LNA design is based on the architecture in Figure 50a, where the input matching of the LNA is implemented using a wideband matching technique, as shown in Figure 51b. LNA is composed of two stages using transistors M1 to M4, where a tunable resonance load consisting of a combination of switching capacitors (C1 and C2) and a variable inductor is utilized.
resonant load is designed in such a way that the minimum variation in the quality factor is observed at all possible tuning states. A similar load is utilized for both cascading stages.

The micrograph of the LNA including GSG pads is shown in Figure 52. The simulated results of the s-parameters and noise figure at different frequency band settings are shown in Figure 53, while the measured s-parameters and noise Figure of the FR-LNA1 at two band settings are shown in Figure 54.

Fig. 51. Schematic of LNA1 (Reprinted, with permission from Publication VIII ©2020 IEEE).
Fig. 52. Chip micrograph of LNA1 (Reprinted, with permission from Publication VIII ©2020 IEEE).

Fig. 53. Simulation results of FR LNA1 (a) s-parameters (b) noise figure (Reprinted, with permission from Publication VIII ©2020 IEEE).

The S21 response of LNA at 39 GHz band setting is aligned well with the simulations, while lower frequency bands are shifted downwards, resulting in a reduced gain. This is due to the fact that the ground return path for resonator capacitors were not modelled perfectly during the design phase, as shown in Figure 55. The input and output resonators see additional and unwanted inductances of 150pH and 90 pH, respectively. These additional inductances in the ground return path caused the downward frequency shift in the load resonance which resulted in a decreased gain. However, the input return loss was close to the simulations due to which noise figure values were not impacted from the aforementioned modelling discrepancy.
4.2.2 **Frequency reconfigurable LNA2**

The FR-LNA2 is designed by using three parallel inputs for each frequency band, a simplified schematic of which is shown in Figure 56. This design is similar to the architecture presented in Figure 50(c), where input circuit consists of a matching network together with its common-source transistor (M1 to M3). The collector terminals of these common-source transistors are connected to a single common-gate transistor (M4). At a given frequency band setting, a single input transistor is biased on keeping the remaining
two transistors in an off-state. A similar tunable resonator load as in FR-LNA1 is used in FR-LNA2; the micrograph of the LNA is shown in Figure 57, the fabricated area of which is $1028\mu m \times 677\mu m$, including the input and output GSG pads. The simulated and the measured results at different band settings are shown in Figures 58 and 59, respectively. The measured results at lower frequency band settings show a similar discrepancy due to the modelling inaccuracy of the ground return path. The input return loss of the different inputs and noise Figure are close to the simulated values.

These two proposed FR-LNAs provide an insight into two different kinds of scalable architectures for 5G (mmWave) communication systems. Furthermore, this work provided an overview for the challenges involved in designing the complex switchable structures for mmWave frequency operations. As the frequency increases, the parasitic components associated with the MOS switches become more critical and, therefore, impose more stringent limitations to their full operations. Having said that, a single tuning element, e.g., switched capacitor, is not sufficient to tune a larger fractional bandwidth at mmWave frequencies. With the help of two tuning elements, i.e., an inductor and a capacitor, it becomes possible to achieve the targeted frequency range. The on-wafer measurement results of both FR-LNAs are compared with the published wideband LNAs around a similar frequency range in Table 7, where it can be observed that the presented LNAs offer sufficient performances for key parameters and could be

![Fig. 56. Schematic of LNA2 (Reprinted, with permission from Publication VIII ©2020 IEEE).](image-url)
Fig. 57. Chip micrograph of FR LNA2 (Reprinted, with permission from Publication VIII ©2020 IEEE).

Fig. 58. Simulation results of an FR LNA2 (a) S-parameters (b) Noise figure (Reprinted, with permission from Publication VIII ©2020 IEEE).
Fig. 59. Measurement results of an FRLNA2 (a) s-parameters (b) noise figure (Reprinted, with permission from Publication VIII ©2020 IEEE).

used towards the implementation of frequency scalable systems for multi-standard 5G mmWave systems.

4.2.3 Comparison of different topologies

Two proposed LNA structures have their respective trade-offs depending on the requirement of the system in which they are implemented. For example, the FR-LNA1 would be a suitable choice for a phased array system due to its compact size that can be integrated into a single receiver channel. However, the wideband input matching of this structure could lead to its intolerance for unwanted blockers at different frequencies that could result in the distortion of the whole system. Contrarily, the FR-LNA2 structure offers good filtering properties at each input which relaxes the overall system requirements for additional filtering at the input but at the cost of increased area footprint.

The design of the three mmWave LNAs covering frequency bands n257, n258 and n260 of 5G New Radio standards were presented in this chapter. The 40 GHz LNA design was based on a cascode device with a multi-gate layout technique which benefits from the reduced resistance of the wiring at the gate terminal of the common-emitter device. Different architecture options for frequency reconfigurable LNAs were also analysed based on the tunable resonant load, the drawn area and implementation complexity for mmWave frequency operations. Two FR-LNAs structures were designed using a similar resonant load that enabled the frequency response tuning from 24 GHz to
39 GHz. These FR-LNA structures provided good insights into the challenges involved in designing other reconfigurable building blocks at mmWave frequencies, e.g., PA, buffers and mixers, etc.

