Analog RFIC and MMIC designs for mm-wave mobile backhaul radios

Marcus Gavell

Microwave Electronics Laboratory
Department of Microtechnology and Nanoscience (MC2)
Chalmers University of Technology
Göteborg, Sweden, 2018
Analog RFIC and MMIC designs for mm-wave mobile backhaul radios

MARCUS GAVELL

© Marcus Gavell, 2018

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 4493
ISSN 0346-718X

Technical Report MC2-402
ISSN 1652-0769

Chalmers University of Technology
Department of Microtechnology and Nanoscience - MC2
Microwave Electronics Laboratory
SE-412 96 Göteborg, Sweden
Phone: +46 (0) 31 772 1000

Printed by Chalmers Reproservice
Göteborg, Sweden 2018
Abstract

The increasing use of mobile services engage prospective challenges in the development of the mobile network infrastructure. Innovative and environmentally sustainable hardware solutions are necessary to accommodate higher data-rates at low cost and to meet authority regulations. This thesis presents advances and innovative circuit topologies in front-end electronics for mobile backhaul to facilitate and encourage the use of wideband communication at mm-wave frequencies. Circuit demonstrators such as power amplifier, analog pre-distorter and modulator MMICs and a variable gain amplifier RFIC have been designed, analyzed and characterized.

Enhancement techniques for increasing the output power, efficiency and linearity are demonstrated with two power amplifiers, one at the V-band (57-64 GHz) and one at E-band (71-76, 81-86 and 92-95 GHz). Using stacked transistors at the output stage, higher output power and gain can be obtained. In microstrip configuration, the stacked transistor is prone to be unstable. This is addressed with a new stacked transistor layout configuration for increased stability. The new stacked transistor layout is supported by simulations and is verified by small and large signal measurements. 25 dBm output power and 15% PAE are demonstrated with a power amplifier MMIC at 60 GHz. Linearity is addressed by designing an analog pre-distorter at the E-band whose nonlinear response mimics the inverse response of the power amplifier. The pre-distorter and power amplifier were characterized by two-tone excitation and a modulated signal, demonstrating carrier to third order intermodulation better than 40 dBc up to 17.5 dBm output power and, reduced spectral regrowth at 17 dBm average output power using 64 QAM.

Targeting the mm-wave wideband point-to-point communication windows at E-band and D-band (141-148.5, 151.5-164 and 167-174.8 GHz), two linear quadrature modulators have been designed. The E-band design covers 70-95 GHz and incorporates an integrated differential branchline coupler and common-mode rejection filters to enhance LO-RF port isolation. LO to RF isolation higher than 30 dB, conversion loss of 11 dB and third order intercept point of 13 dBm have been achieved. At D-band, a novel conversion-efficient subharmonic topology that isolates the carrier from the output has been proposed and demonstrated. Conversion loss of 11 dB and linear output power up to -2 dBm have been measured. Employing input dc offsets to further suppress the $2 \times LO$ at the RF port, the $2 \times LO$ to RF suppression was measured > 70 dB. This enables conversion-efficient topologies to be used in E-band and D-band transmitters.

For high data-rate communication, a wideband variable gain amplifier is de-
signed in a Large-Scale-Integration high-speed InP DHBT process. The design and layout methodology is presented, including frequency bandwidth analysis by means of Zero Value Time constants of the complete circuit. State-of-the-art results are obtained, dc-40 GHz frequency coverage, maximum gain of 31 dB and more than 44 dB gain control range is measured. Using Bi-Phase-Shift-Keying modulated signals, the variable gain amplifier circuit demonstrated 44 Gbit/s high data-rate.

The innovative hardware implementations in this thesis, supported by measurement results, investigate important development areas for future mobile backhaul. The circuit demonstrators aid to address system level issues and assist in future circuit designs.

**Keywords:** Analog pre-distortion, DHBT, GaAs, HEMT, InP, microwave, mixers, mm-wave, modulators, mobile backhaul, PA, Point-To-Point, power amplifiers, stacked HEMT, variable gain amplifier, VGA
# Abbreviations and notations

## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM-AM</td>
<td>Amplitude-to-Amplitude distortion</td>
</tr>
<tr>
<td>AM-PM</td>
<td>Amplitude-to-Phase distortion</td>
</tr>
<tr>
<td>APD</td>
<td>Analog Pre-Distortion</td>
</tr>
<tr>
<td>BEOL</td>
<td>Back End Of Line</td>
</tr>
<tr>
<td>CAD</td>
<td>Computer Aided Design</td>
</tr>
<tr>
<td>C-B</td>
<td>Common-Base</td>
</tr>
<tr>
<td>C-E</td>
<td>Common-Emitter</td>
</tr>
<tr>
<td>C-S</td>
<td>Common-Source</td>
</tr>
<tr>
<td>CI3</td>
<td>Carrier to third order intermodulation</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal Oxide Semiconductor.</td>
</tr>
<tr>
<td>CMRR</td>
<td>Common-Mode Rejection Ratio</td>
</tr>
<tr>
<td>dc</td>
<td>direct current</td>
</tr>
<tr>
<td>DHBT</td>
<td>Double Heterojunction Bipolar Transistor</td>
</tr>
<tr>
<td>DPD</td>
<td>Digital Pre-Distortion</td>
</tr>
<tr>
<td>EM</td>
<td>Electro-Magnetic</td>
</tr>
<tr>
<td>ETSI</td>
<td>European Telecommunications Standards Institute</td>
</tr>
<tr>
<td>EVM</td>
<td>Error Vector Magnitude</td>
</tr>
<tr>
<td>FCC</td>
<td>Federal Communications Commission</td>
</tr>
<tr>
<td>FEOL</td>
<td>Front End Of Line</td>
</tr>
<tr>
<td>FET</td>
<td>Field Effect Transistor</td>
</tr>
<tr>
<td>FOM</td>
<td>Figure-Of-Merits</td>
</tr>
<tr>
<td>GBP</td>
<td>Gain-Bandwidth-Product</td>
</tr>
<tr>
<td>HB</td>
<td>Harmonic Balance</td>
</tr>
<tr>
<td>HEMT</td>
<td>High Electron Mobility Transistor</td>
</tr>
<tr>
<td>IC</td>
<td>Integrated Circuit</td>
</tr>
<tr>
<td>IF</td>
<td>Intermediate Frequency</td>
</tr>
</tbody>
</table>
IM3  Third Order Intermodulation
IM5  Fifth Order Intermodulation
IRR  Image Rejection Ratio
LO   Local Oscillator
LS   Large Signal
LTI  Linear Time Invariant
LTV  Linear Time Variant
mHEMT metamorphic HEMT
MIM  Metal-Insulator-Metal
MMIC Monolithic Microwave Integrated Circuit
MNA Modified Node Admittance
NA Node admittance
NF Noise Figure
OIP3 Output third order intercept point
P2P Point-To-Point
PA Power Amplifier
PAE Power Added Efficiency
PAM Pulse-Amplitude Modulation
pHEMT pseudomorphic HEMT
$P_{SAT}$ Saturated Output Power
QAM Quadrature Amplitude Modulation
RF Radio Frequency
RFIC Radio Frequency Integrated Circuit
RL Return Loss
SNR Signal-to-Noise Ratio
SP Scattering Parameters
TEM Transverse Electro-Magnetic
VCCS Voltage Controlled Current Source
VGA Variable Gain Amplifier
ZVT Zero-Value-Time

Notations

$f_T$ Transition frequency
$f_{max}$ Maximum oscillation frequency
$f_{LO}$ LO frequency
$f_{IF}$ IF frequency
$f_{RF}$ RF frequency
$L_W$ Gate/Emitter width
List of Publications

Appended Publications

This thesis is based on work contained in the following papers:

[A] M. Gavell, I. Angelov, M. Ferndahl and H. Zirath "A V-band Stacked HEMT Power Amplifier with 25 dBm Saturated Output Power in 0.1 \( \mu \)m InGaAs Technology” in IEEE Transactions on Microwave Theory and Techniques, vol. 64, nr. 12, pp. 4232-4240, December 2016. DOI: 10.1109/TMTT.2016.2613849


[D] M. Gavell, S. E. Gunnarsson, M. Ferndahl and H. Zirath ”A D-band second harmonic quadrature modulator MMIC with harmonically isolated RF port” Submitted to IEEE Transactions on Microwave Theory and Techniques

Other Publications

The following papers have been published but is not included in the thesis. The content partially overlaps with the appended papers or is out of the scope of this thesis.


[b] S. E. Gunnarsson, M. Gavell, D. Kuylenstierna and H. Zirath ”60 GHz MMIC double balanced Gilbert mixer in mHEMT technology with integrated RF, LO and IF baluns” ELECTRONICS LETTERS 23rd November 2006 Vol. 42 No. 24


# Contents

<table>
<thead>
<tr>
<th>Abstract</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Publications</td>
<td>v</td>
</tr>
<tr>
<td>Abbreviations and Notations</td>
<td>vii</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Mobile backhaul</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Thesis contribution</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Thesis outline</td>
<td>6</td>
</tr>
<tr>
<td><strong>2 Power amplifier enhancement techniques and designs</strong></td>
<td>7</td>
</tr>
<tr>
<td>2.1 Non-linear GaAs pHEMT transistor model</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Transistor operating classes</td>
<td>10</td>
</tr>
<tr>
<td>2.3 mm-wave power amplifiers</td>
<td>13</td>
</tr>
<tr>
<td>2.3.1 Power combining</td>
<td>15</td>
</tr>
<tr>
<td>2.3.2 Budgeting multiple transistors PAs</td>
<td>17</td>
</tr>
<tr>
<td>2.3.3 Stacked HEMT layout and verification</td>
<td>19</td>
</tr>
<tr>
<td>2.3.4 Stacked HEMT PA design</td>
<td>22</td>
</tr>
<tr>
<td>2.4 Linearization</td>
<td>23</td>
</tr>
<tr>
<td>2.4.1 Implementation overview</td>
<td>24</td>
</tr>
<tr>
<td>2.4.2 Circuit simulation environment</td>
<td>25</td>
</tr>
<tr>
<td>2.4.3 Class C transistor pre-distorter</td>
<td>30</td>
</tr>
<tr>
<td><strong>3 Mixer theory and design of linear quadrature modulators</strong></td>
<td>33</td>
</tr>
<tr>
<td>3.1 Mixing fundamentals</td>
<td>34</td>
</tr>
<tr>
<td>3.2 The resistive FET mixer</td>
<td>35</td>
</tr>
<tr>
<td>3.3 Mixers topologies</td>
<td>37</td>
</tr>
<tr>
<td>3.3.1 Fundamental and harmonic mixers</td>
<td>38</td>
</tr>
<tr>
<td>3.3.2 Balanced mixers</td>
<td>39</td>
</tr>
<tr>
<td>3.3.3 The quadrature mixer</td>
<td>42</td>
</tr>
<tr>
<td>3.3.4 Single sideband mixer</td>
<td>43</td>
</tr>
<tr>
<td>3.4 Quadrature modulator designs for transmitters</td>
<td>44</td>
</tr>
<tr>
<td>3.4.1 Electromagnetic modes and transmission lines</td>
<td>44</td>
</tr>
<tr>
<td>3.4.2 Circuit simulation environment</td>
<td>45</td>
</tr>
<tr>
<td>3.4.3 Fundamental balanced E-band modulator</td>
<td>46</td>
</tr>
<tr>
<td>3.4.4 Subharmonic harmonically balanced D-band modulator</td>
<td>49</td>
</tr>
<tr>
<td>4 Baseband Variable Gain Amplifier for wideband operation</td>
<td>53</td>
</tr>
<tr>
<td>------------------------------------------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>4.1 Design considerations and bandwidth analysis</td>
<td>54</td>
</tr>
<tr>
<td>4.1.1 Transfer response</td>
<td>54</td>
</tr>
<tr>
<td>4.1.2 Zero Value Time constants</td>
<td>55</td>
</tr>
<tr>
<td>4.1.3 Common-mode</td>
<td>58</td>
</tr>
<tr>
<td>4.2 RFIC technologies and 3-D BEOL layout</td>
<td>59</td>
</tr>
<tr>
<td>4.3 Main results</td>
<td>62</td>
</tr>
</tbody>
</table>

| 5 Conclusions and future outlook                           | 65  |

Acknowledgments                                             67

Bibliography                                                69
Chapter 1

Introduction

Communication and dissemination of information have, since the introduction of the Internet, revolutionized our lifestyle and radically changed everyday’s functions. What was initiated in the late 60’s by the Defense Advanced Research Projects Agency (DARPA) program in order to minimize the universities computer systems’ vulnerability by joining them together (codenamed ARPANET at the time), has evolved to become the world’s largest computer network (Internet) used by 48% of the world’s population in the year 2018 [1]. The breakthrough of the Internet and the introduction of wireless internet access through the second generation mobile phone systems (2G or GSM) came in 1991. For the first time, digital communication protocols allowed mobile data services to access Internet from the mobile phone. In 1999, there were 100 millions mobile Internet subscribers worldwide, with a maximum data speed of 9.6 kbps [2,3]. In 2012, there were 6.3 billions subscriptions (1 EB/month), 2017 there were 7.6 billions subscriptions (9.5 EB/month) and it is expected to reach 9 billions subscriptions (71 EB/month) in 2022 [4,5] (Fig. 1.1).

The increasing use of streamed music, TV, films, series, mobile video phone calls and social medias require higher data rates, higher throughput and low latency in the mobile networks. Customer demand to access those services at a low cost are driving factors in the development of future mobile backhaul radios. The operators’ revenue model has over the recent years changed from a profitable model based on fees for calls, messages and data usage to a flat rate subscription model which has led to dramatically decreased revenues. The flat rate model makes revenue increasingly connected to the direct operational costs [10]. License fees, higher power efficiency and higher throughput are factors that play important roles in lowering the operational cost and the cost per bit. Moreover, there are emerging worldwide energy regulations and policies, which incite the development of environmentally friendly products [11].

1.1 Mobile backhaul

The mobile backhaul, the wireless communication link between the radio base stations and the core network of the Internet, has become vital in the expan-
CHAPTER 1. INTRODUCTION

Figure 1.1: Global mobile data traffic 5 year forecast including the three previous years [6–9].

sion of mobile services, both in terms of availability and higher data rates. In particular, it is the microwave backhaul network that is the key solution to provide fast, reliable and cost effective mobile broadband services to the user. The traditional microwave bands (6-42 GHz) have relatively narrow channel bandwidths, up to 112 MHz in the 42 GHz band, whereas, for instance the E-band (71-76 and 81-86 GHz), has defined channel bandwidths of 250 MHz which can be aggregated to a maximum of 5 GHz, allowing for higher data rates to be transmitted [12]. The global trend and the need for higher capacity in the mobile backhaul networks make the E-band microwave point-to-point (P2P) radio and fiber markets to grow [11]. Emerging frequency bands within the W-band (94.1-100 GHz and 102-109.5 GHz) and D-band (141-148.5 GHz, 151.5-164 GHz and 167-174.8 GHz) are given full consideration by the European Telecommunications Standards Institute (ETSI) for P2P communication, shown in Fig. 1.2a. Aggregated channels up to 10 GHz, two times higher than the E-band [13], a low cost license fee model, similar to the E-band and appealing small form factor are incentive to attract operators to choose the D-band for high speed communication. For these reasons, the D-band has large potential of becoming the next generation high speed communication band. It is expected that 65% of all backhaul networks will be microwave based in 2021 [12].

The atmosphere’s low attenuation at microwave and mm-wave frequencies is one enabler of wireless communication. The Electronic Communications Committee (ECC) within the European Conference of Postal and Telecommunications Administrations (CEPT), ETSI, Federal Communications Commission (FCC) and other spectrum organizations around the world organize, allocate and regulate spectra to cater for every applications’ needs [16]. For P2P communication, low attenuation is beneficial in order to make long distance communication possible. In Fig. 1.2a, the atmospheric attenuation of electromagnetic (EM) waves (dB/km) versus frequency is plotted at different rainfall intensities. Three characteristics are observed from the readout: 1)
1.1. MOBILE BACKHAUL

The attenuation increases with frequency, 2) There are absorption peaks at 23, 60, 118 and 183 GHz and 3) Rain plays a significant role in the attenuation level. It comes to no surprise, that the allocated spectra for P2P radio links avoid the strong absorption peaks. The traditional microwave bands are located below the 60 GHz oxygen absorption line, the E-band sits between the 60 GHz and the 118 GHz lines and the D-band above the 118 GHz but below the 183 GHz lines. However, from Fig. 1.2a it becomes obvious that rain will have a significant impact on the link budget, ranging from tenths of dBS to tens of dBS per kilometer depending on the weather conditions. A radio therefore requires a wide gain control range, high output power and low system noise to sustain operational over a wide range of weather conditions. These challenges are addressed in this thesis.

