Integrated millimeter Wave CMOS Power Amplifiers for 5G Systems

Elgaard, Christian

2024

Document Version:
Publisher's PDF, also known as Version of record

Integrated millimeter Wave CMOS Power Amplifiers for 5G Systems

Christian Elgaard

LUND INSTITUTE OF TECHNOLOGY
Lund University

Doctoral Dissertation
Lund, April 2024

Department for Electrical and Information Technology
Lund University
P.O. Box 118
SE-221 00 LUND
SWEDEN

ISSN 1654-790X-172
ISBN 978-91-8039-944-9 (print)
ISBN 978-91-8039-945-6 (electronic)
Series of licentiate and doctoral dissertations.

Produced using LaTeX Documentation System.
Printed in Sweden by *Tryckeriet i E-huset*, Lund.
April 2024.

Abstract

The anticipated continuation of the last three decades of growth of global demand for high-speed high-coverage mobile data, calls for massive investments in cellular infrastructure, for the ongoing roll out of the fifth generation of mobile systems, but also later for the sixth generation. A key enabler, for the high-speed part, is for the cellular system to move up in frequency and support communication at millimeter-wave (mmW) frequencies (24 – 70 GHz), where available wideband spectrum exist. One of the most challenging building blocks for communication at mmW frequencies is the power amplifier, which has the function to amplify the transmission signal before feeding it to the antenna. The power amplifier should ideally provide high output power, without distorting the signal, while consuming as little power as possible. These requirements stand in stark contrast with each other and is particularly troublesome for high-speed communication at high frequencies. This dissertation is about mmW power amplifiers and it starts with introductory chapters, placing the power amplifier in its context and presenting theory about power amplifiers in general. Then follows a brief summary of the scientific contribution with conclusions and some suggestions for future work, of the four papers, which are the foundation of the dissertation.

Paper I presents a CMOS mmW power amplifier and pre-power amplifier, with aim for integration in an antenna array system. To increase the output power, while still reducing the maximum needed supply voltage, the circuit utilizes a "two way" output combiner prior to the load. The PA, measured using continuous wave signals, reached, at the time, state-of-the-art performance for both saturated output power and 1 dB compression point, combined with low AM-PM distortion below the compression point.

Paper II describes a CMOS mmW Transmit/Receive (TRX)-switch, power amplifier, and pre-power amplifier, targeted for integration in an antenna array system. To linearise the output signal the PA input transistors gate bias is adjusted based on the input signal level, i.e. it uses adaptive bias. The TRX-switch provides a downward 1:2 impedance transformation in TX-mode to boost reachable output power, and in RX-mode it provides an upward impedance transformation of 2:1 for optimal noise figure. The adaptive bias brings significant improvement of both saturated output power and 1 dB output compression point, and simulations for the TRX-switch show low insertion loss in both TX and RX mode.

In **Paper III** a CMOS mmW transceiver front-end including a novel TRX-switch is presented. For high efficiency the transmitter is equipped with a Doherty PA, which uses a high bandwidth adaptive bias circuit to reduce the fundamental nonlinearity associated with Doherty PAs. In addition, an innovative method is implemented that breaks the fundamental bandwidth limi-

tation for the input signal of Doherty PAs. The transceiver was extensively measured in both transmit and receive mode. In transmit mode, continuous wave as well as OFDM-modulated measurements were performed. State-of-the-art output power and efficiency for integrated transceivers for high bandwidth OFDM-modulated signals were demonstrated, and for the receiver state-of-the-art noise figure was achieved when compared in the same category. Significant improvements on ACLR and EVM when using the adaptive bias for wideband modulated signals were demonstrated. Furthermore, excellent image rejection ratio and LO leakage suppression were measured.

Paper IV derives fundamental equations related to a Doherty amplifier, using a simplified transistor model suitable for hand calculations, and thus the fundamental nonlinearity of the Doherty amplifier is explained and investigated. Furthermore, the paper analyses the use of adaptive bias, which offers the possibility to mitigate the fundamental nonlinearity as explained by the theory. To verify the theoretical predictions, the design and measurements of the adaptive bias circuit tailored for high PAR high bandwidth modulated signals, for the mmW Doherty amplifier in paper III, are presented in detail. Controllability of the adaptive bias circuit, which is needed to fully benefit from using adaptive bias, was measured using continuous wave tone stimuli. Multiple measurements with wideband OFDM-modulated signals were also conducted which largely verified the predictions by the theory. For increased reliability the measurements were repeated using two different samples.