Table 7. Performance comparison of FR LNAs (Reprinted, with permission from Publication VIII ©2020 IEEE).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>FR-LNA1</th>
<th>FR-LNA2</th>
<th>[130]</th>
<th>[46]</th>
<th>[122]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>45 nm SOI</td>
<td>45 nm SOI</td>
<td>65 nm</td>
<td>45 nm</td>
<td>90 nm</td>
</tr>
<tr>
<td>Configuration</td>
<td>Multi-band</td>
<td>Multi-band</td>
<td>WB</td>
<td>WB</td>
<td>WB</td>
</tr>
<tr>
<td>Freq (GHz)</td>
<td>22/39</td>
<td>24/28/39</td>
<td>16-30</td>
<td>24-28</td>
<td>26-42</td>
</tr>
<tr>
<td>Max. Gain (dB)</td>
<td>8.5/12.9</td>
<td>9.5/12/15.5</td>
<td>10.2</td>
<td>8.5</td>
<td>11.9</td>
</tr>
<tr>
<td>NF (dB)</td>
<td>3.8/4.9</td>
<td>4.5/4.5/5.5</td>
<td>3.3-5.7</td>
<td>4.0</td>
<td>3.6-4.2</td>
</tr>
<tr>
<td>$P_{\text{diss}}$ (mW)</td>
<td>15.5</td>
<td>20.68</td>
<td>12.4</td>
<td>12</td>
<td>40.8</td>
</tr>
<tr>
<td>$P_{\text{1dB}}$ (dBm)</td>
<td>-11/-16</td>
<td>-12/-13/-16.5</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>ChipArea $(\text{mm}^2)$</td>
<td>0.317</td>
<td>0.69</td>
<td>0.135</td>
<td>N/A</td>
<td>0.18</td>
</tr>
</tbody>
</table>


5 Conclusions

The main objective of this thesis was to study the various design techniques of mmWave integrated circuits towards the implementation of complex 5G radio systems. In order to meet the comprehensive research target, the goal was divided into several research questions. These questions mainly target the following research topics: the study of the various types of beamforming architectures and their optimum use in energy hungry 5G systems, the development of a phased array receiver FE IC with multiple data streams and off-chip filtered input matching network design, the conductive measurements of a single channel, the layout design techniques for active and passive devices to achieve optimum performance at mmWave frequencies, electromagnetic modelling techniques, and the design and analysis of various mmWave LNAs covering 5G mmWave frequency bands. The author’s main contributions and those which attempted to answer the aforementioned research questions are summarized in the original peer-reviewed research articles I - VIII.

First, various beamforming architectures were studied to understand different key parameters, e.g., area, power consumption, implementation complexity. Then, a fully integrated four-elements phased array receiver FE was designed supporting two simultaneous data streams implemented within the same silicon die. In order to reduce its complexity, this design was based on RF beamforming architecture. An IQ vector modulator topology was utilized to realize the programmable phase shift with fine step tuning. Input matching of the receiver channels was implemented by using off-chip transmission lines based impedance transformers which provided an insight into how to implement impedance transformation in a multi-technology environment.

Second of all, understanding the layout related effects from interconnects in active and passive devices has a significant importance in designing integrated circuits for mmWave frequencies. Various layout techniques of these devices were studied and designed to find the major constraints to reach their optimum performance. Two cascode connected active devices with separate layout styles were designed and characterised, and their maximum frequency operations were compared. The accurate electromagnetic modelling of passives is very important to model the parasitic effects (e.g., unwanted coupling) of a passive device with its neighbouring environment. A strategy of port definition, selection of an appropriate ground plane, and an overall methodology was proposed in this thesis. A 40 GHz 2-stage LNA was designed based on the aforementioned techniques for active and passive devices. A differential reflection-type phase shifter based on a vertically-coupled CPW line coupler was demonstrated. The
design of this phase shifter was an attempt to integrate big $\lambda/4$ based structures to fit in a phased array channel with a compact area. MmWave signal combiners for phased arrays receiver are briefly discussed on the basis of drawn area, power consumption, isolation and insertion loss. Furthermore, a 4-way compact signal combiner based on a current-combining technique using transformers is proposed and compared with a traditional Wilkinson power combiner.