In Fig. 1.2b, a microwave backhaul radio installation is shown. The radio is a complex system that consists of many digital and analog circuits and components, digital and analog interfaces, see Fig. 1.3. In this context, the focused area and the scope of this thesis is analog circuit designs, whose function in the radio is to convert baseband data-streams to a carrier frequencies and back by transmitter and receiver circuits (front-ends Fig. 1.4). The front-end circuits are situated on a higher level than semiconductor materials and processes but below system level. In Fig. 1.4 the line-up of components such as modulators, local oscillators and various types of amplifiers are shown. The same figure shows the circuits that form the basis for this thesis.

Semiconductor materials and fabrication processes have a direct impact on the system performance. Different technologies are constantly under development, targeting different niches. From the development of the first fabricated
CHAPTER 1. INTRODUCTION

SYSTEM

PACKAGE

CIRCUIT DESIGN

SEMICONDUCTOR TECHNOLOGIES

Frequency, modulation, data-rates

High frequency interfaces, digital interfaces, thermal...

Digital: Modem, ASICs, FPGAs

Analog: Modulators, Amplifiers, VGA...

Semiconductor materials

Processing

Figure 1.3: System hierarchy of a mobile backhaul radio, where the focus area of this thesis is marked in red.

Figure 1.4: Schematic of a typical radio front-end, where the circuits in this thesis are shown.

Field Effect Transistor (FET) to today’s advanced Complementary Metal Oxide Semiconductor (CMOS), SiGe BiCMOS, High Electron Mobility Transistor (HEMT) or Double Heterojunction Bipolar Transistor (DHBT) technologies is astonishing. The design of analog and digital Integrated Circuits (ICs) could not be more interesting at the moment. Computer Aided Design (CAD) tools allow implementation of complex models of both passive and active components to simulate and realize complex integrated circuits with sufficient accuracy. The enrollment of the fifth generation (5G) mobile systems will require great engineering skills and efforts to realize attractive microwave and millimeter wave (mm-wave) circuit designs to solve the challenging requirements.

1.2 Thesis contribution

In this thesis, circuit theory and design of monolithic microwave integrated circuits (MMICs) in the millimeter wave frequency range for P2P communication are presented. This work helps to improve radio performance and to develop future radio equipment.

Power combining is an important task in the design of PAs for achieving wideband frequency operation and high output power level. Present GaAs technologies targeting mm-wave frequencies, suffer from low breakdown voltage and as consequence, class AB GaAs PAs require high direct current (dc) bias currents and present low optimum load impedance. The stacked HEMT power combining technique is appealing because it combines by voltage rather
than current. This reduces the dc current and increases the optimum load impedance for the same output power. Further advantageous is the higher efficiency the technique offers for multiple gain stages. Using this technique in [A], a record in output power per chip area of any other GaAs PA in the same frequency range is demonstrated. A proposed layout of the stacked HEMT transistor offering enhanced high frequency stability is presented and analyzed.

Modern wireless communication has stringent requirements of spectral regrowth. Spectral regrowth is caused by distortion, which originates from a non-linear transfer function from baseband data to a carrier frequency. Without linearization techniques, the Power Amplifier (PA) needs to operate far below its maximum capability to comply with the spectral mask requirement, making it very energy inefficient. Digital pre-distortion (DPD) is considered the most appealing approach in terms of efficiency and implementation complexity at low channel bandwidths. However, at wide channel bandwidths, such as at the E-band, DPD needs fast processing and will for this reason degrade in efficiency. In the analog domain, analog pre-distortion (APD) has the advantage of being efficient for wide bandwidths. In [B], the fundamental linear system is explained and a class C amplifier is proposed to mimic the inverse properties of a class AB PA, and therefore has the ability to linearize the transfer function. An experiment based on a PA and APD chip-set demonstrated improved linearity and reduced spectral regrowth using modulated signals. Circuit theory and implementation of the APD is presented.

The fundamental resistive mixer in [C] presents in line with the modulator in [D] state-of-the-art results. The design comprises innovative passive components such as a common-mode reject filter and differential branchline coupler. Both help to minimize chip area and to increase the critical LO to RF isolation.

Sub-harmonic mixers are often favorable at mm-wave frequency systems in order to relax the local oscillator (LO) multiplication factor. Among known mixer topologies, the resistive FET mixer is widely spread among HEMT and FET technologies because of its high linearity and good conversion efficiency. In [D] a novel circuit topology that combines multiple conversion efficient subharmonic resistive mixers to form a harmonically balanced configuration is designed, theoretically explained and demonstrated. The radio frequency (RF) port is harmonically isolated from the LO, including the important second harmonic, which allows this mixer topology to be used in transmitters with no carrier leakage. The topology was demonstrated with a 135-170 GHz second harmonic quadrature modulator, obtaining unprecedented performance, > 70 dB2 × LO to RF suppression at dc offset injection and order of magnitude higher linearity than previously reported harmonic mixers at D-band.

High speed communication in the emerging W-band and D-band mm-wave P2P radio bands and beyond, or in optical communication equipment, require baseband components with tens of Gigahertz bandwidths. In [E] a variable gain amplifier (VGA) is designed based on a Gilbert-cell topology in one of the fastest semiconductor processes to date. By means of careful and compact layouting, taking full advantage of the process’s advanced back-end-of-line (BEOL), and pole analysis by Zero Value Time (ZVT) constants, state-of-the-art performance was achieved.
1.3 Thesis outline

This thesis is divided in four parts. In Chapter 2, transistor operation and models are covered, together with power combining techniques with emphasis on voltage combining used in [A]. General linearization objectives and methods, the APD in [B] and a co-simulation environment for the APD and PA are also presented. In Chapter 3, mixing theory, mixer topologies and the operation of the resistive mixer are covered. The circuit design methodologies in [C] and [D] follow, focusing on the phase distribution networks being a critical part of the designs. In Chapter 4, network theory and the method of ZVT constants along with design considerations for wideband operation in [E] are included. Final words and future work concludes this thesis in Chapter 5.
Chapter 2

Power amplifier enhancement techniques and designs

This chapter focuses on the realization of efficient mm-wave PAs and linearizers. The principles of PA operation and design philosophy for multi-stage designs are covered. Power combination techniques, both in current and voltage are given attention in Section 2.3.1, where the benefits and challenges of the two techniques are discussed.

Output power, linearity and efficiency are three important parameters for PAs. Quadrature Amplitude Modulation (QAM) signals used in communications, have large Peak-to-Average Power Ratio, distributing the statistical power levels across a large span. For PAs, it is therefore desirable to achieve high efficiency across that same output power range. Techniques such as load modulation and supply modulation are interesting implementations to achieve high efficiencies at back-off, where the PA mostly will operate. At microwave frequencies, efficiencies over 70% with maintained efficiency at back-off using load modulation have been demonstrated [17]. Doherty amplifiers have been demonstrated in CMOS at E-band frequencies, achieving 12% Power Added Efficiency (PAE) at peak power and twice the drain efficiency at 6 dB back-off compared to a class A amplifier [18]. At frequencies approaching $f_t/f_{max}$, load modulation becomes inefficient at back-off because of the auxiliary transistor’s lossy output reflection. On the other hand, supply modulation is an effective method to improve the efficiency of non load-modulated amplifiers at back-off, but it is difficult to implement efficiently for wide bandwidths. In [19], a 5.7 times enhancement in efficiency was demonstrated with a 20 MHz signal bandwidth at 44 GHz, using envelope tracking. At E-band frequencies, 27% PAE at 26 dBm output power at 76 GHz, and 35 dBm and 14% PAE at 74 GHz have been demonstrated in InP DHBT [20] and GaN [21], respectively. The outlook is positive for the advances in efficiency and output power by the use of the InP DHBT and GaN HEMT technologies. In addition, enhancement techniques such as stacked transistors, not limited by technology, contribute to increased output power and efficiency, which is addressed in [A], obtaining 15% PAE and
25 dBm output power in GaAs pseudomorphic HEMT (pHEMT) technology.

In the previous paragraph, an overview of output power and efficiency was presented. Hereby changing focus towards linearization; an overview and discussion follow. Under spectrum constraints or Error Vector Magnitude (EVM) requirements, high efficiency and high output power are of little use unless the response is linear. Unfortunately, efficiency and linearity are strongly correlated parameters, being their antithesis. Ultimately, it is necessary to add functionality to a system in order to achieve linear, high output power and high efficiency simultaneously. Linearization is the missing piece. While many applications and requirements exist, there is also a wide variety of implementation solutions for linearization suggested in the literature, each having its niche. Video broadcasts from satellites use wideband signals and therefore, analog linearization implementations were consolidated early within this field [22,23], while digital implementations were consolidated at an early stage within communications. A paradigm shift in communications, where wideband signals are the essence of 5G, make analog implementations attractive. In [B], we address the non-linear response from a class AB amplifier by compensating its compressed output power response with an expanding response class C pre-distorter. The circuit and theory of operation, simulation environment and measurements are covered in Section 2.4. In addition, the large signal (LS) nonlinear transistor model used in [A] and [B] is presented in Section 2.1. The emphasis is on the static non-linear parameters, to relate bias operating condition with high frequency performance and sources for distortion.

2.1 Non-linear GaAs pHEMT transistor model

Successful MMIC designs rely on accurate models. Semiconductor fabrication is expensive, time consuming and once the circuit is manufactured, there are no or few rework possibilities. For this reason, it is important to accurately describe the included components’ physical properties by models to endeavor successful MMIC designs at first attempt. Probably the most challenging and difficult component to model is the transistor. It must be precise, and accurately reproduce both the linear and non-linear transistor behavior and simultaneously be computational friendly. Today’s CAD circuit simulation tools rely on behavioral models, described by mathematical functions to express the physical behavior [24]. In this section, the chosen model for the designs in [A] and [B] is presented and a survey of model follow.

The non-linear Chalmers model [25,26], Fig. 2.1a, describes the intrinsic elements $C_{GS}, C_{GD}, C_{ds}, R_i, R_j$ and current source $I_{ds}$ with non-linear functions and the extrinsic elements $R_S, R_G, R_D, L_S, L_G$ and $L_D$ as linear elements. It considers the channel as one section (lumped) and the gate’s and drain’s reference planes are empirically adjusted by $L_G, L_D, C_{PG}$ and $C_{PD}$, making transistor scaling uncertain for a large span of gate width and number of fingers.

Although more complicated models than the Chalmers model exist, where multiple sections (distributed effects) of the channel are considered [27,30], temperature intercoupling between fingers and parallel fingers [29], temperature dependent source and drain access resistances [h], or charge distribution
along the channel \[31\] are taken into account, the Chalmers model provides a good balance between linearity, frequency range and, simulation computation time for most purposes. As a rule-of-thumb, the lumped approach is adequate when the device size is \(< \lambda/10\), i.e. electrical length of 36 degrees. It is the basis of the PAs in [A] and [B] and the pre-distorter circuit in [B]. Rather than using one scalable model, several specific sizes were developed to improve model precision and to incorporate many of the mentioned effects without the need for complicated equations, or to model interaction between transistors. The transistor sizes used in [A] and [B] were \(2 \times 37.5 \mu m\), \(2 \times 50 \mu m\), \(4 \times 50 \mu m\), \(6 \times 50 \mu m\), and models were developed accordingly. A photograph of the transistor to model is shown in Fig. 2.1\(b\).

![Non-linear electrical model of the HEMT.](image1)

![Photograph of a 4 × 50 µm pHEMT from WIN Semiconductors.](image2)

Figure 2.1: Electrical model and photograph of a 4 × 50 µm large transistor.

To assess the transistor’s high frequency performance, transition frequency \(f_T\) and maximum oscillation frequency \(f_{max}\) were used. Their definitions are unity current gain and unity power gain respectively \[32\]. Their interpretation for circuit designs is however indicative because the requirement on margin towards \(f_T/f_{max}\) varies from design to design. Logically, for small signal amplification, \(f_{max}\) is the important upper limit to consider, and although many stages have been used to reach satisfactory gain, amplification at 67% and 75% of \(f_{max}\) has been demonstrated respectively \[33, 34\]. At high drain bias, the intrinsic bias dependent parameters, Fig. 2.2, show lower peak transconductance \((g_m)\) and higher gate-source capacitance \((C_{GS})\), which affect both \(f_T\) and \(f_{max}\) negatively, Fig. 2.2. Breakdown voltage for this process is \(\simeq 10 V\), Fig. 2.3. Therefore, for optimum output power, drain bias should be around 4V. However, since a high drain voltage affects the available gain negatively, the quiescent bias point was chosen to 3.3 V for the designs in [A] and [B] in favor of gain and at the expense of output power.
Figure 2.2: Bias dependency of the intrinsic model parameters for a $2 \times 50 \, \mu m$ transistor at $V_d = 4.4 \, V$ (--), $3.2 \, V$ (---), $2.0 \, V$ (----) and $0.8 \, V$ (-----).

Figure 2.3: Modeled IV characteristics of the pHEMT.

## 2.2 Transistor operating classes

PA operating classes are idealized, but it is still useful to define the operating conditions. In this context, the transistor model is based on a piecewise linear voltage controlled current source (VCCS) with output current levels limited between zero and $I_{\text{max}}$. It is constrained by the knee voltage $V_{\text{knee}}$ and breakdown voltage $V_{\text{max}}$, being the sole model element. The objective is to utilize the maximum current and voltage swings from the transistor to deliver maximum output power.

The idealized dynamic and time domain waveforms of the $240^\circ$ degrees conduction angle ($\alpha$) class AB amplifier, using the transistor’s maximum voltage and current constraints, Fig. 2.4. Under these conditions, the output power is expressed in (2.1), dc power consumption in (2.2) and the maximum ef-
2.2. TRANSISTOR OPERATING CLASSES

Efficiency in (2.3). At 240° degrees conduction angle, the output power of a class AB is approximately 6% higher than class A (360° degrees) and class B (180° degrees), and stems from the higher fundamental current amplitude. This also contributes to a slightly lower optimum load resistance (2.4) than for class A and class B, also referred to as Cripps’s load [35]. The clipped current waveform contains plentiful harmonics which requires specific short circuited load conditions to simultaneously acquire a pure sinusoidal voltage waveform at the transistor output. By Fourier transform, the harmonic content of the current waveform is analyzed and plotted versus the conduction angle in Fig. 2.5, leaving some interesting information about the harmonic content. In class AB, the sign of the third order harmonic is negative, creating a compressed gain response versus output power whereas in class C it is positive, creating the opposite expanding behavior. In [B], the pre-distorter is based on this observation.

\[
P_{\text{out}} = \frac{8\pi - 3\sqrt{3}}{12\pi} I_{\text{max}} \frac{V_{\text{max}} - V_{\text{knee}}}{\sqrt{2}} \approx 0.13 I_{\text{max}} (V_{\text{max}} - V_{\text{knee}}) \quad (2.1)
\]

\[
P_{\text{dc}} = \frac{\sqrt{3} + 2\pi/3}{3\pi} I_{\text{max}} \frac{V_{\text{max}} + V_{\text{knee}}}{2} \approx 0.2 I_{\text{max}} (V_{\text{max}} + V_{\text{knee}}) \quad (2.2)
\]

\[
n = \frac{P_{\text{out}}}{P_{\text{dc}}} \quad (V_{\text{knee}}=0) \quad 66\% \quad (2.3)
\]

\[
R_{\text{opt}} = \frac{V_{\text{max}} - V_{\text{knee}}}{I_{1}} \quad (2.4)
\]

(a) Dynamic plot in Class AB operation. (b) Time domain plot of a Class AB operation.