Populärvetenskaplig sammanfattning

I takt med att vi människor använder våra mobiltelefoner mer och mer och att vi samtidigt efterfrågar en allt högre datahastighet måste även mobilnätverken, som mobiltelefonerna kommunicerar med, anpassas för att klara av att leverera och ta emot den ökande datahastigheten och datamängden. Utvecklingen av de mobila nätverken och mobiltelefonerna går mycket snabbt och både datatrafiken och den genomsnittliga datahastigheten har fördubblats de senaste tre åren, och förväntningarna är att denna utveckling kommer att fortsätta.Varje mobiltelefon har både en sändare och en mottagare och det samma gäller basstationerna i mobilnätverken. I sändaren, både i basstationen och i telefonen sitter det, något förenklat, en effektförstärkare precis innan antennen. Effektförstärkaren i sändaren, som är huvudfokus för den här avhandlingen, har som uppgift att förstärka radiosignalen som skall skickas ut av antennen, så att signalen kan tas emot även om mottagaren befinner sig långt bort.

Ett sätt att kunna skicka mycket data snabbt är att använda en hög bandbredd. Det är tyvärr väldigt begränsat med tillgänglig bandbredd, men vid mycket höga frekvenser finns det fortfarande tillgängligt. Tyvärr finns det många svårigheter med att skicka trådlös information vid höga frekvenser. Det kanske mest problematiska är att eftersom antennstorleken minskar med ökad frekvens så kommer även mottagen signalstyrka att minska. Det är även svårare för signalen att penetrera byggnader, regn eller till och med löven på ett träd och elektroniken som sänder och tar emot signalerna fungerar generellt sämre. Frekvensområdet som ligger vid ca 30 GHz till 300 GHz kallas för millimetervågor eftersom våglängden för elektromagnetiska vågor i det området är 1 till 10 mm. För att kompensera för den stora signalförlusten mellan basstation och mobiltelefon vid millimetervågskommunikation använder man många antenner, flera hundra eller kanske till och med tusentals så kallade antennelement, vilket gör att man kan öka den mottagna signalstyrkan väldigt mycket

Den här avhandlingen handlar om effektförstärkare för millimetervågor konstruerade i en vanlig och billig kiselbaserad teknik som kallas för CMOS, för femte generationens mobilsystem, men även för kommande generationer. I avhandlingen ingår fyra vetenskapliga publikationer baserade på tre uppmätta effektförstärkare med gradvis ökande komplexitet där den tredje även innefattar i princip en hel sändare och även delar av en mottagare. Den tredje kretsen är uppmätt med en signal där effektförstärkaren skickar hela 9.6 Gbit/s, vilket är en väldigt hög datahastighet. För att exemplifiera skulle det motsvara att ladda ner cirka en hel timmes video på bara en sekund. För att klara av att skicka en signal med hög uteffekt och med så hög datatakt, utan förvränga den så att det blir svårt eller till och med omöjligt för mottagaren att avkoda