Thirdly, multiple frequency bands for 5G mmWave systems have been allocated by 3GPP standards, which has created the need for designing a separate FE system for each frequency band. However, a single wideband FE system can serve all the frequency bands at the same time, or alternatively, a frequency reconfigurable FE system can also be helpful with additional benefits of filtering. Two separate frequency reconfigurable LNA structures were proposed to switch all possible frequency bands, i.e., 24GHz, 28GHz and 39 GHz. The first structure (FR-LNA1) employed a single input and single output structure with a wideband input matching network. The second approach (FR-LNA2) used separate inputs with their respective matching networks for each frequency band, i.e., 24GHz, 28GHz and 39 GHz. A tunable resonant load is proposed based on a magnetically tuned variable inductor and switched capacitors. Two structures utilized a similar resonant load, the performance of which were verified from their implemented prototypes. FR-LNA1 offered better noise performance because there was no switch utilized at the input. However, FR-LNA2 demonstrated better filtering properties at the input that could relax the input filtering requirements of the overall system. Due to the compactness of the FR-LNA1, it could be the preferred choice for integration in a phased array channel.

Successful development of silicon-based complex 5G mmWave system requires prior understanding of the peak performance of active devices offered by the process technology, design methodologies to reduce the layout related parasitic effects, and modelling of these unwanted parasitic effects. Methods and design techniques proposed in this thesis will assist the seamless processes in the development of future silicon-based mobile communication systems. For example, the design of the 4 × 2 input matching network presented in section 2.5 offers insights into the co-designing approach for those designs consisting of multiple technologies. The signal combiners presented in sections 2.5 and 3.10 provide a good understanding of the compact implementation of such signal combiners in phased array systems. Similarly, the observations in the design of FR-LNAs presented in section 4.2 are beneficial in the implementation of multi-standard radios for mmWave communication systems. Moreover, these findings should be extended further to improve the design methodologies of the complex mmWave systems. One suggestion for the future development of such systems is to co-design of the silicon
ICs together with their relevant die packages, where part of the input or output matching networks can be implemented in the package to improve the transition losses.
References


B. K. Esfeh, S. Makovejev, D. Basso, E. Desbonnets, V. Kilchytska, D. Flandre, and J. Raskin, “RF SOI CMOS technology on 1st and 2nd generation trap-rich high resistivity


Original publications


Reprinted with permission from IEEE (I, III, IV, V, VI, VIII) and Springer (II, VII)

Original publications are not included in the electronic version of the dissertation.
<table>
<thead>
<tr>
<th>Number</th>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>794</td>
<td>Oppenlaender, Jonas</td>
<td>(2021) Crowdsourcing creative work</td>
</tr>
<tr>
<td>796</td>
<td>Liu, Chen-Feng</td>
<td>(2021) Resource allocation for ultra-reliable and low-latency 5G communication</td>
</tr>
<tr>
<td>797</td>
<td>Kijpjärvi, Joni</td>
<td>(2021) RF-microwave sensor development for cell and human in vitro and ex vivo monitoring</td>
</tr>
<tr>
<td>798</td>
<td>Hukari, Jarkko</td>
<td>(2021) 2D CMOS SPAD array techniques in 1D pulsed TOF distance measurement applications</td>
</tr>
<tr>
<td>799</td>
<td>Peyvasteh, Mostahareh</td>
<td>(2021) Polarimetric and spectral imaging approaches for quantitative characterization of inhomogeneous scattering media including biotissues</td>
</tr>
<tr>
<td>800</td>
<td>Shehab, Mohammad</td>
<td>(2021) Energy efficient QoS provisioning and resource allocation for machine type communication</td>
</tr>
<tr>
<td>802</td>
<td>Nellattukuzhi Sreenivasan</td>
<td>(2021) Synthesis and alkali activation of Magnesium-rich aluminosilicates</td>
</tr>
<tr>
<td>803</td>
<td>Kallio, Johanna</td>
<td>(2021) Unobtrusive stress assessment in knowledge work using real-life environmental sensor data</td>
</tr>
<tr>
<td>804</td>
<td>Meriö, Leo-Juhani</td>
<td>(2021) Observations and analysis of snow cover and runoff in boreal catchments</td>
</tr>
<tr>
<td>805</td>
<td>Lovén, Lauri</td>
<td>(2021) Spatial dependency in Edge-native Artificial Intelligence</td>
</tr>
<tr>
<td>806</td>
<td>Törmänen, Matti</td>
<td>(2021) Improved analysis of tube flow fractionation data for measurements in the pulp and paper industry</td>
</tr>
<tr>
<td>807</td>
<td>Rusanen, Annu</td>
<td>(2021) Catalytic conversion of sawdust-based sugars into 5-hydroxymethylfurfural and furfural</td>
</tr>
<tr>
<td>808</td>
<td>Yaraghi, Navid</td>
<td>(2021) Analyzing human impacts on the quality and quantity of river water</td>
</tr>
<tr>
<td>809</td>
<td>Tarakanchikova, Yana</td>
<td>(2021) Multilayered polyelectrolyte assemblies as delivery system for biomedical applications</td>
</tr>
</tbody>
</table>

Book orders: Virtual book store
http://verkkokauppa.juvenesprint.fi
Rana Azhar Shaheen

DESIGN ASPECTS OF MILLIMETER WAVE MULTIBAND FRONT-ENDS