Figure 2.4: Ideal class AB waveforms.
In contrast to the piecewise linear single element VCCS transistor model, the actual transistor behaves in quite a different way. Firstly, the non-linear transconductance has a smooth Gaussian shape versus control voltage. In Section 2.1, the extracted shape of the intrinsic transconductance is plotted versus gate-source voltage and peaks for 3.3 V drain-source ($V_{ds}$) operation around $V_{gs} = -0.35 V$, which in quiescent current bias corresponds to around 0.4 $I_{max}$ or 0.35 mA/µm. The PAs in [A] and [B] have been designed to operate at 0.25 mA/µm, providing an important improvement in efficiency at the expense of gain. Second, the parasitic capacitance at the drain node ($C_{DS}$) and feedback capacitance between gate-drain ($C_{GD}$) make waveforming and harmonic tuning more difficult at high frequencies. Fortunately, the output capacitance ($C_{DS}$) contributes to an already low output impedance at the harmonic frequencies, resembling class AB operation without specific design targets for the harmonic impedances in the output matching network. The capacitor $C_{DS}$ loads the current source, changing the optimum extrinsic waveforms to lower resistance and higher susceptance. The extrinsic load-lines from the all four stages of the PA in [B] are plotted in Fig. 2.6. Even at compression, the extrinsic dynamic load-lines are fairly straight, indicating the higher harmonic currents to be short-circuited through $C_{DS}$. The intrinsic voltage and current of the current source $I_{ds}$ would be more interesting to observe because of the current source’s constraints. Unfortunately they are inaccessible in the pre-configured Chalmers model in Keysight’s ADS electronic simulation software and therefore less focus on the fundamental observations could be made.
2.3 mm-wave power amplifiers

Today’s state-of-the-art solid state E-band GaAs PAs approach 1 W of saturated output power ($P_{SAT}$)\cite{36}. This has been achieved by parallel combination of multiple small sized (2-finger, gate width < $\lambda/10$) transistors operated in class A/AB. Class A/AB provides the necessary gain and linearity, and in addition, the small sized (gate width and number of fingers) transistors have superior high frequency performance than large transistors. Effects such as higher effective source inductance and higher temperature for multiple finger transistors, and larger distributed effects of large gate width transistors impair their high frequency performance\cite{29,37}. Mm-wave PA designs need multiple gain stages to provide sufficient gain which impact efficiency and linearity negatively. The size of the output stage and the scaling factors of the driver stages are prominent in the trade-off between efficiency and linearity. On average the scaling factor is 1.5-2 times for the PAs in Table\ref{tab:table2.1}. In this context, the highest reported PAE for an E-band PA is 27%\cite{20} at a $P_{SAT}$ of 26 dBm. This has been possible because of the process’s high $f_T/f_{max} = 370/650$ GHz. The high gain per stage relaxes the number of driver stages and the driver scaling factor, and therefore also the overall dc power consumption\cite{38}. The downside of such technology for PAs is the low breakdown voltage, caused by aggressive scaling, which limits the voltage swing and therefore the output power. SiGe has demonstrated 27 dBm output power, although the lossy combining network deteriorate the PAE\cite{39}. GaAs technology has for many years been the dominant technology because of its good high frequency properties and stable
manufacturing, but emerging wide bandgap technologies such as GaN, offering higher power density than GaAs, are in development. $P_{SAT}$ beyond 1 W at E-band and W-band frequencies has been demonstrated already [21, 40]. The state-of-the-art V-band, E-band and W-band PAs in GaAs, GaN, SiGe and InP are compared for $P_{SAT}$, PAE, gain, power density, scaling and technology in Table 2.1. The scaling factor is the ratio between the normalized output stage (12) and the driver stages. Power density refers to the output power divided by the total gate/emitter width of all chip transistors.

Table 2.1: Comparison of state-of-the-art V-band, E-band and W-band PAs.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Freq. (GHz)</th>
<th>$P_{SAT}$ (dBm)</th>
<th>PAE (%)</th>
<th>Gain (dB)</th>
<th>Power density (W/mm)</th>
<th>Scaling</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>[36]</td>
<td>71-76</td>
<td>29</td>
<td>14</td>
<td>17</td>
<td>0.33</td>
<td>4:8:12</td>
<td>GaAs pHEMT</td>
</tr>
<tr>
<td>[21]</td>
<td>71-76</td>
<td>35</td>
<td>14</td>
<td>10</td>
<td>1.80</td>
<td>3:6:12</td>
<td>GaN HEMT</td>
</tr>
<tr>
<td>[40]</td>
<td>92-96</td>
<td>33</td>
<td>18</td>
<td>18</td>
<td>1.10</td>
<td>3:6:12</td>
<td>GaN HEMT</td>
</tr>
<tr>
<td>[41]</td>
<td>59-64</td>
<td>30</td>
<td>21</td>
<td>15</td>
<td>0.22</td>
<td>na</td>
<td>InP HEMT</td>
</tr>
<tr>
<td>[39]</td>
<td>68-88</td>
<td>27</td>
<td>12</td>
<td>19</td>
<td>0.05</td>
<td>3:6:12</td>
<td>SiGe HBT</td>
</tr>
<tr>
<td>[B]</td>
<td>71-76</td>
<td>27</td>
<td>18</td>
<td>27</td>
<td>0.19</td>
<td>3:4:8:12</td>
<td>GaAs pHEMT</td>
</tr>
<tr>
<td>[A]</td>
<td>55-64</td>
<td>25</td>
<td>15</td>
<td>22</td>
<td>0.18</td>
<td>3:4:12:12</td>
<td>GaAs pHEMT</td>
</tr>
</tbody>
</table>

Semiconductor technologies for 60+ GHz mm-wave PAs and mm-wave circuit implementation are limited to a handful of available technologies: InP DHBT, InP HEMT, GaN HEMT, GaAs pHEMT and SiGe HBT. Semiconductor material properties and manufacturing determine power density (mW/mm), $f_T/f_{max}$, breakdown voltage and integration capabilities. Wide bandgap materials such as GaN, can withstand high critical electrical field (V/cm) and at the same time offer high mobility ($cm^2/V s$), which by the commercially available processes in Table 2.2 is evinced. To date, GaAs is the most used technology in mm-wave front-end transmitters and receivers because of its stable and good high frequency performance. It is the main reason why this technology was chosen for the designs.

$^1$The width includes all the chip transistors.
Table 2.2: Summary of technology parameters for commercially available mm-wave processes.

<table>
<thead>
<tr>
<th>Technology/Parameter</th>
<th>GaAs pHEMT [42]</th>
<th>GaN HEMT [43]</th>
<th>SiGe HBT [44]</th>
<th>InP DHBT [38]</th>
<th>InP HEMT [43]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breakdown [V]</td>
<td>10</td>
<td>25</td>
<td>1.7</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$f_T/f_{max}$ [GHz]</td>
<td>135/200</td>
<td>170/250</td>
<td>250/340</td>
<td>370/650</td>
<td>300/450</td>
</tr>
<tr>
<td>Power density [mW/mm]</td>
<td>700</td>
<td>1000</td>
<td>100$^2$</td>
<td>1600$^2$</td>
<td>100</td>
</tr>
<tr>
<td>Node</td>
<td>100 nm</td>
<td>60 nm</td>
<td>130 nm</td>
<td>250 nm</td>
<td>70 nm</td>
</tr>
<tr>
<td>Wafer size</td>
<td>6 in</td>
<td>3 in</td>
<td>8 in</td>
<td>4 in</td>
<td>3 in</td>
</tr>
</tbody>
</table>

2.3.1 Power combining

Transistors with short gate widths were mentioned to have better high frequency properties than wide gate widths in Section 2.3. As a consequence, parallel transistor combining solutions are implemented for increasing the total output gate width for improving the power handling and linearity at the lowest cost of high frequency performance.

Many designs adopt 90° hybrid combining networks, principally at the input and output networks of the amplifier because of their advantage of isolating the two branches and providing low input and output return loss (RL). Since low RL and power match are not necessarily the same, 90° hybrid couplers bring ideal power match for the transistors while providing ideal system match at the output. Their loss is also reasonably low which makes this solution attractive as an input/output combiner. This combining method has been adopted in [36].

Power combiners provide isolation between branches and let $N Z_0$ branches of the same $Z_0$ impedance to be combined. The Wilkinson power combiner is used in [20].

The simplicity of the tee combiner, makes it the least lossy power combiner and is used in [A], [B] and in [21,39,40]. It does not provide isolation between the branches in contrast to hybrids and power combiners and must therefore consider the excitation on the other port. To obtain the transistor’s optimum load impedance, the combiner network must be optimized under common-mode excitation. The common-mode output impedance ($S_{c22}$), which should equal the optimum impedance, is either calculated from the three port network in Fig. 2.7 (left) using (2.5) or directly optimized from the two port network under equivalent common-mode excitation, Fig. 2.7 (right) [45].

$^2$Normalized to the emitter width, including the emitter length in the calculation.
\[ \begin{align*}
S_{c11} &= S_{11} \\
S_{c21} &= \frac{S_{21} + S_{31}}{\sqrt{2}} \\
S_{c22} &= \frac{S_{22} + S_{33} + S_{32} + S_{23}}{2}
\end{align*} \]

(2.5)

Figure 2.7: Equality between linear 3-port to common-mode translation.

Parallel combining multiple devices have been the de facto implementation at mm-wave frequencies to maintain good high frequency performance while increasing output power. The power increase stems from the increase in current, as illustrated in Fig. 2.8a. Power combination by voltage is possible when the transistor is configured to amplify in-phase, which occurs when the signal input is on the source terminal and the output is on the drain terminal, i.e. the configuration of a stacked HEMT and common-gate, as illustrated in Fig. 2.8b. Stacking transistors was first mentioned in 1964 [46] for a high-voltage video amplifier, but its characteristics did not generate interest until about ten years ago, when extreme down-scaled processes appeared, increasing $f_T/f_{\text{max}}$ of the transistor, but suffering from low breakdown voltage, low output impedance and power density. By series connecting both the ac and dc signals, theoretically, both higher output impedance and higher operating voltage can be achieved, which may simplify large, complicated combining networks and facilitate biasing.

Figure 2.8: Current and voltage combining methods.
2.3.2 Budgeting multiple transistors PAs

Mm-wave PAs may need multiple gain stages in contrast to PAs in the low microwave frequency region, where single transistor designs may be sufficient for providing gain, efficiency and output power. In this context, budgeting the amplifier with respect to number of stages and gate width is important. In Fig. 2.9, the 4-stage, 4-parallel PA from [B] is shown. The size of the output stage determines, unless limited by the driver stages, the maximum output power. Scaling the transistor parameters per $\mu$m and calculating for class A operation and assuming $V_{knee} = 0$ for simplifying calculations, the output power of this particular design is estimated to: 

$$\text{gatewidth (}\mu\text{m}) \cdot I_{\text{max}} (\text{mA}/\mu\text{m}) \cdot V_{\text{max}} (\text{V}) / 8 = 1200 \mu\text{m} \cdot 0.75 \text{mA}/\mu\text{m} \cdot 6 \text{V} / 8 = 675 \text{mW},$$

using Cripps’s load [35], which is close to the achieved output power of 500 mW.

![Block diagram and transistor sizing budget.](image1)

![Chip photograph.](image2)

**Figure 2.9:** Block diagram and photograph of the PA in [B]

The balance between linearity and efficiency is determined by the sizing of the driver stages. At a minimum, for not saturating the driver stages before the output stage, the delivered power needs to be $P_{\text{OUT}} (\text{dBm}) - \text{Gain (dB)}$. However, taking gain saturation into account, the output power needs to be higher. The budget for [B] was calculated from output third order intercept point (OIP3) and by simulating the dynamic load-lines to visually confirm the proper scaling, shown in Table 2.3 and Fig. 2.6 respectively.

It is in this context when budgeting for multiple transistors, that voltage combining (stacked HEMTs) has a positive impact on the efficiency. Under the assumption that an additional 3 dB gain and, 3 dB output power can be obtained from one stacked HEMT. Then, by adding one stacked HEMT to the output transistor and by removing one branch and the input driver stage of the amplifier in [B] similar output power and gain can be obtained, while increasing PAE and reducing chip area. In Fig. 2.10, the equivalent configured PAs, based on common-source (C-S) transistors [B] and the stacked HEMT [A] for equivalent performance are shown.
Table 2.3: Budget of the PA in [B] at 73.5 GHz.

<table>
<thead>
<tr>
<th>Stage/Parameter</th>
<th>Comb</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>Comb</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain (dB)</td>
<td>-0.5</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>-1</td>
<td>25.5</td>
</tr>
<tr>
<td>OIP3 (dBm)</td>
<td>Inf</td>
<td>31</td>
<td>32</td>
<td>35</td>
<td>37</td>
<td>Inf</td>
<td>33.4</td>
</tr>
<tr>
<td>δOIP3/δN (%)</td>
<td>0</td>
<td>3</td>
<td>14</td>
<td>28</td>
<td>55</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>PDC (mW)</td>
<td>0</td>
<td>250</td>
<td>330</td>
<td>660</td>
<td>990</td>
<td>0</td>
<td>2230</td>
</tr>
<tr>
<td>LW (mm)</td>
<td>NA</td>
<td>0.3</td>
<td>0.4</td>
<td>0.8</td>
<td>1.2</td>
<td>NA</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Figure 2.10: Equivalent translation from C-S PA to using stacked HEMT at the output. Block diagram and layout of the stacked HEMT PA in [A]. The chip size measures 2 × 1.6 mm².

Table 2.4: Budget of the PA in [A] at 60 GHz.

<table>
<thead>
<tr>
<th>Stage/Parameter</th>
<th>Comb</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>Comb</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain (dB)</td>
<td>-0.5</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>-0.5</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>OIP3 (dBm)</td>
<td>Inf</td>
<td>26</td>
<td>27</td>
<td>34</td>
<td>Inf</td>
<td>33.4</td>
<td></td>
</tr>
<tr>
<td>δOIP3/δN (%)</td>
<td>0</td>
<td>8</td>
<td>31</td>
<td>61</td>
<td>0</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>PDC (mW)</td>
<td>0</td>
<td>165</td>
<td>200</td>
<td>900</td>
<td>0</td>
<td>1265</td>
<td></td>
</tr>
<tr>
<td>LW (mm)</td>
<td>NA</td>
<td>0.2</td>
<td>0.25</td>
<td>1.2</td>
<td>NA</td>
<td>1.65</td>
<td></td>
</tr>
</tbody>
</table>
2.3.3 Stacked HEMT layout and verification

Voltage combining at mm-wave frequencies is not unproblematic. Several conditions must be met for ideally increasing the output power \( n \) times with \( n \) stacked FETs \((2.6)\). 1) The voltage phase rotation across the \( n^{th} \) stacked FET must be zero. 2) The \( n^{th} \) HEMT needs to be optimally loaded. Since the current through all transistors is equal, assuming \( I_G = 0 \), the voltage magnitude across the \( n^{th} \) device needs to be the same as the first FET. When these conditions are met, the result is \( n \) times higher voltage swing for the same current \((2.7)\). For a two stacked FET, the voltages and currents along the FETs are illustrated in Fig. 2.8b. Analysis and theories on how to configure the stacked FETs for gaining output power has appeared in many publications.

\[
|V_{DS,n}| = |V_{DS,1}|
\]

\[
\angle V_{DS,n} = 0, \forall n > 1
\]

\[
I_{D,n} = I_{D,1}
\]

\[
|V_{D,n}| = n \times |V_{D1}|
\]

Perhaps the most influential paper on how to configure stacked FETs \([47]\), presents a two step procedure. 1.) tuning the capacitor at the stacked transistors’ gates (\( C_{A1} \) in Fig. 2.8b) to obtain the optimum resistance or conductance of \( Z_{IN2} \) and 2.) introducing a reactive element, either between the transistors or in parallel with the stacked FETs, to tune the reactive part such that the optimum impedance for output power is obtained. In a demonstrator design comprising three stacked FETs, the output power that they achieve is 5.7 dB higher than that of a single C-S transistor at 41 GHz. Their design frequency is far below the \( f_T/f_{max} \) of the process which allows them to use a simplified small signal transistor model with no extrinsic elements in the analysis, achieving satisfactory agreement between simulations, theory and measurements.

In \([48]\), the optimum number of stacked FETs is considered for high frequency operation. By taking into account the loss of current along the stack towards the gate(s), their hypothesis is that there exists a maximum number of stacked FETs that increase output power. The loss of gate current is important, because a stacked HEMT has unity or less current gain. An expression based on the ratio between the operating frequency and the process’s \( f_T \) to indicate the optimum number of stacked FETs is derived. In \([49, 50]\), their hypothesis has been tested on a set of broadband PA designs using Fraunhofer IAF’s 50nm metamorphic HEMT (mHEMT) process \([51]\), and delivers 17.5 dBm (PAE 6%) and 22.5 dBm (PAE 8%) output power respectively at W-band frequencies.