de digitala bitarna, och samtidigt förbruka så lite ström som möjligt, används en Dohertyförstärkare kombinerat med adaptiv bias. Dohertyförstärkaren är en välkänd förstärkartyp som uppfanns redan 1936, och den har den speciella egenskapen att den, jämfört med andra förstärkartyper, förbrukar väldigt lite ström när den förstärker signaler med kraftigt varierande amplitud, vilket är en en egenskap som signalerna i 5G (och 6G) har. För att spara ström har Dohertyförstärkaren två förstärkare som samarbetar. En av dessa är på hela tiden (huvudförstärkaren), medan den andra (hjälpförstärkaren) bara är på vid höga amplituder, vilket minskar förbrukningen. Det är dock en svår utmaning att konstruera en Dohertyförstärkare vid så höga frekvenser som millimetervågor vilket gör det intressant ur ett forskningsperspektiv. För att transistorerna skall förstärka signalen på önskvärt sätt behöver man ställa in en lämplig arbetspunkt, vilket kallas för att man förspänner transistorerna. Normalt använder man en konstant arbetspunkt, men vid adaptiv bias justerar man istället transistorernas arbetspunkter i takt med att signalen ändrar sin amplitud. Som framgått av forskningen i avhandlingen är Doherty förstärkare med adaptiv bias för hjälpförstärkaren ett bra sätt att minska de problem som uppstår när en effektförstärkare på ett strömsnålt sätt skall hantera de komplexa signalerna i 5G och framtidens mobilsystem. I den fjärde artikeln undersöks och förklaras de teoretiska aspekterna för hur en adaptiv bias signal skall utformas för att fungera optimalt ihop med hjälpförstärkaren, och artikeln innehåller även en detaljerad beskrivning av konstruktionen av en krets som klarar av skapa en sådan adaptiv bias signal och därmed kan ändra arbetspunkten för hjälpförstärkaren mycket snabbt, dvs. med hög bandbredd. Kretsen skapar adaptiv bias signalen genom att först extrahera amplitudinformationen ifrån Dohertyförstärkarens insignal. Ett viktigt teoretiskt resultat är att en ideal adaptiv bias signal sedan skall konstrueras genom en olinjär överföringsfunktion ifrån amplitudinformationen. Mätningar och simuleringar visar att kretsen klarar av att göra detta på ett effektivt sätt.

Contents

Preface

This dissertation concludes my academic contribution for a Ph.D degree in Electrical Engineering at Lund University, Sweden. The Ph.D studies were performed in the form of an industrial Ph.D in close cooperation with the section of RF Frontend and Power Amplifier Design at Ericsson Research in Lund, Sweden. The dissertation consists of two parts, the first contains a general overview and introduction to the field of power amplifier design, and the second part consists of the published research papers, listed below.

Included Research Papers

The contribution to the research field is summarized by the following publications:

Paper I C. Elgaard, A. Axholt, E. Westesson and H. Sjöland, A 26 GHz 22.2 dBm Variable Gain Power Amplifier in 28 nm FD-SOI CMOS for 5G Antenna Arrays, 2018 Asia-Pacific Microwave Conference (APMC), Kyoto, Japan, 2018, pp. 965-967, doi: 10.23919/APMC.2018.8617416. [1]

Paper II C. Elgaard, S. Andersson, P. Caputa, E. Westesson and H. Sjöland, A 27 GHz Adaptive Bias Variable Gain Power Amplifier and T/R Switch in 22nm FD-SOI CMOS for 5G Antenna Arrays, 2019 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), Boston, MA, USA, 2019, pp. 303-306, doi: 10.1109/RFIC.2019.8701819. [2]

Paper III C. Elgaard, M. Özen, E. Westesson, A. Mahmoud, F. Torres, S. Bint Reyaz, T. Forsberg, R. Akbar, H. Hagberg and H. Sjöland, Efficient Wideband mmW Transceiver Front-End for 5G Base Stations in 22 nm FD-SOI CMOS, IEEE Journal of Solid State Circuits (JSSC), pp. 1-16, 2023. [3]

Paper IV C. Elgaard and H. Sjöland, Analysis and Design of a GHz Bandwidth Adaptive Bias Circuit for an mmW Doherty Amplifier, IEEE Transactions on Microwave Theory and Techniques, Regular Papers, Submitted Feb. 2024. [4]

Related publications

I have also authored and co-authored two papers, listed below, that are not considered a part of this dissertation.

Paper V C. Elgaard and L. Sundström, A 491.52 MHz 840 uW Crystal Oscillator in 28 nm FD-SOI CMOS for 5G Applications, ESSCIRC 2017 - 43rd IEEE European Solid State Circuits Conference, Leuven, Belgium, 2017, pp. 247-250, doi: 10.1109/ESSCIRC.2017.8094572. [5]

Paper VI S. Ek, T. Påhlsson, C. Elgaard, A. Carlsson, A. Axholt, A. Stenman, L. Sundström, and H. Sjöland, A 28-nm FD-SOI 115-fs Jitter PLL-Based LO System for 24–30-GHz Sliding-IF 5G Transceivers, in IEEE Journal of Solid-State Circuits, vol. 53, no. 7, pp. 1988-2000, July 2018, doi: 10.1109/JSSC.2018.2820149. [6]