In \([52]\), by using a transistor model without \( C_{GD} \) and having no extrinsic elements, this work shows that there exist an optimum configured stacked HEMT for output power by introducing a gate capacitor and a bypass capacitor between source and drain on the second HEMT. The values are easy to derive since they are related to the device parameters. At 10 GHz, their theory is verified and delivers 1.3 W output power at 3 dB gain saturation, which is around 1.5 dB higher than the simulated single HEMT. [A] is based on this approach, although a more complicated transistor model is used in the design.
In [53], the analysis in [52] is complemented by introducing the gate-drain feedback capacitor to the transistor model. With the same elements and configuration as in [52], the analysis of the configuration shows that the proposed drain-source bypass capacitor becomes gyrated, creating the effect of a negative capacitance seen at its source input. The negative capacitance compensates for the previous transistor’s drain susceptance, creating both a conjugate match and minimized voltage phase rotation along the stack, but more importantly, the configuration is frequency independent. However, as is commented in [52], the small value of the capacitor is difficult to implement and as shown and commented in [47], the proper phase matching is achieved but requires higher inductive load at the top device, making the output matching network challenging to realize.

In [47–50,52,53], all assume an ideal lumped capacitor to ground, ignoring any distributed effects from the capacitor or via hole inductance to ground. The distributed effects of the capacitor or inductance to ground can generate a negative input resistance to the stacked HEMT ($Z_{IN2}$), and possibly cause stability problems (Fig. 2.11a). The implementation solution in [A] is to improve stability when the distributed effect from the gate capacitor is accounted for. An innovative transistor layout in microstrip is proposed in Fig. 2.11b to increase the self-resonance frequency by connecting the gates on both side and distribute the capacitance $C_{A1}$ on four vias, as shown in Fig. 2.11b. The design contributions in [47, 49, 50, 53] were implemented with co-planar waveguides, having lower ground inductance than a microstrip via holes and therefore higher self-resonance frequency.

![Figure 2.11: Simulated input resistance towards the stacked HEMT and the proposed layout having increased stability.](image)

(a) Simulated input resistance of $Z_{IN2}$ versus frequency for n number of vias on the gate terminal of the 2nd HEMT. All are simulated with $C_{A1} = 300\, \text{fF}$.

(b) Layout of the proposed stacked HEMT cell, showing the $6 \times 50\, \mu\text{m}$ C-S HEMT to the left and $4 \times 75\, \mu\text{m}$ 2nd HEMT with four gate vias to the right.

The fabricated output transistors $6 \times 50\, \mu\text{m}$ (used in [B]), and the pre-configured stacked HEMT $6 \times 50\, \mu\text{m}$ and $4 \times 75\, \mu\text{m}$ with $100\, \text{fF}$ gate capacitance $C_{A1}$ (used in [A]) were characterized in an active load pull measurement set-up developed at Chalmers [54] for two purposes. 1.) To measure the maximum output power at ideal load conditions and 2.) to use the data to verify the
transistor model. The two devices were measured at a fundamental frequency of 10 GHz while sweeping the load impedance actively at the device using the load pull setup, the results for the two devices are plotted in Fig. 2.12. The gain and PAE for the two transistors are plotted in Fig. 2.13, where the pre-configured stacked HEMT achieves 3 dB higher output power at a load impedance close to twice that of the C-S transistor. The higher output impedance of the stacked HEMT is attracting for further parallel combining, since the impedance transformation to a 50 Ω is lower and therefore more broadband. The transistor model was verified by comparing the measured waveforms for impedances between 5 Ω and 500 Ω, such that the maximum voltage and current were captured, shown in Fig. 2.14. The two waveforms, the simulated and the measured, are compared on top of each other, indicating that the model is a good fit.

![Figure 2.12](image1)

(a) Stacked HEMT.

(b) C-S HEMT.

Figure 2.12: The contour lines of the measured output power $P_{\text{OUT}}$ of the stacked and C-S HEMTs versus $Z_L$ (red continuous lines) and the sampled $Z_L$ points (blue dots) at 4 V and 0.25 mA/µm. In the measurement, the input power was 2 dBm. In both plots, the normalized impedance is $Z_0 = 50 \, \Omega$.

![Figure 2.13](image2)

Figure 2.13: Measured power gain (continuous blue) and PAE (dashed black) versus output power at optimum load impedance of the stacked HEMT (squares) and the C-S HEMT (circles) at 10 GHz.

![Figure 2.14](image3)

Figure 2.14: Measured and simulated waveforms while varying the load impedance of the stacked HEMT between 5 and 500 Ω. Blue traces are the simulated waveforms and red traces are measurements.
2.3.4 Stacked HEMT PA design

The power combining methods of the PA in [A] are a blend of voltage and current combining, comprised of two stacked HEMTs and one tee-combiner. The higher optimum output impedance for power of the stacked HEMT in contrast to the C-S HEMT, $17 + j22 \Omega$ and $12 + j10 \Omega$, respectively, alleviate the design of the output combiner network in terms of lower impedance transformation and broadband matching. In Fig. 2.15, the complete circuit schematic and the circuit realization in Fig. 2.10 are shown. The output combiner network is realized with two sections of inductors and capacitors to increase the stacked HEMTs’ common-mode source impedance of $8.5 - j11 \Omega$ to $30 \Omega$ ($L_4$, $L_7$ and $C_6$) and, with a quarter wavelength transformer to match to $50 \Omega$ ($T_L_2$). The small iterations of impedance transformation, ensure low-Q matching and enlarge the bandwidth.

![Figure 2.15: Circuit schematic of the stacked HEMT PA.](image)

Ideally, the output impedance of the stacked HEMT should be twice that of the C-S HEMT, indicating that the configuration is not optimally power matched. In Fig. 2.16, the time domain and voltage amplitudes of the C-S HEMT, second HEMT and stacked HEMT are shown. Although the voltage swings across the C-S HEMT and the second HEMT are similar in magnitude, their phases are not perfectly aligned (left plot), lowering the combined output magnitude Fig. 2.16 (right).

In Fig. 2.17a, the measured frequency response is shown. The gain and return loss measure flat response of 22 dB and low return loss better than 10 dB within the 60 GHz ISM band. In Fig. 2.17b, the gain, output power and PAE are plotted. A maximum PAE of 15% and 25 dBm $P_{SAT}$ are obtained. Because of the small chip area, the circuit’s output power density per chip area is higher than the average GaAs PA.
Figure 2.16: Simulated time domain intrinsic voltage waveforms of $v_{D1}$ and $v_{D2}$ (left) and current waveforms of $i_{D1}$ and $i_{D2}$ (middle). (Right) The simulated intrinsic drain-source voltage swings $v_{DS1}$ of the C-S HEMT (dotted red line), $v_{DS2}$ of the 2\textsuperscript{nd} HEMT (dashed blue line) and $v_{OUT}$, the output voltage swing from the stacked HEMT (black line).

Figure 2.17: Measured (continuous) and simulated (dashed) responses of the stacked HEMT PA.

2.4 Linearization

PAs are often considered to be one of the more important components in a transmitter, where output power, linearity and efficiency are highlighted Figure-Of-Merits (FOMs). However, these parameters diverge from each other, making the combination efficient and linear PAs difficult to achieve. To make use of efficient PAs, or enhance the output power from an ordinary PA by operating it closer to saturation, linearization is necessary. At the expense of increased system complexity, either in the digital or analog domain, and by hardware or software, the reward is higher RF performance. This ideal response is coveted by pre- and postdistorters.
2.4.1 Implementation overview

Linearization of PAs has been a topic of interest ever since 1923, when Harold Black was intrigued to solve distortion in transatlantic communications. By accepting the inherent non-linear characteristics of amplifiers (electron tubes at the time), he presented the feed-forward linearization technique [55]. The method is advantageous and ingenious in the sense that "the error", the difference between the non-distorted input signal and the distorted output signal is inversely added to the output signal, linearizing the response of the PA [56]. Unfortunately, the invention did not materialize at the time because of the difficulties of maintaining the amplifier to work linearly over time. Another invention from H. Black came in the interim, the feedback amplifier [57], which under correct feedback conditions was shown to cancel the distortion products [58]. However, the feedback amplifier holds plenty of features appreciated by circuit designers, such as increased stability, limit of gain, ease of matching etc. The feed-forward technique was demonstrated to work many years later [59] and has found use in the low microwave frequency domain and, for example, demonstrated 65 dBC carrier to third order intermodulation (CI3) at 7% efficiency [60]. The main drawback is the power overhead in the auxiliary branch that reduces efficiency [61]. While many alternative linearization techniques have been presented over time (many are summarized in [56, 58]), the scope of this section is on analog pre-distorters.

Pre-distortion, as the name implies, distorts the input signal such that a linear response is achieved after the PA or the complete transmitter as shown in Fig. 2.18. By inversely mimicking the non-linear response of the PA, a linear response up to \( v_{IN, max} \) is achieved and, as such, the higher order polynomials cancel (2.10). The non-linear response of the pre-distorter and PA are shown in Fig. 2.18a and Fig. 2.18b respectively, whose non-linear complex response gives rise to amplitude and phase skew. Because pre-distortion is a cascaded implementation, it can be implemented at the very first component in a transmitter chain, in the modem, digitally. This has several advantages: no additional hardware is needed and, since the polynomial parameters need to be updated depending on signal characteristics, system temperature drift, output power and frequency response. This solution is very flexible and does not suffer from constraints in the same manner that may occur in an analog implementation. During the last two decades, DPD has been widely adopted on narrow-band communication signals and has become the de facto solution to overcome non-linearities because of its low operational cost (opex) and capital cost (capex) [61]. For applications using wideband signals such as 5G, broadband backhaul (E-band) and television broadcasts, DPD requires sample rates multiple times the signal bandwidth. In addition complex calculations is required, that even the more efficient CMOS node (28 nm) matches the power consumption of the PA in [A] making DPD unsuitable as the linearizing implementation [62]. On the contrary, APD of PAs can be implemented with negligible power consumption to enhance the linearity over large bandwidths [22]. Recent publications on the topic, the work in [23] presents the simulations and improvements of using a GaAs APD for linearizing traveling wave tube amplifiers at the E-band and W-band. In [63], a varactor is used to linearize the non-linear gate-source capacitance and in [35]64, a cold FET
and anti-parallel diodes are used to form gain expansion respectively.

\[ v_{APD}(v_{IN}) = a_1 v_{IN} + a_3 v_{IN}^3 \]  
\[ v_{PA}(v_{APD}) = g_1 v_{APD} + g_3 v_{APD}^3 + \cdots \]  
\[ v_{PA}(v_{IN}) = a_1 g_1 v_{IN} + (g_1 a_3 + g_3 a_1^3) v_{IN}^3 + \cdots \]

2.4.2 Circuit simulation environment

In order to develop and simulate the pre-distorter, it was necessary to suppress the heavy computational efforts of the PA itself, without the need of co-simulating the PA. For this reason, a simulation platform was developed in Keysight’s Advanced Design System (ADS) circuit simulator software that simulates only the pre-distorter under LS excitation, and calculates the combined non-linear response based on either the measured or simulated non-linear PA response. The simulation environment was developed for continuous wave and for two-tone excitation, simulating amplitude-to-amplitude distortion (AM-AM), amplitude-to-phase distortion (AM-PM) and intermodulation distortion. The simulation time was shortened from 84 seconds to 11 seconds, allowing the design phase to be more efficient and, more importantly, each components’ waves are separated such that the individual contribution and combined response can be displayed simultaneously. From a design perspective, this is valuable informative information. In the following paragraphs, the simulation environment, calculus and comparisons between the calculated and the combined simulated responses are presented.

1. Both the pre-distorter and PA need to have the same Harmonic Balance (HB) setup and sweeps, i.e. the same center frequency and the same tone separation frequency such that the harmonic components are the same, shown in Fig. 2.19. In particular, the power sweeps are of interest to improve the AM-AM and AM-PM responses and intermodulation distortion at high output power levels. The frequency tone separation sweep is also of interest to determine the maximum signal bandwidth for the circuit combination but, is not shown here.

2. Normally when cascading 2-port linear blocks, ABCD matrices are used. However, they cannot be applied to a non-linear circuit where the LS com-
pressed response depends on the load condition. Since the circuits in [B] have been designed, simulated and measured in a 50Ω environment for large and small signals characterization, Scattering Parameters (SP) are a suitable basis to work from [32]. SP also have the flexibility to choose system impedance, also complex, which is useful if the LS response needs to be characterized under other load impedance conditions. If the circuits are assumed to be non-reflective ports, have reverse isolation and designed in the same system impedance, the LS through response ($L_{S21}$) can be cascaded when $b_{2APD}$ equals $a_{1PA}$ (see Fig. 2.20) (2.11). These observations are useful and reasonable simplifications to make in this case for estimating the combined performance. While SP are useful in the characterization of circuits and devices, they cannot directly be cascaded. On the other hand, transfer parameters allow direct cascading of blocks, including reflections and reverse transmission and are based on the same normalized power waves used for SP (2.12) [32]. While more exact combined non-linear response can be achieved if load-pull data are available to reflect the response under specific load impedances, the purpose of this simulation environment is not to reach perfection, but rather to estimate the performance during the design phase.
2.4. LINEARIZATION

\[
\begin{bmatrix}
    b_1^{APD} \\
    a_1^{APD} \\
    b_2^{PA}
\end{bmatrix}
\begin{bmatrix}
    T_{11}^{APD} & T_{12}^{APD} \\
    T_{21}^{APD} & T_{22}^{APD}
\end{bmatrix}
\begin{bmatrix}
    T_1^{PA} \\
    T_2^{PA}
\end{bmatrix}
\begin{bmatrix}
    a_2^{PA} \\
    b_2^{PA}
\end{bmatrix}
\Rightarrow
\begin{bmatrix}
    S_{21}^{APD} \\
    S_{22}^{APD}
\end{bmatrix}
\]

(2.11)

3. Read the PA’s already simulated or measured input and output voltages and currents from a saved dataset for the frequency components of interest. In this case the fundamental component and the third (IM3) and fifth order intermodulation (IM5) products from the 2-tone sweep were of interest. In the previous paragraph, it was mentioned that \( b_2^{APD} \) needs to equal \( a_1^{PA} \) for cascading the LS responses, which puts the reference plane between the pre-distorter and the PA, at \( P_{APD} \) in Fig. 2.20. Since the PA response already depends on the input power at this reference plane, the LS SP are directly calculated from the dataset (2.13) and (2.14). In addition to the fundamental gain \( L_{S21}^{PA}(P_{APD}) \), the gain of the IM3 and IM5 products, \( L_{S21,IM3}^{PA}(P_{APD}) \) and \( L_{S21,IM5}^{PA}(P_{APD}) \), are calculated with reference to the input signal \( a_1(P_{APD},1,0) \) (2.13). The dependent variables are the power level at \( P_{APD} \) and the frequency, obtained from the mixing product \( n_f1 + m_f2 \).

\[
a_1^{PA}(P_{APD}, n_f1, m_f2) = \frac{1}{2 \sqrt{Z_0}} \left( v_{APD}(P_{APD}, n_f1, m_f2) + Z_0 \cdot i_{in2}(P_{APD}, n_f1, m_f2) \right)
\]

\[
b_1^{PA}(P_{APD}, n_f1, m_f2) = \frac{1}{2 \sqrt{Z_0}} \left( v_{APD}(P_{APD}, n_f1, m_f2) - Z_0 \cdot i_{in2}(P_{APD}, n_f1, m_f2) \right)
\]

\[
a_2^{PA}(P_{APD}, n_f1, m_f2) = \frac{1}{2 \sqrt{Z_0}} \left( v_{out}(P_{APD}, n_f1, m_f2) + Z_0 \cdot i_{out}(P_{APD}, n_f1, m_f2) \right)
\]

\[
b_2^{PA}(P_{APD}, n_f1, m_f2) = \frac{1}{2 \sqrt{Z_0}} \left( v_{out}(P_{APD}, n_f1, m_f2) - Z_0 \cdot i_{out}(P_{APD}, n_f1, m_f2) \right)
\]

(2.13)

\[
L_{S21}^{PA}(P_{APD}) = \frac{b_2^{PA}(P_{APD},1,0)}{a_1^{PA}(P_{APD},1,0)}
\]

\[
L_{S21,IM3}^{PA}(P_{APD}) = \frac{b_2^{PA}(P_{APD},2,-1)}{a_1^{PA}(P_{APD},1,0)}
\]

\[
L_{S21,IM5}^{PA}(P_{APD}) = \frac{b_2^{PA}(P_{APD},3,-2)}{a_1^{PA}(P_{APD},1,0)}
\]

(2.14)

4. During the power sweep of the pre-distorter, the response will depend on the input power at \( P_{IN} \) (2.15). For the calculations to work, the output power level must match the input power level of the PA, such that \( b_2^{APD} \) equals \( a_1^{PA} \). To accomplish this, the LS through response of the pre-distorter \( L_{S21}^{APD}(P_{IN}) \) is interpolated to match the same power level grid as the PA (2.16). A new
common grid that spans across the common low and high power levels of the
two sweeps is generated. Finally, by rotating the phase of $b_2^{APD}$ to match $a_1^{PA}$, $b_2^{APD} = a_1^{PA}$. Since we assume no reflections, the phase of $a_1^{PA}$ is zero.

$$LS_{21}^{APD}(P_{IN}) = \frac{b_2^{APD}(P_{IN},1,0)}{a_1^{APD}(P_{IN},1,0)}$$

$$LS_{21,IM3}^{APD}(P_{IN}) = \frac{b_2^{APD}(P_{IN},2,-1)}{a_1^{APD}(P_{IN},1,0)}$$

$$LS_{21,IM5}^{APD}(P_{IN}) = \frac{b_2^{APD}(P_{IN},3,-2)}{a_1^{APD}(P_{IN},1,0)}$$

(2.15)

5. The combined LS gain response is obtained by multiplying $LS_{21}^{APD}(P_{APD})$ and $LS_{21}^{PA}(P_{APD})$. The results of both the combined simulation (black) and separate simulation (red) APD+PA responses and the standalone PA (blue) are shown in Fig. 2.21. The separated and combined AM-AM and AM-PM responses, obtained from the LS gain responses, are shown in Fig. 2.22. The gain expansion and opposite phase rotation are positive effects towards a linear response. The intermodulation contributions from the APD, PA and the combined APD+PA are shown in Fig. 2.23. The output $b$-waves of the APD and APD+PA are obtained by (2.17) and (2.18), and the $b_2$ parameter from the PA is obtained from (2.13). Since no reflections are assumed in the calculations, $|b_2|^2$ equals the output power. The simulation platform provides in general good agreement with the combined simulations, but improves simulation time and gain understanding in the design of an APD circuits. There is much more interesting information to show, for example IM5, phase and amplitude mismatch in the intermodulation products, but it is beyond the scope of the thesis.

$$b_{2,IM3}^{APD} = \frac{u^{APD}}{\sqrt{Z_0} LS_{21}^{APD}} LS_{21,IM3}^{APD} \cdot LS_{21}^{PA}$$

(2.17)

$$b_{2,IM3}^{COMB} = \frac{u^{APD}}{\sqrt{Z_0} LS_{21}^{APD}} LS_{21,IM3}^{APD} \cdot LS_{21}^{PA} + a_1^{PA} \cdot LS_{21,IM3}^{PA}$$

(2.18)
2.4. LINEARIZATION

Figure 2.21: Simulated gain and output power responses using separate simulations and calculating the response (red) and combined simulation (black). The PA response is plotted in blue.

Figure 2.22: Simulated AM-AM and AM-PM responses using separate simulations and calculating the response (red) and combined simulation (black). The PA response is plotted in (blue) and the APD in (pink).

Figure 2.23: Simulated intermodulation using separate simulations and calculating the response (red) and combined simulation (black). The PA response is plotted in (blue) and the APD in (pink).
2.4.3 Class C transistor pre-distorter

The analog pre-distorter in [B] comprises two branches, one for the linear coefficient \( a_1 \) and one for the third order coefficient \( a_3 \), as shown in Fig. 2.24. The linear branch consists of a meandered transmission line having equal time delay \( \tau_1 \) in Fig. 2.24\(^{2,3}\) as the non-linear branch to ensure wide signal bandwidth operation. With no tunable elements in this branch, \( a_1 \) is therefore fixed. The design concept is to let each branch have one dominant coefficient over the other branches in order to have independent control of the coefficients. The fundamental leakage of the class C biased transistor (EA) requires the linear branch to be the dominant contributor to the overall \( a_1 \) coefficient. It is important that the ratio \( a_1/a_3 \) can be tuned, such that changes in the PA’s characteristics can be compensated for by the pre-distorter. Since \( a_1 \) is fixed, it is important that \( a_3 \) can be tuned, zeroing the combined third order coefficient \( \tau_2 \). By adjusting the gate bias of the EA and by adjusting its delay (VDL), the magnitude and phase of the complex coefficient \( a_3 \) change. A drawback with this implementation is that the non-desired higher harmonic components also change when tuning the biases. Their significance to the overall performance was sufficiently low and the spectral regrowth was measured lower with the pre-distorter than without for the same output power, as shown in Fig. 2.26\(^{2,3}\).

A critical part of this design is the isolation between the two branches, which if carefully considered, reduces undesirable feedback and increase stability. In [65], layout considerations for microstrip Wilkinson power combiners at millimeter wave frequencies were investigated. By accounting for the distributed effects of the resistor, coupling between the lines and adjusting the transmission lines lengths not considered in the original design [66], they show that a variety of solutions exist in order to obtain ideal performance (no return loss and isolation between the output ports). Our Wilkinson power combiner design, comprises orthogonal transmission lines exiting from the resistor in order to minimize the coupling between lines. Moreover, the resistor’s width-length aspect ratio (100 \( \Omega \)) is a compromise between being physically short (higher coupling between lines) and long (larger distributed effects). The limitation of using a one section Wilkinson power combiner is its relative bandwidth of 36% for 20 dB isolation, but if increasing the number of sections to two or higher, the relative bandwidth can be increased to 67% for 27 dB isolation or multiple octaves bandwidth [67,68]. In [69], they analyze by means of SP the "error function" for increasing the isolation at the expense of return loss and propose a compensation network that increases the relative bandwidth to 83% for 20 dB isolation. An alternative more wideband hybrid coupler to the Wilkinson power combiner is the Lange coupler [70], which is attractive in terms of size.
The source for generating the IM3 products originates from a transistor biased in class C and it has the inverse cubic response in contrast to a Class AB transistor [35], Fig. 2.5. This creates the wanted gain expansion and is the type of function that compensates the PA’s gain compression, shown in Fig. 2.18. Although the third order polynomial is the desired function in this design, there will be a plethora of harmonics and intermodulation products, small and large, which makes it impossible to separate and isolate each harmonic coefficient. At class C bias, the third harmonic is relatively large whereas the higher odd order harmonics are relatively small, which makes this bias point attractive. The other consideration in the pre-distorter design is the harmonic coefficients dependency on input power. When the two circuits have different power dependent parameters they will need to updated for different power levels. This effect is observed in Fig. 2.26a, where for a two-tone measurement, "sweet spots" of canceled IM3 products appear for specific settings of the coefficient when sweeping the input power.
(a) Measured CI3 versus output power per tone at 10 MHz tone separation frequency and 73.5 GHz center frequency.

(b) Measured output spectrum from a 64 QAM modulated signal with 250 MHz modulation bandwidth at 17 dBm average output power. The mask follows ETSI EN 302-217-2.

Figure 2.26: Measured results of the combined pre-distorter and PA.
Chapter 3

Mixer theory and design of linear quadrature modulators

This chapter focuses on the designs of quadrature modulators. An overview of mixer classes, the operation of the resistive mixer and mixing theory are also presented and discussed. The conception of linear, non-linear, time invariant and time variant systems processes, of which the non-linear and time variant processes give rise to mixing are presented. Circuit complexity and mixer classes, such as fundamental, sub-harmonic, balanced and quadrature are introduced. The realization of two balanced quadrature modulators (in reality a mixer, defined to mix modulated signals to a carrier), one sub-harmonic [D] and one fundamental [C], both sharing the same balanced mixer class and based on the resistive mixer [71] are presented.

Wideband communication, such as at the E-band, W-band and D-band, requires linear quadrature modulators to retain the deteriorated Signal-to-Noise Ratio (SNR). Thermal noise increases linearly with the signal bandwidth, making it an increasingly important parameter to consider in the transmitter design and for minimizing EVM. Simultaneously, high order modulation formats have tougher requirements on spectral purity. Both noise and intermodulation requirements are controlled by regional regulatory organs for radio deployments, such as ETSI and FCC, making them compulsory to comply with and, non-negotiable in order to preclude interference with adjacent channels. The quadrature modulators in [C] and [D] address this problem by optimizing their linearity, to allow high input signal power and thereby increase the SNR. Other important parameters for modulators are the carrier (LO) to RF suppression and orthogonality (quadrature balance). Particular emphasis on the phase and amplitude distribution networks is taken, that to a large extent determine the performance of these two parameters.
3.1 Mixing fundamentals

All circuit elements show non-linear characteristics to some degree, even for components such as resistors or capacitors at extreme excitation. With this perspective, to generate new frequency components from their non-linear behavior sounds trivial. The key point is, that specific frequency mixing or harmonic products are wanted. From a circuit designer’s perspective, this is a complicated task that requires a solid basis about non-linear components’ properties, circuit designing skills and circuit topologies. Although non-linear elements for generating new frequency components were mentioned, the term “linear mixers” may sound contradictory, but it can in fact be true from a theoretical standpoint. To clarify the terms, in particular with respect to non-linearity and time variance, this will be explained.

A linear system, per definition, satisfies the principle of superposition [72], that is, the output is a replica of the input signal, only scaled in amplitude. The system in (3.1), a polynomial function to the $n^{th}$ degree with coefficients $a_N$ is therefore linear when $n = 1$ and arbitrary scaled by $a_1$. When the input signal (3.2) passes the linear system, the output response is $v_{out} = a_1 b_1 \cos(\omega_1 t) + a_1 b_2 \cos(\omega_2 t)$. If $a_1$ in such system is a constant, the system is considered to be Linear Time Invariant (LTI) but if $a_1(t)$ depends on time, the system is considered to be Linear Time Variant (LTV). For example, if $a_1(t) = a_1 \cos(\omega_{LO} t)$ the output becomes $v_{out} = a_1 \cos(\omega_{LO} t) b_1 \cos(\omega_1 t) + a_1 \cos(\omega_{LO} t) b_2 \cos(\omega_2 t)$, a perfect linear frequency translation. This discloses the term ”linear mixers”.

$$v_{out}(v_{in}) = \sum_{n=0}^{\infty} a_n v_{in}^n$$

$$v_{in} = b_1 \cos(\omega_1 t) + b_2 \cos(\omega_2 t)$$

If a simple polynomial of the second degree ($n = 2$) in (3.1) is considered. When applying the same input signal (3.2) to the system, plenty of new frequency components are generated by the non-linearity (3.3). This system does not fulfill the principle of superposition, and therefore cannot be defined as a linear system. In this context, the polynomial degree $n$ determines the maximum mixing order $n = k + l$, where $k$ and $l$ are real integers . . . , −2, −1, 0, 1, 2 . . . and therefore determines the frequency harmonic content based on $\omega_1$ and $\omega_2$, as $\omega_{k,l} = \pm k \omega_1 \pm l \omega_2$. A second order polynomial gives rise to second order intermodulation (IM2) products, a third order polynomial to IM3 products and so on, thereof the terms second and third order intermodulation.

$$v_{out} = a_0 + \frac{b_1^2 + b_2^2}{2} + a_1 b_1 \cos(\omega_1 t) + a_1 b_2 \cos(\omega_2 t) + \frac{a_2 b_1^2 \cos(2 \omega_1 t) + a_2 b_2^2 \cos(2 \omega_2 t)}{2} + a_2 b_1 b_2 \left( \cos \left( (\omega_1 + \omega_2) t \right) + \cos \left( (\omega_1 - \omega_2) t \right) \right) + \ldots$$

(3.3)
3.2 The resistive FET mixer

The resistive FET mixer, presented by S. Maas in 1987 [71], is a widely used mixer topology due to its high linearity, simple implementation and zero direct current (dc) consumption. The FET is operated in the linear region of the FET, with quiescent 0 V drain bias and gate bias close to the threshold voltage. This operational region ($I_D$ vs $V_{DS}$) is shown in Fig. 3.1a, with the "on" and "off"-resistances highlighted. In the proximity of the threshold voltage, around -0.7 V in this example, the channel resistance varies strongly as a function of $V_{gs}$, see Fig. 3.1b. The operation of the resistive FET mixer is based on the external control of the channel resistance, where the large signal LO is applied on the gate, in principle switching the channel resistance between the "on" and the "off" states. Simultaneously, the RF or IF signal is applied to the drain, where the switched channel resistance mix the RF or IF signal.

Figure 3.1: Channel resistance versus $V_{GS}$.

The quadrature modulator in [C] follow this design principle, such that the conductance waveform has a large fundamental amplitude, and therefore, mixing efficiently on the fundamental frequency, Fig. 3.2a. In [73], a sub-harmonic FET resistive mixer relies on the same operation, but optimizes the quiescent bias such that the second harmonic of the conductance waveform is maximized. By combining two FETs in parallel, and driven by complementary LO signals, the channel conductance waveform will vary every two LO cycles as shown in Fig. 3.2b. This pulse train of variation in channel conductance every half cycle varies the second harmonic channel conductance more efficiently and therefore improves conversion efficiency and linearity [74]. The design in [D] relies on this mixing principle.

Consider the idealized resistive mixer in Fig. 3.3, where the channel conductance only depends on the LO $v_{LO}$ and source ($Z_S$) and load ($Z_L$) are ideally behind lossless complementary bandstop and bandpass filters. Outside their pass bands, both filters are idealized with infinite input impedances. In this example, the source impedance $Z_S$ is set to 0 Ω for simplifying the mathematics. The impact of the filters makes the source current $i_S = v_{RF} \times Y_S \parallel g_{max}$, and $i_L(t) = i_S(t) - i_G(t)$ [3.4].
\[ G(t) = \frac{g_{\text{max}}}{2} + \frac{2 g_{\text{max}}}{\pi} \sum_{k=1}^{\infty} \frac{\sin((2k-1)\omega_{\text{LO}} t)}{2k-1} = \frac{g_{\text{max}}}{2} + \frac{2 g_{\text{max}}}{\pi} \left( \sin(\omega_{\text{LO}} t) + \frac{1}{3} \sin(3\omega_{\text{LO}} t) + \frac{1}{5} \sin(5\omega_{\text{LO}} t) + \ldots \right) \] (3.5)
\[ i_L(t) = i_S(t) - G(t) \times v_{RF}(t) = -\frac{2g_{max} v_{RF}}{\pi} \sin(\omega_{LO} t) - \frac{2g_{max} v_{RF}}{3\pi} \sin(3\omega_{LO} t) - \ldots \] (3.6)

In Section 3.1, the definitions for a linear, non-linear, time invariant and time variant system were described. When the channel conductance \( G(v_{LO}, v_{RF}) \) solely depends on the LO signal \( v_{LO} \), i.e. it does not depend on the RF signal \( v_{RF} \), then, it categorizes the resistive mixer as an LTV system. Analysis of linear frequency translation systems is described in [76]. However, if current saturation is present, like in Fig. 3.1a, the channel conductance also depends on the signal drive level. When this non-linear behavior is taken into account and the conductance waveform depends on both \( v_{LO}(t) \) and \( v_{RF}(t) \) such that \( G(v_{LO}(t), v_{RF}(t)) \), the system is characterized as non-linear and time variant. For explanation, the problem is simplified to model \( g_{max} \) in (3.5) with dependence of \( v_{DS} \). The mixing products of this non-linear system are best illustrated with Taylor series expansion around the operating point, here \( v_{DS} = 0 \) (3.7). For further and detailed analysis of third order intermodulation in mixers, the reader is referred to [77].

\[ g'_{max}(v_{DS}) \approx g_{max}(0) + \frac{\delta g_{max}(0)}{\delta v_{DS}} (v_{DS} - 0) + \frac{\delta^2 g_{max}(0)}{\delta v_{DS}^2} \frac{(v_{DS} - 0)^2}{2} + \]
\[ + \frac{\delta^3 g_{max}(0)}{\delta v_{DS}^3} \frac{(v_{DS} - 0)^3}{6} + \ldots \] (3.7)

This causes signal distortion, and the highlighted term in (3.7) is what generates IM3 products. For a modulated signal, this mixing products fall inside the signal’s spectrum and cause increased EVM and as a result, higher bit error rate. Because it is within the signal bandwidth, this term is not filterable and requires either improved linear designs or linearization techniques as demonstrated in [13] and Section 2.4, i.e. to lower \( g_3 \) in order to reduce the distortion, [35]. Furthermore, the IM3 spans three times the original signal bandwidth, causing interference with adjacent channels. Using the test signal in (3.2), the important third order term is the mixing products of \( b_1 \cos(\omega_1 t) \) and \( b_2 \cos(\omega_2 t) \), producing (3.8).

\[ \left( b_1 \cos(\omega_1 t) + b_2 \cos(\omega_2 t) \right)^3 = \frac{3b_1^2 b_2}{4} \cos \left( (2\omega_1 - \omega_2) t \right) + \frac{3b_1 b_2^2}{4} \cos \left( (2\omega_2 - \omega_1) t \right) + \ldots \] (3.8)

### 3.3 Mixers topologies

In analogy with amplifiers, mixers are in a similar manner classified to their function, to describe their properties. The second order intermodulation example in (3.3) captures the spectral components that arise from second order
signal mixing, frequency components that in a transmitter or receiver may cause problems. Mixer classes such as balanced mixers, quadrature mixers, image reject mixers and harmonic mixers are combinations of multiple mixers, phase shifters and, quadrature and out-of-phase signals, configured to suppress unwanted frequency mixing components and to isolate the ports from each other. Based on the simplified resistive mixer operation described in Section 3.2, and considering no other elements than the voltage dependent conductance $G(v_{GS})$, the fundamental mixer configurations of the modulator designs in [D] and [C] are reviewed in the following sections by mixing terms and port isolation. In reality, the model is much more complicated. Capacitors and non-linear behavior of components contribute to leakage between ports and harmonic generation that are not considered here. For an accurate prediction of the performance, complex behavioral models and HB must be used.