Patent applications

1. C. Elgaard "Oscillator Circuit" US2016308491A1 US9692354B2 2013-12-16
2. J. Peltonen, C. Elgaard, and A. Stenman "Oscillator Circuit with Bias Current Generator" US10333526B2 US2017237441A1 2014-06-02
3. C. Elgaard and L. Sundström "Low-Noise Oscillator Amplitude Regulator" HK1250847A1 2015-06-05
4. C. Elgaard, A. Carlsson, and H. Sjöland "A Reference Signal for Noise Reduction" WO2017162275A1 2016-03-22
5. C. Elgaard, S. Andersson, A. Axholt, and I. Ud Din "Power Amplifier" US10848108B2 US2019245493A1 2016-10-14
6. H. Sjöland, A. Axholt, and C. Elgaard "An Integrated Transformer" WO2018188719A1 2017-04-10
7. L. Sundström, S. Ek, and C. Elgaard "Transceiver Circuit" US11095335B2 US2020145041A1 2017-06-22
8. C. Elgaard, M. Åström, and F. Tillman "Integrated Circuit With Clock Distribution" WO2019029819A1 2017-08-11
9. C. Elgaard, S. Andersson, and L. Sundström "Power Amplifier Arrangement" US11258413B2 US2020389136A1 2018-01-17
10. C. Elgaard and H. Sjöland "Power Amplifier Circuit" WO2020126054A1 2018-12-21
11. C. Elgaard and H. Sjöland "Bias Circuit and Power Amplifier Circuit" EP3909127A1 2019-01-10
12. M. Hamid, C. Elgaard, A. Al-Qamaji, and F. Abdalrahman "Interleaved Antenna Array Configuration in a Radio Node" US2023155290A1 2020-04-09
13. F. Torres, C. Elgaard, S. Andersson, and H. Sjöland "Transmitter Arrangement, Transceiver, Radio Communication System and Method" WO2021209151A1 2020-04-17

14. C. Elgaard, S. Mattisson, and S. Andersson "Envelope Controlled Biasing of an Auxiliary Transmitter of a Doherty Power Amplifier" WO2021228396A1 2020-05-14

15. C. Elgaard and H. Sjöland "Transmitter Circuit" WO2021259448A1 2020-06-22

16. I. Ud Din, C. Elgaard, and S. Andersson "A Transceiver Switch Circuitry" WO2022069037A1 2020-09-30

17. I. Ud Din, C. Elgaard, and S. Andersson "A Transceiver Switch Circuitry" WO2022069038A1 2020-09-30

18. I. Ud Din and C. Elgaard "A Differential Transceiver Switch Circuitry" WO2022111814A1 2020-11-27

19. C. Elgaard and H. Sjöland "Mitigation of Memory Effects in a Power Amplifier" WO2023117083A1 2021-12-22

20. C. Elgaard and H. Sjöland "Digital Phase Locked Loop Amplitude Modulation-to-Phase Modulation (AM-PM) Compensation" WO2023126054A1 2021-12-28

21. C. Elgaard, H. Sjöland, and M. Hamid "Concurrent Digital Post-Distortion of Multiple Signals" WO2023160783A1 2022-02-24

22. C. Elgaard, R. Kasri, F. Mesquita, and H. Sjöland "Receiver Beamforming" WO2023208362A1 2022-04-29

23. C. Elgaard, H. Sjöland, and M. Åström "Co-Located Back-Lobe Cross Link Interference (CLI) Digital Canceler Interface" WO2023217680A1 2022-05-09

24. C. Elgaard, H. Sjöland, I. Ud Din, and H. Hagberg "Antenna Array Power Amplifier Mismatch Mitigation" WO2023241806A1 2022-06-16