### 3.3.1 Fundamental and harmonic mixers

In Fig. 3.4, the most simple configuration of a mixer is shown. It has two input ports, the LO and the intermediate frequency (IF) port for up-conversion mixing, and one output port, the RF. When the RF, IF and LO frequencies are sparsely separated from each other, this configuration is often used with simple matching networks and filters at each port [78]. Image frequency suppression and port isolation are solved with filters.

$$\text{Figure 3.4: Standard mixer.}$$

$$\begin{align*}
\text{Frequency (Hz)} & \quad P_{\text{OUT}} (\text{dBm}) \\
\text{f}_f & \quad f_{IF} & \quad f_{LO} - f_{IF} & \quad 2 f_{LO} - f_{IF} \\
\text{f}_{LO} & \quad 2 f_{LO} & \quad 2 f_{LO} + f_{IF} & \quad 2 f_{LO} + f_{IF}
\end{align*}$$

$$\text{Figure 3.5: Up-converted frequency composition of a standard mixer.}$$

Consider the LTV system in (3.1) to represent the mixer in Fig. 3.4, where the channel conductance depends on $v_{GS}$ as shown in Fig. 3.1b. By using Taylor series around the operating point, $G(v_{GS})$ can be expressed with a general polynomial function (3.9). When applying the input signals $v_{in}(t)$ and $v_{LO}(t)$ to the mixer (3.10), then, the time varying conductance waveform takes the shape (3.11), and as stated in Section 3.1, when multiplied with
3.3. MIXERS TOPOLOGIES

a time varying coefficient, \(v_{in}(t)\) will be linearly translated to the sidebands around the LO harmonics. Let \(b_0\) represents an input dc level such that leakage between the LO and RF ports is present. The frequency composition of the output signal \((3.12)\) contains, the linear frequency translation to the lower and upper side bands \((\omega_{LO} \pm \omega_{IF})\) and the LO and IF frequencies (Fig. 3.5). This explains the most simple linear mixer, of which two relevant problems for direct conversion (homodyne) systems arise: LO leakage and double sideband mixing.

\[
G(v_{GS}) = \sum_{n=0}^{\infty} g_n v_{GS}^n \quad (3.9)
\]

\[
v_{in}(t) = b_0 + b_1 \sin(\omega_{IF} t)
\]

\[
v_{LO}(t) = \sin(\omega_{LO} t) \quad (3.10)
\]

\[
G(v_{LO}(t)) = g_0 + \sum_{n=1}^{\infty} g_n \sin^n(\omega_{LO} t) \quad (3.11)
\]

\[
i_{out} = G(v_{LO}(t)) v_{in}(t) = \left( g_0 + \sum_{n=1}^{\infty} g_n \sin^n(\omega_{LO} t) \right) \left( b_0 + b_1 \sin(\omega_{IF} t) \right)
\]

\[
= g_0 b_0 + g_1 b_0 \sin(\omega_{LO} t) + g_0 b_1 \sin(\omega_{IF} t) +
\]

\[
+ \frac{g_1 b_1}{2} \left( \cos \left( (\omega_{LO} + \omega_{IF}) t \right) + \cos \left( (\omega_{LO} - \omega_{IF}) t \right) \right) + \ldots \quad (3.12)
\]

3.3.2 Balanced mixers

When the LO, RF and IF frequencies are too close to be separated by filters in MMIC technology, a balanced configuration can provide the necessary isolation between ports, without the need for filters to separate the signals. The balanced mixer in Fig. 3.6, is a suitable configuration in up-conversion systems when the LO and RF frequencies are closely separated or in homodyne systems where the LO appear as the carrier frequency. This topology relies on out-of-phase IF \((v_{in}(t))\) and LO \((v_{LO}(t))\) signals such that the mixing product of the IF and LO adds constructively, while the IF and LO signals add destructively at the RF port. Using the LTV system in \((3.6)\) to represent one mixer, then, as shown in Fig. 3.7, the frequency composition of the output signal contains the lower and upper sidebands, with all odd LO harmonic components isolated from the RF port \((3.13)\). This mixer serves as the basis of the design in \([C]\).

\[
i_{out} = G(v_{LO}(t)) v_{IF}(t) + G(v_{LO}(-t)) v_{IF}(-t) = b_0 \sum_{n=1}^{\infty} 2 g_{2n} \sin^{2n}(\omega_{LO} t) +
\]

\[
+ g_1 b_1 \left( \cos \left( (\omega_{LO} + \omega_{IF}) t \right) + \cos \left( (\omega_{LO} - \omega_{IF}) t \right) \right) + \ldots \quad (3.13)
\]
Figure 3.6: Single balanced mixer.

Figure 3.7: Up-converted frequency composition of a single balanced mixer.

For a second harmonic transmitter mixer, isolation towards the RF port is important for the second LO harmonic. In the balanced configuration (Fig. 3.6), there is no isolation towards the RF port for the even harmonics, instead the periodicity of the non-isolated LO harmonics for the balanced mixer repeats with $2 \times n$ (3.13). To suppress and isolate the critical second LO harmonic from the RF-port, the balanced mixer requires in addition to complementary signals, supplementary $90^\circ$ and $270^\circ$ degrees LO signals, comprised into the second harmonic balanced mixer, shown in Fig. 3.8. The additional supplementary $90^\circ$ and $270^\circ$ degree LO signals make the even order LO harmonics of $2, 6, 10, \ldots$ to add destructively. Therefore, the second harmonic balanced mixer generates constructive interference for all LO harmonics with a periodicity of $4 \times n$ (3.14). In sum, this arrangement isolates the LO harmonic numbers of $1, 3, 5, 7, \ldots$ from the RF-port, including the important second LO harmonic.

$$i_{out} = b_0 \sum_{n=1}^{\infty} g_n (1 + j^n) \cdot (\sin^n (\omega_{LO} t) + \sin^n (-\omega_{LO} t))$$

$$= b_0 \sum_{n=1}^{\infty} 4 g_{4n} \sin^{4n} (\omega_{LO} t)$$

(3.14)

In order to mix with the second LO harmonic in the balanced mixer cell
while maintaining $2 \times$ LO isolation towards the RF port, the sub-harmonic balanced mixer must employ complementary IF signals for constructively interfering with the second LO harmonic as illustrated in Fig. 3.8. The mixing terms for the balanced mixer cell is shown in (3.15), where the fundamental out-of-phase IF signal is added to the terms. Those mixing terms instead have a periodicity of $4 \times n - 2$. The second harmonic balanced mixer has the targeted properties for a direct conversion up-converting sub-harmonic mixer, i.e. the RF port is isolated from the second LO harmonic while mixing at the second LO harmonic as shown in Fig. 3.9. This mixer serves as the basis of the design in [D].

![Second harmonic balanced mixer diagram](image)

**Figure 3.8: Second harmonic balanced mixer.**

$$i_{out} = b_1 \sum_{n=1}^{\infty} g_n \left( \sin (\omega_{IF} t) + \sin (-\omega_{IF} t) j^n \right) \cdot \left( \sin^n (\omega_{LO} t) + \sin^n (-\omega_{LO} t) \right)$$

$$= b_1 \sum_{n=1}^{\infty} 4 g_{4n-2} \sin (\omega_{IF} t) \sin^{4n-2} (\omega_{LO} t)$$

(3.15)

![Frequency composition graph](image)

**Figure 3.9: Frequency composition of a second harmonic balanced mixer.**
3.3.3 The quadrature mixer

Both the fundamentally and sub-harmonically balanced mixer configurations in Section 3.3.2 translate the baseband signal to the LO carrier frequency. However, the signal folds on both the positive and negative frequency planes, around the carrier frequency. Apart from occupying twice the spectra needed, down-converting both sidebands is problematic in terms of signal reconstruction. This is because the positive and negative sidebands may add destructively if the receiver LO is orthogonal to the carrier LO, zeroing the wanted signal \( v_{\text{in}}(t) \) (3.16). In this case, the carrier phase must be recovered and synchronized. This already cumbersome problem is aggravated since the carrier is an unwanted spur, suppressed by the balanced topologies in Figs. 3.6 and 3.8. It is due to these reasons that the commonly used heterodyne or super-heterodyne architectures (using intermediate frequencies) rely on selective filtering to suppress the image frequency to avoid signal cancellation and phase recovery [79]. In homodyne systems, the frequency separation between the lower and upper side-bands is zero, and therefore, they are inseparable by filters.

\[
v_{\text{out}}(t) = v_{\text{in}}(t) \cos(\omega_{LO} t) \sin(\omega_{LO} t) = 0 v_{\text{in}}(t) + \frac{v_{\text{in}}(t)}{2} \sin(2\omega_{LO} t) \quad (3.16)
\]

Direct conversion mixers therefore mandate quadrature LO (cosine and sine), with two independent input/output signals, called in-phase (I) and quadrature (Q) [79,80]. The quadrature mixer topology is shown in Fig. 3.10, where the LO is split in 0 (cosine) and 90 degrees (sine) waves and two input ports (I) and (Q). Because there is no correlation between the input signals, again, the signal data will fold on both sides of the carrier (3.17). The difference between the quadrature mixer and the standard mixer lies in the signal demodulation, where the I and Q signals can be separated from each other, independently of the carrier phase. Mathematically, following trigonometric rules, the I and Q signals can be recovered at any LO phase state, only phase rotation is necessary to sort \( v_I(t) \) from \( v_Q(t) \) (3.18). The implementation and the basis of modern communication are based on complex numbers and the complex plane [81]. The beauty and simplicity of QAM modulation and the symbiosis with the simple direct conversion quadrature transmitter and receiver architectures has made direct conversion the preferred architecture in today’s radio equipment [80]. In addition to balanced configurations, the modulators in [C] and [D] rely on the quadrature principle.

\[
v_{\text{out}}(t) = \left( v_I(t) \cos(\omega_{LO} t) + v_Q(t) \sin(\omega_{LO} t) \right) 
\]

\[
v_{\text{out}}(t) \cos(\omega_{LO} t) = \frac{1}{2} \left( v_I(t) + v_I(t) \cos(2\omega t) + v_Q(t) \sin(2\omega t) \right) = \frac{v_I(t)}{2} 
\]

\[
v_{\text{out}}(t) \sin(\omega_{LO} t) = \frac{1}{2} \left( v_Q(t) + v_I(t) \sin(2\omega t) - v_Q(t) \cos(2\omega t) \right) = \frac{v_Q(t)}{2} 
\]

(3.18)
3.3.4 Single sideband mixer

If the two input signals $v_I(t)$ and $v_Q(t)$ are identical, only shifted 90 degrees, the up-converted signal appears only on one side of the carrier as shown in Fig. 3.11 and (3.19). The single sideband mixer is therefore an incremental extension of the quadrature mixer, Fig. 3.12. In the verification of the quadrature modulators in [D] and [C], external 90° and 180° hybrids are used, configuring the setup as a single sideband mixer. The measure of quadrature balance is therefore expressed as Image Rejection Ratio (IRR) and is measure of upper side band divided by the lower side band.

$$v_{\text{out}}(t) = v_{\text{in}}(t) \cos(\omega_{\text{LO}} t) + j v_{\text{in}}(t) \sin(\omega_{\text{LO}} t) = v_{\text{in}}(t) e^{j \omega t} \quad (3.19)$$
3.4 Quadrature modulator designs for transmitters

The balanced mixer topologies described in Section 3.3.2, in combination with the quadrature topology in Section 3.3.3 form the basis of the designs in [C] and [D]. Four phase states of 0°, 90°, 180° and 270° degrees with equal amplitude are needed for providing the necessary function. As a consequence, the transmitter quadrature modulators have many passive components in common for realizing the phase distribution networks. In this section, special emphasis is made on the phase distribution networks. The circuit designs, transverse electro-magnetic (TEM) and transmission line modes for passive circuits, mixer simulation environment and the measurement results are presented in the pursuit of high performing circuits.

3.4.1 Electromagnetic modes and transmission lines

Transmission lines suspended on dielectric materials, such as in MMIC designs, are used for propagating quasi TEM waves. Quasi TEM because the microstrip mode has a relatively strong longitudinal electric component due to the large dielectric mismatch in-between air and the substrate [82]. The designs in [D] and [C] rely on microstrip and coupled lines to guide the microstrip, common- and differential-modes, shown in Fig. 3.13. Their characteristics, such as characteristic impedances and electrical lengths depend on the substrate thickness, transmission line width(s), signal to signal spacing and their effective dielectric constants. In the design of a Marchand balun in Section 3.4.3, the analysis on the common- and differential modes’ propagation characteristics, shows vital for the performance.

![TEM modes' electric field distribution](image)

Figure 3.13: TEM modes’ electric field distribution suspended on a dielectric substrate (light blue). Gray-colored areas are ground.
3.4.2 Circuit simulation environment

Mixers are demanding components to simulate. In comparison, amplifiers can, to a large extent, be designed based on small signal simulations only, based on SP. Frequency response, port RL and, to some degree, even optimum large signal conditions can be designed based on load-pull data in a small-signal environment. For the resistive mixer, the large variation of channel resistance during the LO cycle described in Section 3.2, requires large signal simulations to assimilate the operation. A mixer simulation template, considering ports’ return losses at large LO drive levels, including other common mixer parameters such as CL and linearity versus power, frequency and bias sweeps was developed with HB simulation engine \cite{83,84} in Keysight ADS software and shown in Fig. 3.14. In symbiosis with \eqref{eq:2.13}, where expressions for scattering waves were adopted to HB, port matching can as such be determined under the influence of large signals \eqref{eq:3.21}. Considering reflection on port one, the source \(v_{RF}\) needs to be turned off in order not to interact with the mixing products \eqref{eq:3.21}.

\begin{equation}
\begin{align*}
a_1(nf_{IF}, mf_{LO}) &= \frac{1}{2\sqrt{Z_0}} \left( v_1(nf_{IF}, mf_{LO}) + Z_0 \cdot i_1(nf_{IF}, mf_{LO}) \right) \\
b_1(nf_{IF}, mf_{LO}) &= \frac{1}{2\sqrt{Z_0}} \left( v_1(nf_{IF}, mf_{LO}) - Z_0 \cdot i_1(nf_{IF}, mf_{LO}) \right)
\end{align*}
\tag{3.20}
\end{equation}

\begin{equation}
LS_{11} = \frac{b_1(1,0)}{a_1(1,0)} \bigg|_{v_2=0} \tag{3.21}
\end{equation}

Figure 3.14: Simulation environment of an up-converting mixer.
In equality with amplifier designs, the performance of the resistive mixer is determined by its load impedance. Parameters such as $RL$, conversion efficiency, linearity and $P_{SAT}$ can be obtained by load-pull simulations or measurements. The method has, however, not attracted much or any attention for mixers. From a system perspective it is understandable since their efficiency and output power have little impact on the overall performance. However, it is a highly appreciated tool from a circuit design perspective. For this reason, the amplifier load-pull simulation environment (2-Tone Nonlinear Simulations/Load Pull - PAE, Output Power, IMD Contours) in ADS was modified to suit mixers. In Fig. 3.15, the CL and OIP3 versus load impedance have been plotted. The load-pull results show resemblance with amplifiers, where the maximum gain and maximum output power in most cases occur at different load impedances. It is also apparent from Fig. 3.15a that the drain-source capacitance ($C_{ds}$) lowers the optimum output impedance ($R_{OPT}$) for maximum conversion efficiency (minimum CL) at high frequencies. These two simulation environments allow for efficient and systematic mixer designs.

![Diagram of CL and OIP3 vs load impedance](image)

**Figure 3.15:** Simulated CL and OIP3 versus swept load-impedances of the mixer-cell in [C]. In (a) Minimum CL is 8 dB, each contour step is 0.5 dB. In (b) Maximum OIP3 is 15 dBm and each contour step is 1 dB.