25. H. Sjöland and C. Elgaard "Dual Polarity Power Amplifier With Nonlinear Supply Modulation" WO2024008306A1 2022-07-08

Acknowledgments

Ever since I joined Ericsson Research I have felt a sense of pride to be part of such an amazing highly talented group. One of my first assignments after joining the group was to design a high frequency crystal oscillator and after completing the design my colleague **Lars Sundström** said "if it works we can write a paper" and since it worked we wrote a paper. So, thanks Lars for directing my in a way that came to become the first spark of this dissertation. Somewhere around this time I realized that maybe there is a possibility for me to formalize the research work I do at Ericsson in collaboration with academia. Sometimes later I asked another colleague, Professor **Henrik Sjöland** if he and the university would be willing to formally take me on as a PhD student. After some internal discussions at Ericsson I had a final meeting about the matter with **Fredrik Tillman** and **Sven Mattisson** which ended with Sven saying the words "Then I will be invited to a dissertation party". I am sincerely grateful for given the rare opportunity to pursue an industrial PhD at Lund University while still being able to continue and grow in my research position in the Integrated Radio System RF frontend and PA group at Ericsson Research. Thanks Fredrik and Sven for making this happen. After some time I got a new manager, **Stefan Andersson** who has been kind enough to turn a blind eye to the fact that the somewhat optimistic plan, which stated that my PhD studies should be completed in four years with 50 % allocation (and 50 % on other Ericsson stuff) was long overdue. Thanks a lot Stefan for, as far as possible, allocating me on projects which allowed me to advance my dissertation. Through this whole time, I would like to thank my closest team members which were, and still are, always very supportive, no matter what the problem was. Nevertheless, I want to mention **Imad Ud Din** who has been my companion on this journey and especially thanks for the good times we had completing those early PhD courses.

However, as I embarked on my endeavor, **Henrik Sjöland** became not only my colleague but also my supervisor during my PhD studies. For this I'm very grateful and I cannot think of a better choice of supervisor. Henrik, with extremely impressive hands-on knowledge, on all kind of matters that relate to our field, is always willing to help and spend the time to do it, even though it means sacrificing his own time during the weekend. There is a long list of truly impressive qualities and behaviours that could be listed.

In the background to the research conducted at Ericsson there was always support from family and friends. I would like to start by expressing gratitude to **Johan Löfgren** who inspired me to switch from finance/economics to physics by simply saying "Engineering Physics is considered to be the hardest to study"- back in 1996, upon which I immediately started preparing for getting accepted to LTH. I also want to thank **Pierre**, **Henrik**, and **Magnus** for being

the best possible friends. Pierre who never would hesitate to any request for support, Henrik for those daily small adventures, such as discussing polarized light or group delays while sailing, and Magnus for those never-ending deep discussions late at night and racket sport matches.

I would also like to thank, my father **Bent**, mother **Majvi**, and sister **Jenny**, with whom I grew up. I think that at this time my interest for natural science was developed that lead to my choices later in life. When deciding to pursue the PhD I was already a father of four wonderful children **Oscar, August, Molly**, and **Elliot** together with my fantastic, always generous and supportive fiancé **Maria**. My goal was to absorb the time needed for the PhD studies during office hours to minimize the absence from my family. I hope that that has also been the case, at least for most of the time, even though there has been some late evenings towards the end. Anyhow, to my family, I want you to know that I love you and thank you for your support. When having a bad day at the office I know that you are always there when I come home.

List of Acronyms

AAS	Antenna Array System
AC	Alternating Current
ACLR	Adjacent Channel Leakage Ratio
ACPR	Adjacent Channel Power Ratio
ADC	Analog-to-Digital Converter
AM-AM	Amplitude Modulation to Amplitude Modulation
AM-PM	Amplitude Modulation to Phase Modulation
BB	Baseband
BPF	Band Pass Filter
BW	Bandwidth
CMOS	Complementary Metal Oxide Semiconductor
CW	Continuous Wave
DAC	Digital-to-Analog Converter
DE	Drain Efficiency
DC	Direct Current
DCO	Digital Controlled Oscillator
DPLL	Digital Phase Locked Loop
EIRP	Effective Isotropic Radiated Power
EV	Error Vector
EVM	Error Vector Magnitude
DIBL	Drain Induced Barrier Lowering
DPD	Digital PreDistortion
EVM	Error Vector Magnitude
FD-SOI	Fully Depleted Silicon on Insulator
FSPL	Free Space Path Loss
IBW	Instantaneous Bandwidth
IF	Intermediate Frequency
IIPn	n-th order Input-Referred Intercept Point