In [C] the optimum load impedance for linearity at the RF port was realized by coupled lines [85] and in [D] by quarter-wavelength transformers, accepting worse RL. The I and Q ports on the other hand, need low RL from dc, which limits the available matching tools to LO power, gate bias and transistor size.

### 3.4.3 Fundamental balanced E-band modulator

The balanced quadrature modulator MMIC in [C] consists of two balanced resistive mixer cells and several passive elements such as matching networks, filters [86], Wilkinson power combiners [66], 90° degree branch-line coupler [82] and 180° degree Marchand balun [87] to covert the desired function. The circuit schematic and layout are shown in Fig. 3.16, including the mixer cells and passive components. The phase distribution network is located on the gates’ side (LO). In contrast to the sub-harmonic modulator in [D] where
the 90° degree hybrid is located on the RF side, this implementation can tolerate a higher quadrature amplitude imbalance because CL saturates for high LO power level. On one hand, this is advantageous for wideband signals, since the phase offset is constant across the signal bandwidth. On the other hand, the 90° degree hybrid provides low RL at the RF port in [D] which the power combiner in [C] does not. In this section, special attention is given to the differential branch-line coupler and the common-mode reflection filter. Finally, the most significant measurement results are presented.

Figure 3.16: Chip photograph and schematic of the single balanced E-band quadrature modulator presented in [C]. The chip area measures 2.5 × 1.5 mm².

The differential branch-line coupler’s principle of operation is, in theory, the same as microstrip branch-line coupler [82], although the propagating mode is different. It consists of four quarter wavelength coupled transmission lines with odd mode characteristic impedances of $Z_{0o} = 71 \, \Omega$ and 100 Ω and termination resistance 100 Ω, shown in Fig. 3.17. The introduction of a differential load resistance ($Z_L$) is beneficial for the frequency response. The differential resistance contributes with indefinitely small inductance in comparison to resistance to ground, using a via-hole. When fed with a differential source, all four required phase states of 0°, 180°, 90° and 270° degrees at port 2 and port 3 are obtained. It occupies the same chip area as the microstrip branch-line coupler and as a positive consequence, this relaxes the need for chip area for the complete circuits.

Figure 3.17: Differential branch-line coupler in GaAs technology.
Amplitude and phase balances are the most important FOMs for a branch-line coupler, while for differential designs, mode conversion is important. For quadrature modulators, these FOMs are reflected in the IRR and in the LO to RF port isolation. Differential excitation at port 1, should ideally provide differential signals to port 2 and port 3, but due to asymmetries in the differential branch-line coupler, transformation from differential-mode to common-mode is inevitable. The simulated differential phase and amplitude balance between port 2 and port 3, RL and mode-conversion from port 1 to ports 2 and 3 responses are shown in Fig. 3.18. The results show better than 1 dB amplitude error, 5° degree phase error and 10 dB RL over a frequency span from 65 to 92 GHz. Differential to common-mode conversion in Fig. 3.18d is better than 23 dB for both $S_{2c1d}$ and $S_{3c1d}$.

![Figure 3.18: Simulated performance of the differential branch-line coupler.](image)

To improve the RF to LO port isolation, a common-mode filter was proposed and implemented in the design. The filter relies on a $\lambda/2$ long microstrip line connected between the differential lines, shown in Fig. 3.19a. Along the symmetry line, the electrical conditions are a perfect magnetic boundary (open circuit) for the common-mode and on the contrary for the differential-mode, a perfect electric boundary (short circuit). The effects are, after transforming the two boundary conditions via the $\lambda/4$ long microstrip line, an open circuit for the differential-mode and a short circuit for the common-mode, re-
fecting the common-mode signal. The simulated differential-mode (blue) and common-mode (red) return losses and transmissions are shown in Figs. 3.19c and 3.19d respectively. The circuit schematic and layout of the common-mode rejection filter are shown in Fig. 3.19a and Fig. 3.19b respectively.

Figure 3.19: Schematic, layout and simulation results of the common-mode filter.

The incorporated differential branch-line coupler and common-mode rejection filters are reflected on the measurement results, especially in terms of high IRR and high LO to RF isolation in Fig. 3.20a. Across the three E-bands (71-76 GHz, 81-86 GHz and 92-95 GHz), IRR, LO to RF isolation and CL are measured higher than 20 dB, 30 dB and better than 11 dB respectively. The high OIP3 of 13 dBm (Figs. 3.20b and 3.20c) and low CL (Fig. 3.20a) form the basis for wideband communications and permits high IF input power to ensure high SNR ratio.

Figure 3.20: Measurements results when $f_{LO} = 86$ GHz, $p_{LO} = 13$ dBm, $f_{IF1} = 0.99$ GHz, $f_{IF2} = 1.01$ GHz and $V_G = -0.7$ V.

### 3.4.4 Subharmonic harmonically balanced D-band modulator

The circuit topology in [D] is based on the mixer cell presented by [74], but combines four mixer cells, driven at different phases to achieve second harmonic isolation. In Fig. 3.21a, the circuit schematic of the quadrature harmonic modulator is shown, including the encircled mixer and balanced mixer sub-cells in blue short dash-dots and red dashes respectively. In Fig. 3.21b, the
mixer cell and balanced mixer cell are shown in more detail, including the transistors and input and output matching networks. The resemblance with the modulator design in [C] is significant. The phase distribution network on the LO side is principally equal. It consists of the same phase states \((0^\circ, 90^\circ, 180^\circ\), and \(270^\circ\)) although two-folded and comprises four mixer cells rather than two. The passive networks consist of two differential branchline couplers, one Marchand balun and one Lange coupler, all together combined into a 3.2 mm\(^2\) chip area (Fig. 3.22).

![High level circuit schematic.](image)

![Circuit schematic of the input and output matching networks of the balanced mixer cell.](image)

**Figure 3.21:** Circuit schematic of the second harmonic quadrature modulator.

![Layout of the second harmonic quadrature modulator.](image)

**Figure 3.22:** Layout of the second harmonic quadrature modulator. The chip size measures 2 × 1.6 mm\(^2\).

The LO to RF isolation depends on symmetry and the balance on the phase
distribution networks. The design and performance of the differential branch-line coupler was described in Section 3.4.3 and in this section, the design and analysis of a three wire high common-mode rejection ratio (CMRR) Marchand balun is presented. The design of high CMRR Marchand baluns requires some fundamental understanding about the passive circuit in order to approach the problem. In a planar technology like the GaAs processes, the inhomogeneous media of air and dielectric, sandwiched around the signal conductor, makes different modes propagate at different velocities. Depending on the modes in Figs. 3.13a to 3.13c, the electric field is distributed differently between air and the dielectric, which results in different effective dielectric constants. In a coupled structure such as the Marchand balun or directional coupler, the difference in group velocity between the even and the odd mode has a direct negative impact on the performance, resulting in low CMRR for the balun and poor directivity for the directional coupler. Numerous analysis and compensation techniques are demonstrated to overcome the problem, such as additional capacitive coupling in-between lines [88] and wiggy lines [89]. The approach in [90] by combining N unique coupled lines with different widths, spacing and lengths result to have the same even-mode and odd-mode group velocities. The balun in [D] is based on this approach. The even-mode and odd-mode propagation velocities on a three-wire coupled line in [D] show almost identical group velocities and therefore automatically achieve high CMRR in the Marchand balun structure Fig. 3.23. The schematic of the Marchand balun is shown in Fig. 3.24a and the circuit implementation of the 80 GHz Marchand balun is shown in Fig. 3.24b. The results show over wide bandwidth 60-100 GHz, RL better than 23 dB, CMRR higher than 36 dB and therefore better than 0.2 dB amplitude match and 1° degree phase imbalance Fig. 3.25.

![Figure 3.23: Simulation of propagation velocity for the differential-mode and common-mode of the coupled lines.](image)

(a) 2-wire coupled line.  (b) 3-wire coupled line.  (c) 4-wire coupled line.

Figure 3.23: Simulation of propagation velocity for the differential-mode and common-mode of the coupled lines.

![Figure 3.24: Circuit schematic and layout of the implemented Marchand balun.](image)

(a) Circuit schematic.  (b) Circuit layout.

Figure 3.24: Circuit schematic and layout of the implemented Marchand balun.
CHAPTER 3. MIXER THEORY AND DESIGN OF LINEAR QUADRATURE MODULATORS

Figure 3.25: Simulated results of the implemented Marchand balun.

The rational designing and high performance passive components have resulted in state-of-the-art measured performance. From 135-170 GHz, IRR, LO to RF isolation, linear output power and CL are measured higher than 15 dB, 30 dB (70 dB with dc-offsets), -2 dBm and 11 dB respectively. In Fig. 3.26a, the CL, IRR and 2 × LO isolation are plotted versus RF frequency at the optimum $V_G$ for minimizing CL. Unlike fundamental mixers, subharmonic mixers depend both on gate bias and LO power for achieving the lowest CL [91]. By optimizing both the $V_G$ and $P_{LO}$ for the lowest CL, a flat CL response is achieved over a range of LO input power levels. In Fig. 3.26b, the simulated and measured CL, IRR and 2 × LO isolation are plotted versus IF input power. The CL is flat up to -2 dBm. In Fig. 3.26c, the measured 2 × LO isolation is shown with and without applied dc offset on the I+, Q+ and Q- input ports. By individual adjustments of the voltage levels between -200 mV and 300 mV, the 2 × LO isolation can be suppressed an additional 40 dB, to 70 dB.

Figure 3.26: Measured performance of the sub-harmonic quadrature modulator.
Chapter 4

Baseband Variable Gain Amplifier for wideband operation

Communication systems opting for Multi-Gbit/s, require several Gigahertz of signal bandwidth even for spectral efficient modulation formats [92]. On the electronics side, transmitters and receivers at 110-170 GHz have demonstrated the feasibility of wireless transmission at 44 Gbit/s using quadrature phase shift keying (QPSK) modulation [93,94]. For that, a baseband bandwidth of 22 GHz or higher is required, depending on the signal’s pulse shaping [72]. In optic communications, the 2nd generation 100 Gigabit Ethernet (100 GbE) [95,96], uses four channels with pulse-amplitude modulation (PAM)-4 modulation at 25 Gbit/s sample rate per channel, which translates to 12.9 GHz signal bandwidth [97]. The development of the 3rd generation 100 GbE, is on a similar basis based on PAM-4 but divided by 2 channels. With data rates of 50 Gbit/s per channel, the required signal bandwidth is 25.8 GHz [98]. Future Terabit Ethernet aim for 100 Gbit/s per channel [99,100], which will require even wider bandwidths. In this context, the development of wideband baseband circuits is essential for future communication equipments.

Variations in transmission conditions, for example, rain in a microwave link, or circuits’ gain variation with temperature, make it necessary to have gain adjustable components, to compensate for external influence. The VGA is such a component, where manifold topologies exist. For example, variable active feedback [101], variable source degeneration [102], variable attenuator [103] or the Gilbert cell [104], of which the latter form the basis of this design.

In the context of wideband communication, several techniques for enhancing the bandwidth and the gain bandwidth product (GBP) of baseband Gilbert cell VGAs have been reported. In [105], a negative resistance towards the emitter is generated by a cross-coupled pair on compensate for the contact resistance [106]. A comparison between negative emitter resistance and 0 Ω, report 354 GHz GBP and 176 GHz GBP, respectively. In [107], capacitively loaded emitter followers at the input stage contribute to frequency peaking. Additional techniques such as inductive loading and emitter degeneration, compris-
ing a parallel resistance and capacitance, increased GBP to 113 GHz. In [108], a transimpedance load circuit, featuring low input impedance and complex poles, contributed to increase the GBP to 146 GHz. The VGA in [E] has reported 1.42 THz GBP, which is the highest of VGAs based on the Gilbert cell topology.

In this chapter, the circuit design and design methodology of [E] are presented. Dominant pole analysis, using ZVT constants to identify parts of the design, accountable for the 3dB bandwidth is explained. Technologies for Radio Frequency Integrated Circuit (RFIC) designs and layout in 3-D BEOL are presented.

4.1 Design considerations and bandwidth analysis

The VGA in [E] has reported the highest GBP of VGAs, based on the Gilbert cell topology. The great results are primarily due to the high-speed technology [38], but also assimilated by classic design techniques in combination with careful analysis and layouting. The circuit design comprises transistors in all configurations; common-emitter (C-E), common-base (C-B) and common-collector and pn-diodes based on the transistor. Moreover, design techniques such as inductive peaking in the gain control amplifier, cascode current sources, cascode output stage and bias reference ladder are examples of implemented configurations in the wideband VGA. The schematic in Fig. 4.1 shows the complexity of the circuit and reveals the different configurations.

4.1.1 Transfer response

This section introduces the transfer response and its relationship with network circuit theory. Node admittance (NA) and modified node admittance (MNA) matrices are established methods for describing electrical networks of any type [109,110], such as the VGA schematic in Fig. 4.1, and are used by circuit simulators to construct and calculate responses of electrical networks. Each node corresponds to one row and one column in the NA based matrix, and a minimum of one row and one column per node for the MNA matrix. By matrix operations on the NA and MNA matrices, information about different responses can be obtained. The transfer response can, for example, be expressed by the ratio of co-factors and determinants, and can be derived from general N-port network theory [109,111]. Thus, it becomes a rational function with zeros and poles in the complex frequency plane \( s \) and consequently the polynomial order may be as high as the number of nodes (4.1). While equation (4.1) solves the frequency response, it is an immense computational task to symbolically calculate the poles and zeros for a circuit of the scale in [E]. Even worse, it would only bring incomprehensible equations that would be impossible to decipher. Circuit simulators such as Keysight ADS are powerful tools for circuit design, however, one weakness is related with numerically solved responses and therefore, they do not provide information about the zeros and poles, neither to which components they originate.
4.1. DESIGN CONSIDERATIONS AND BANDWIDTH ANALYSIS

Figure 4.1: Detailed schematic of the VGA showing the full circuit implementation including the gain control amplifier, cascode amplifier and Current sources.

\[ H(s) = \frac{v_{out}(s)}{v_{in}(s)} = \frac{G(s-z_1)(s-z_2)\ldots(s-z_N)}{1(s-p_1)(s-p_2)\ldots(s-p_N)} \] (4.1)

4.1.2 Zero Value Time constants

ZVT constants is a method that eases the calculation of the transfer function \((v_{out}/v_{in}\) for example) to a first order denominator polynomial, and therefore can determine the dominant time constant of the pole and estimate the 3 dB bandwidth (4.2) [112]. In [E], this method for analyzing the dominant pole is applied, and by tuning design parameters and bias, information about the time constants is retrieved to improve the 3 dB bandwidth. The great benefits of this method, apart from estimating the 3dB bandwidth, are the details each time-constant \((R_i C_i)\) provides about their contribution to the circuit’s bandwidth.

\[ |H(\omega)| \simeq \frac{G}{\sqrt{1 + \left(\frac{\omega}{p_1}\right)^2}} \Rightarrow \omega_{3\,dB} \simeq p_1 \quad \forall p_1 \in \mathbb{R} \] (4.2)
Firstly, the poles originate from the determinant of the admittance matrix \( \Delta \) and second, that the first order pole stems from the diagonal elements \[112, 113\]. Each capacitor \( C_i \), discharging the Thevenin equivalent resistance across the same capacitor gives rise to time constant \( \tau_1 \), where \( p_1^{-1} \) is the sum of all first order time constants \((4.3)\). The 3 dB bandwidth of the transfer function can therefore be estimated even for complex networks utilizing \((4.2)\) and \((4.3)\) \[113\].

\[
\tau_1 = \sum_{i=1}^{n} R_i C_i \\
p_1 = \tau_1^{-1} \quad \forall p_1 \in \mathbb{R} \quad (4.3)
\]

The ZVT analysis was conducted on the VGA using the small signal model in Fig. \[4.2\] with three sets of parameters for three different current settings in Table \[4.1\]. The VGA topology employs two types of circuits, the cascode amplifier and the emitter follower. The Gilbert cell is configured similarly to the cascode amplifier and is therefore treated equally throughout the analysis. By evaluating the expressions for each time constant, the analysis showed that the base-emitter capacitance \( C_\pi \) in the C-E transistors in the gain control amplifier \((Q_1, Q_2)\) had the largest contribution to the pole, 1.8 ps. Since those transistors were already accompanied by a C-B transistor, and thereby low loading, the effects on bandwidth and GBP were investigated for load resistances \((R_6, R_7)\), emitter degeneration \((R_2)\) and current bias. To achieve the highest GBP; it was concluded that high load resistances, low degeneration and high bias current should be used. A high bias current increase the GBP because 1) \( f_T/f_{max} \) increase with current density and 2) the input impedance of the emitter followers decreases because of the higher transconductance, lowering the times constants related with \( C_\pi \). However, a high current bias and high load impedance will result in large differences in collector-emitter voltage between \((Q_3 \text{ and } Q_4)\), and \((Q_5 \text{ and } Q_6)\) at maximum gain. Therefore, in this design, the load resistors \( R_6 \) and \( R_7 \) were selected to 200 \( \Omega \) and the gain control amplifier bias current was selected to 10 mA. To regain higher 3 dB bandwidth from this stage, 300 pH inductors \((L_1 \text{ and } L_2)\) were introduced to peak the gain in proximity of the gain drop. In Fig. \[4.3\] the calculated gain responses from ZVT constants are compared with the measured and simulated. The low frequency gain is almost identical, and the estimated 3 dB bandwidth is 22 GHz compared to measured 40 GHz. In sum, the method was useful for identifying the limiting part of the design, in order to concentrate efforts on that part of the circuit. However, to achieve accuracy between the simulations and measurements, CAD tools and accurate models are necessary.
4.1. DESIGN CONSIDERATIONS AND BANDWIDTH ANALYSIS

Figure 4.2: Small signal model of the InP DHBT.