IMn	n-th order Intermodulation distortion
IPn	n-th order Intercept Point
IRR	Image Reject Ratio
LNA	Low-Noise Amplifier
LO	Local Oscillator
LPF	Low Pass Filter
LTI	Linear Time-Invariant
LTV	Linear Time-Variant
mmW	millimeter Wave
NF	Noise Figure
NMOS	n-channel metal-oxide-semiconductor field-effect transistor
OCP	Output-referred Compression Point
$N^{th}G$	N^{th}-Generation Technology for Cellular Networks System
OFDM	Orthogonal Frequency Division Multiplexing
PA	Power Amplifier
PAE	Power Added Efficiency
PAR	Peak-to-Average Ratio
PCB	Printed Circuit Board
PLL	Phase Locked Loop
PMOS	p-channel metal-oxide-semiconductor field-effect transistor
PPA	Pre Power Amplifier
PSD	Power Spectral Density
PWM	Pulse Width Modulation
RMS	Root Mean Square
RF	Radio Frequency
RX	Receiver
SCS	Signal Component Separator
SMPA	Switch Mode Power Amplifier
SSBW	Small Signal Bandwidth
SSG	Small Signal Gain

TDD	Time Division Duplex
TRP	Total Radiated Power
TRX	Transceiver
TTD	True Time Delay
ZVDS	Zero Voltage Derivative Switching
UE	User Equipment
ZVS	Zero Voltage Switching
QAM	Quadrature Amplitude Modulation

List of Symbols

A	Amplitude in polar coordinate system
A_I	Amplitude of (fundamental) current [A]
B'_s	Normalized shunt susceptance [S]
C_{gs}	Gate to source capacitance [F]
C_{in}	Input capacitance [F]
C_{out}	Output capacitance [F]
f_T	Cut-off frequency [Hz]
g_m	Transconductance [S]
HD_n	n^{th} order harmonic distortion
h_n	n^{th} order convolution term
$h(s)$	Transfer function
I_{dc}	Direct current [A]
i_{DS}	Drain to source current [A]
$i_{D,max}$	Maximum drain current [A]
I_{fund}	Current at fundamental harmonic [A]
IM_n	n^{th} order intermodulation distortion
IIP_n	Input-referred n^{th} order intercept point
IP_{ndB}	Input-referred n dB compression point
i_S	Switch current [A]
k_n	n^{th} order power series expansion coefficient
$Nbr_{transistors}$	Number of transistors
NF	Noise figure [dB]
OIP_n	Output-referred n^{th} order intercept point
OP_{ndB}	Output-referred n dB compression point
P_{DC}	Power from DC source [W]
P_{diss}	Dissipated power [W]
P_{EV}	Power of error vector [W]
P_{fund}	Power at fundamental harmonic [W]

P_{in}	Input power [W]
P_{loss}	Power of losses [W]
P_N	Normalized power output capability
P_{out}	Output power [W]
P_{Ref}	Power of reference vector [W]
$P_{R_{Load}}$	Power dissipated in R_{Load} [W]
$P_{R_{ON}}$	Power dissipated in on resistance [W]
P_{sat}	Saturated output power [W]
R_{Load}	Load resistance [Ω]
R_{on}	On resistance [Ω]
T	Signal period [s]
V_{dd}	Supply voltage [V]
v_{DS}	Drain to source voltage [V]
$v_{DS,max}$	Maximum drain to source voltage [V]
V_{in}	Input voltage [V]
V_{out}	Output voltage [V]
V_{pp}	Peak to peak voltage [V]
V_{th}	MOS transistor threshold voltage [V]
Z_{aux}	Output impedance of auxiliary amplifier [Ω]
Z_{load}	Load impedance [Ω]
Z_{main}	Output impedance of main amplifier [Ω]
$Z_{\lambda/4}$	Characteristic impedance of $\lambda/4$-transmission line [Ω]
Q	Quality factor
γ	Ratio of transconductance for main and auxiliary amplifier
ε	Fractional gain mismatch
η	Efficiency [%]
θ	Phase mismatch, outphasing angle, or polar phase [rad]
λ	Wavelength of Electromagnetic Wave [m]
ϕ	Common mode phase [rad]
2Φ	Conduction angle [rad]

Introduction

Chapter 1

Introduction

1.1 Motivation

Figure 1 shows recent years development of global mobile network data traffic and global subscriptions by technology and predictions of how they will evolve to 2028, and Fig. 2 shows the average mobile data rate and how it has increased in recent years [7, 8].

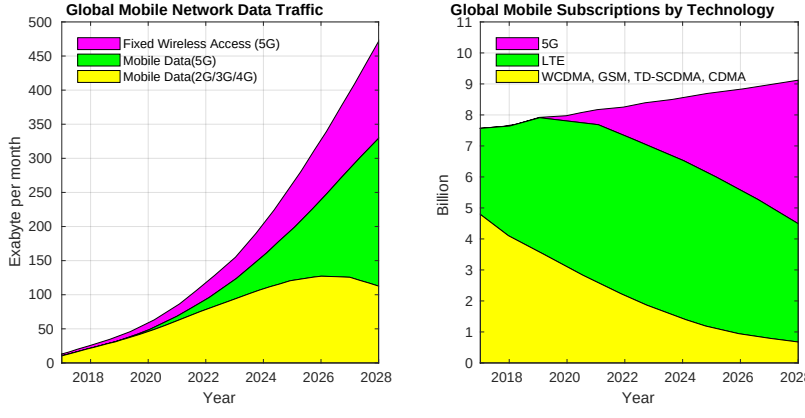

Figure 1: Left: Global mobile network data traffic per month 2017-2023 and predictions 2023 - 2028. Right: Global number of mobile subscribers 2017 - 2023 and predictions 2023 - 2028

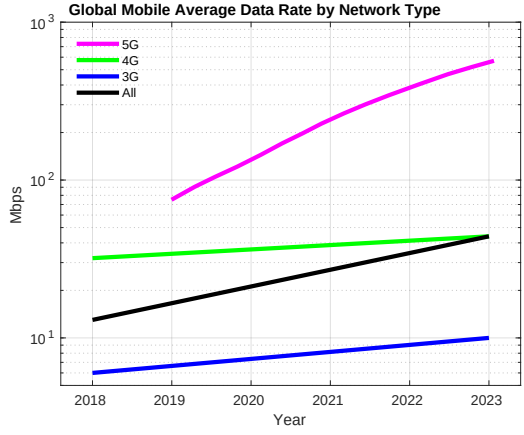

Figure 2: Global mobile average data rate by network type

The tremendous anticipated growth in mobile data traffic and data rate calls for massive research, development, and investments in cellular network infrastructure.

In 1948 Shannon formulated his famous equation for the capability to transfer information with an electronic signal [9]:

$$C = B \cdot \log_2 \left(1 + \frac{S}{N_t}\right) \tag{1}$$

Where C is the channel capacity in bits per second, B is the bandwidth (BW) of the channel in Hz, and S and N_t are the power of the signal and the total noise including interference, respectively. Shannons channel capacity equation combined with the anticipated mobile network data traffic growth and increased user data rates pinpoint the motivation for this dissertation. To increase the BW of the communication channel, to satisfy the need for more network data traffic and higher user data rates, higher and higher carrier frequencies are constantly being explored. The main driver is that at higher carrier frequencies there is more available spectrum. The increasing demand for data traffic and data rates, with a chase for more available bandwidth, has been a key driving force within the telecommunications industry for many years. This was thus the case at the beginning of my PhD studies, when exploration of millimetre wave (mmW) frequencies (about 24-70 GHz) for 5G cellular communication was still quite new, and it is true now when writing this dissertation as sub-THz frequencies are more frequently being discussed for high BW 6G cellular communications.