Table 4.1: List of intrinsic parameters for a 0.25 × 10 μm² large transistor biased at $V_{ce} = 2$ V for three different current densities.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>2mA/μm²</th>
<th>4mA/μm²</th>
<th>8mA/μm²</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_b$</td>
<td>1.4 Ω</td>
<td>1.3 Ω</td>
<td>1.2 Ω</td>
</tr>
<tr>
<td>$r_c$</td>
<td>5.3 Ω</td>
<td>5.3 Ω</td>
<td>5.3 Ω</td>
</tr>
<tr>
<td>$r_e$</td>
<td>5.2 Ω</td>
<td>5.2 Ω</td>
<td>5.2 Ω</td>
</tr>
<tr>
<td>$r_π$</td>
<td>4 kΩ</td>
<td>2.14 kΩ</td>
<td>1.13 kΩ</td>
</tr>
<tr>
<td>$C_π$</td>
<td>110 fF</td>
<td>180 fF</td>
<td>330 fF</td>
</tr>
<tr>
<td>$C_μ$</td>
<td>6.8 fF</td>
<td>6.4 fF</td>
<td>5.5 fF</td>
</tr>
<tr>
<td>$g_m$</td>
<td>170 mS</td>
<td>330 mS</td>
<td>606 mS</td>
</tr>
</tbody>
</table>

Figure 4.3: The calculated frequency response from the ZVT analysis in solid lines and the ADS simulation and measurement at 4 mA/μm² current density in dashed lines.
4.1.3 Common-mode

A differential circuit’s resistance to amplify common-mode signals can be quantified as the CMRR and is defined as \(|S_{d21}/S_{c21}|\). This is an important FOM for a system’s perspective, to prevent dc amplification and even-order intermodulation products such as IM2 to propagate through the amplifier. The differential topology relies on virtual grounds rather than common ground as reference, and in this perspective, the nodes’ load impedances are of less importance because there is no current for the differential-mode signal. For the corresponding common-mode, the same node is on the contrary perceived as a virtual open circuit. For this reason, the nodes’ load impedances, including asymmetries or virtual ground displacements, reduce CMRR. The differential pair transistors \(Q_1, Q_2\) and \(Q_7 - Q_{14}\) rely on this principle and therefore, the loading of the common nodes play a significant role in the common-mode response. The cascode current source is a better current source than the C-E because of its higher output impedance and is implemented in \(Q_{22} - Q_{29}\). However, neither current source possesses desirable ideal open circuit characteristics. Parasitics contribute to lowering the output impedance as frequency increases, leading to deteriorated performance. The simulated small signal output admittance of the cascode \((Q_{23})\) and C-E \((Q_{20})\) current sources are shown in Fig. 4.4. From a practical perspective, the cascode current source requires higher bias voltage since there are additional C-B transistors following the C-E. For this reason, it was implemented where the voltage headroom was high enough, at the emitter-followers \(Q_7 - Q_{10}\). The measured CMRR of the VGA is plotted in Fig. 4.5 and as expected, because of the capacitive loading at the current sources, the CMRR decrease with frequency.

![Figure 4.4: Simulated output admittance of the cascode and C-E current sources.](image)

The bias rails consist of ground (M2) and a single -7 V bias supply. All internal dc leveling is handled by on-chip current sources and reference sources to ensure that all transistors are operated in the forward-active region. The common-mode input and output voltages are -3.5 V and -1.5 V respectively. Whether this is acceptable or not depends on the type of application and
interface, and on ac or dc coupling. Ac-coupling is always possible, but puts tough requirements on the external capacitors to cover a wide frequency range. Dc-coupling must meet the same common-mode input and output voltage levels, only then, the full frequency coverage of the VGA in [E] can be obtained.

4.2 RFIC technologies and 3-D BEOL layout

The circuit designs in Chapters 2 and 3 consist only of a few transistors each, four and eight transistors in [C] and [D] respectively and the PAs in [B] and [A] consist of 16 and 8 transistors with a total transistor gate width of 2.7 mm and 1.65 mm respectively. They contain many passive circuit elements in terms of matching elements, hybrids, power combiners and baluns in order to provide the desired function and high performance, occupying 3.75 mm², 3.2 mm², 6 mm² and 3.2 mm² die area, respectively. In contrast, the baseband VGA in [E] consists of 47 transistors and only a handful of passive reactive elements, in all, occupying 0.77 mm². The design methodology is therefore quite different between a baseband circuit and an RF circuit, which is also reflected in the chosen technology.

The BEOL of Teledyne’s 0.25 μm DHBT InP process [38] consists of five stacked metal layers embedded in a low permittivity dielectric material and a thin layer of Silicon Nitride between M1 and M2 to form Metal-Insulator-Metal (MIM) capacitors, shown in Fig. 4.6. Integrated in the process’s front-end-of-line (FEOL) are resistor and npn transistors.

The integration of 47 transistors into 200 × 200 μm² circuit area in the InP technology is feasible because of the multi-layer BEOL and small-sized, high transconductance transistors. The chip area, including probing pads occupies 800 × 960 μm², see Fig. 4.7a. In comparison, a baseband amplifier with 9 transistors (and 16 gate Schottky diodes) in WIN Semiconductor’s 0.1 μm pHEMT process [42] requires an active area of 600 × 400 μm², Fig. 4.7b. To achieve competitive NF, gain and linearity, large transistors are necessary. Moreover, a 1-metal layer BEOL make transistor interconnection difficult. Large-Scale-Integration (LSI) and Very-Large-Scale-Integration (VLSI) processes such as Teledyne’s 0.25 μm InP DHBT [38] and IHP’s SiGe 0.13 μm HBT (SG13S) [44] share a similar BEOL and FEOL, which allows large numbers of transistors to be connected in dense areas, utilizing all metal layers for short interconnects.
between transistors or to passive elements. For analog circuits, a few metal layers are perfectly sufficient and provide added value in terms of compact designs and performance.

In Fig. 4.8, the compact active core of the VGA exposes a jumble of interconnected wires, transistors, inductors, capacitors and TFRs. To minimize
the inductive effect of the narrow interconnecting wires and capacitive inter-
coupling with other wires, the wire lengths were kept as short as possible. A
square transistor dense layout in combination with the utilization of all metal
layers for interconnection, provided the best layout implementation for this
purpose.

A further advantage of a multi-layer BEOL is the ability to layout passive
components vertically. For inductors, this is beneficial because they can be
layouted in the form of a helix in the BEOL. The magnetic flux is stronger
and more concentrated through the helix core for the same current and there-
fore increases the inductance per unit length and unit area. In [114], different
layout configurations of the multi-layer inductor were investigated for induc-
tance and quality factor using a blend of layout parameters such as number
of vertical layers and number of horizontal layers. The helix layout achieves
higher quality factor but lower resonant frequency because of the higher ca-
pacitive intercoupling between the lines. The 300 pH inductors ($L_1$ and $L_2$ in
Fig. 4.1 and shown in the lower part of the photography in Fig. 4.8) use a helix
design for increasing the inductance and minimizing the area. In Fig. 4.9, the
simulated inductance and quality factor of the inductor are shown, including
the self resonance frequency at 205 GHz, far above the 3 dB bandwidth of the
VGA.
4.3 Main results

The VGA was characterized on wafer with 4-port SP up to 50 GHz for true differential and common-mode characterization. The 4-port SP were recalculated to differential and common-mode SP, which is a better representation of the performance of this differential circuit [45]. The measured (solid lines) and simulated (dashed lines) differential gain versus frequency with gain steps of approximately 10 dB is shown in Fig. 4.10a. Across the whole bandwidth, the gain control range measures 44 dB, and the minimum Noise Figure (NF) measures 6.2 dB, plotted in Fig. 4.10b. The maximum gain was measured to 31 dB and the 3 dB roll-off frequency to 40 GHz. The combination of these two parameters resulted in a GBP of 1.4 THz, which is the highest reported GBP of a Gilbert-cell based VGA.

![Figure 4.9: Simulated inductance (L) and quality factor (Q) of the 300 pH inductor L1 and L2](image)

![Figure 4.10: Measured and simulated gain and NF versus frequency and voltage control VGC.](image)

(a) Differential gain in steps of approximately 10 dB interval in the following order (◊, Δ, +, V, o, □). Maximum gain occurs at \( V_{GC} \) -3.3 V (◊).

(b) Measured and simulated gain and NF versus voltage control \( V_{GC} \) at 10 GHz.

In Fig. 4.11a, the constellation diagram of a 64-QAM signal at 6.5 Gsym/s is shown. The measurement results in this setup reached a throughput of
39 Gbit/s at 5.9% EVM. In this test the input signal level was 25 mV Vpp and the input EVM was 3.8%. On/off keying with PRBS-31 test pattern was used to test the transmission of the VGA using wide signal bandwidth. Again, the VGA was set to operate at maximum gain and the input signal level was 60 mV Vpp. In Fig. 4.11b, the eye-diagram is clear and open at 44 Gbps.

Figure 4.11: Measured response of modulated signals at max gain.
Chapter 5
Conclusions and future outlook

This thesis has addressed important challenges in the development of future mm-wave front-end electronics for mobile backhaul. Wideband operation, efficiency, linearity and high output power are subjects covered in frequency modulators, VGA and PAs in the pursuit for improved radio performance. Theory, realization and experiments have been practiced.

In [A], a combination of stacked HEMTs and parallel power combining was evaluated to simultaneously improve output power and efficiency. The stability of the stacked HEMT was addressed by introducing multiple via-holes and anchoring the gates at both edges, improved stability was obtained. This was supported by model verification. The PA design delivered 25 dBm output power and 15% efficiency at 61 GHz which is the highest output power per chip area in GaAs technologies. Simulations predicted a peak efficiency than 22%, much higher than measured. This loose end is worth following up. Stacked HEMTs are conceptually very attractive and further studies to understand the effects when operating close to $f_T/f_{max}$ would be an interesting line of inquiry. Theoretical analysis, topologies and transistor layouting are areas to further pioneer. Furthermore, it is expected that the output reflection close to $f_T/f_{max}$ of stacked HEMTs in deep class C are more reflective than common-source transistors. This opens up possibilities for load-modulation to improve efficiency at back-off rather than at peak power. This is uncharted territory and utterly rewarding from a system perspective.

Linearization of PAs in [B] described theory and circuit implementation for a class C APD. The realization of a joint linear APD and PA combination was successfully demonstrated under two-tone and modulated signal experiments. At 17 dBm average output power, spectral regrowth was reduced with APD. The APD also demonstrated tuning ability in the CI3 which is important for compensating for temperature, output power and frequency variations. The effects of frequency and temperature variations were not covered in [B] but the limited tuning range of the class C APD, suggests further studies on intermodulation generation. In this context, better polynomial selectivity and larger polynomial coverage are related research topics. The ultimate system would be self-sustained, adapting polynomial parameters for a linear response.
Two linear balanced quadrature modulators for the E-band and D-band communication windows were designed and verified in \([C]\) and \([D]\) both with state-of-the-art performance. In \([D]\) a novel topology incorporating a conversion-efficient second-harmonic mixer cell to isolate the important second LO harmonic from the RF port, enables this component to be used in transmitters. High LO-power in resistive mixers is advantageous for efficient and linear frequency conversion which requires high LO to RF isolation. Innovative phase distribution networks, including differential branch-line couplers, common-mode rejection filters and high CMRR Marchand baluns were realized in the circuit designs for this purpose. Conversion loss of 11 dB, 2 × LO to RF isolation better than 30 dB (70 dB with dc injection) and linear output power up to -2 dBm were measured at the D-band. At the E-band, conversion loss of 11 dB, LO-RF isolation > 30 dB and OIP3 of 13 dBm were measured. Linear mixers by both topology and technology are interesting research areas with promising prospects. Graphene and GaN are potential technologies to support higher linearity. Stacked FETs is unexplored territory for resistive mixers that is obliged to be researched.

Classical circuit topologies such as the Gilbert cell and the cascode amplifier can, with compact layouting and high speed semiconductor processes, provide high bandwidths. A vertical 3-D BEOL allowed the use of short interconnects and small-sized high inductive coils in the design. In \([E]\) exploiting the full capabilities of the process, the VGA demonstrated state-of-the-art performance. By analyzing the dominant pole for the complete circuit by means of ZVT constants, a good estimation of the 3 dB bandwidth was achieved. Furthermore, the dominant time constant of the circuit was revealed and, allowed focused efforts treat this critical part of the design to further increase the GBP. State-of-the-art results were achieved, dc-40 GHz bandwidth and 44 dB gain control range. VGAs are critical components in a link budget and, in the context of future wideband communications, system noise will be more prominent, and increasingly more important to budget for. Topologies with large gain control range and generating low noise figure at all settings are an important area of research, both on the receiver and transmitter side.
I would like to thank the people who has made this possible:

To my examiner and supervisor Prof. Herbert Zirath for giving me the opportunity to pursue a Ph.D., and for all the technical and rewarding discussions and encouragement.

To my supervisor Sten E. Gunnarsson and co-supervisor Mattias Ferndahl for your assistance, loyal support, technical discussions, friendship and for valuable guidance.

To Iltcho Angelov for broadening the context of this work and his enthusiasm and desire to teach.

I also want to thank my colleagues and friends over the years; and all the people at MEL and Gotmic for making them great and inspiring places to work at. It has been a privilege to be a PhD student and a great time of my life.

I express my deepest gratitude to my lovely wife Helena and our beloved Alfonso. To my family and friends for all the support and love I get from them.

Finally, I would like to acknowledge the financing institutions and collaboration partners which made this work possible. This work has been supported by “Vinn Verifiering” which is funded by Swedish Governmental Agency of Innovation Systems (VINNOVA) and “Lösningar för trådlös kommunikation med hög datatakt” which is funded by Swedish Foundation for Strategic Research (SSF).
Bibliography


[16] “The European table of frequency allocations and applications in the frequency range 8.3 kHz to 3000 GHz (ECA TABLE),” CEPT Electronic Communications Committee (ECC), Tech. Rep. ERC REPORT 25.


A V-band Stacked HEMT Power Amplifier with 25 dBm Saturated Output Power in 0.1 µm InGaAs Technology
Marcus Gavell, Iltnco Angelov, Mattias Ferndahl, Herbert Zirath

Paper B

An E-band Analog Pre-distorter and Power Amplifier MMIC chipset

Marcus Gavell, Göran Granström, Christian Fager, Sten E. Gunnarsson, Mattias Ferndahl, Herbert Zirath

*Microwave and Wireless Components Letters*, vol. 28, nr. 1, pp. 31-33, January 2018.
A linear 70-95 GHz differential IQ modulator for E-band Wireless Communication

Marcus Gavell, Herbert Zirath, Mattias Ferndahl, Sten E. Gunnarsson

Paper D

A D-band second harmonic quadrature modulator MMIC with harmonically isolated RF port

Marcus Gavell, Sten E. Gunnarsson, Mattias Ferndahl, Herbert Zirath

submitted to IEEE Transactions on Microwave Theory and Techniques
Paper E

Design and analysis of a wideband Gilbert cell VGA in 0.25 µm InP DHBT Technology with DC-40 GHz frequency response

Marcus Gavell, Sten E. Gunnarsson, Iltcho Angelov, Zhongxia Simon He, Mattias Ferndahl, Herbert Zirath