However, reaching more available spectrum by increasing the carrier frequency comes at a high cost. Firstly, at higher frequencies the attenuation

of the signal power from the transmit to receive antenna becomes much more severe. This was first captured by H. T. Friis in 1946 by his transmission formula [10], which can be used to derive what is commonly known as the Friis free space path loss (FSPL) formula. If assuming isotropic transmit and receive antennas, i.e. antennas with no directivity, the FSPL becomes:

$$FSPL = \left(\frac{4\pi d}{\lambda}\right)^2 \tag{2}$$

where d is the distance between the antennas, and λ is the electromagnetic wavelength. Friis path loss formula dictates that when the frequency increases tenfold, e.g. from 3 GHz to 30 GHz, so that λ decreases tenfold, the path loss increases a hundredfold.

Secondly, in 1965 Johnson concluded that the optimal trade-off between the maximum usable frequency, the cut-off frequency (f_T), and the maximum voltage (V_m), for a transistor, can be expressed as [11]:

$$V_m \cdot f_T = \frac{E \cdot v_s}{2\pi} \tag{3}$$

Where E and v_s are material constants that depend on the used semiconductor material. For silicon $E \cdot v_s/2\pi = 2 \cdot 10^{11} V \cdot Hz$, which gives a maximum voltage of 1 V for an f_T of 200 GHz. This is particularly troublesome for power amplifiers (PAs) operating at high frequencies, since to increase f_T for a transistor in a given semiconductor material, to boost gain and efficiency, one must reduce the voltage levels in the device, which limits the output power that the PA can deliver.

To summarize the discussion so far; the increased mobile network data traffic and increased user data rates call for more available spectrum, which is found at higher frequencies. Higher frequencies, however, gives significantly more path loss, and the analog circuits will, for a given semiconductor material, have lower gain and deliver a lower maximum output power.

To overcome the aforementioned obstacles, mmW cellular communication relies on using antenna arrays that increases transmit output power and receive and transmit antenna gain. Many antennas are then used, arranged in rows and columns, often with a spacing of half a wavelength, which at 30 GHz becomes just 5 mm. On the base station side the antenna arrays can thus be rather large when it comes to number of antennas, in the order of a hundreds or even thousands. Even the user equipment (UE), most commonly a mobile phone, will typically be equipped with a small antenna array. A large antenna array is often referred to as an antenna array system (AAS). Assuming a separate PA to be used for each antenna element, the total radiated power (TRP) from an AAS increases linearly with the number of antenna elements, which relaxes the output power requirement on each PA. Furthermore, and perhaps even more important, if the individual antenna signals are phase shifted to produce

constructive interference in a desired direction, the antenna gain also increases linearly in that direction, effectively producing antenna directivity. The antenna directivity, accomplished by phase shifting the signals to the individual antennas, is called beamforming gain and by adjusting the phase gradient over the array it is possible to steer the beam. The basic principle of beamsteering in an AAS is illustrated in Fig. 3.

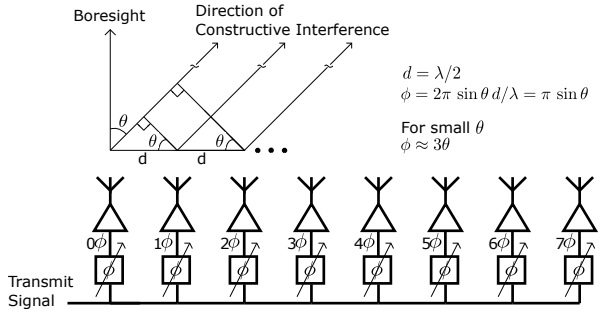

Figure 3: Principle of beamsteering for an 8x1 antenna array.

The effective isotropic radiated power (EIRP), which is the how much power an isotropic antenna would have to radiate to give equal radiated power in the direction of the beam from the AAS, will increase quadratically with the number of antennas in an AAS. Firstly, it increases as more PAs provide increased radiated output power. Secondly, it increases as the beamforming gain increases as more antenna elements are added. Since the AAS is rather small and contains many antennas, each connected to a transceiver (TRX) or at least a front-end of a transceiver, not much space is available for the electronics that form each TRX. This necessitates a high level of integration both due to the limited space, but also to reduce the overall cost as the number of transceivers becomes large. Both integration level and cost push in a direction towards CMOS technologies, which have excellent integration level and physical size, as well as low cost per unit in large scale production. However, it is challenging to integrate the PA in a CMOS TRX, due to multiple reasons, of which the most severe ones are its high output power requirements combined with the CMOS poor ability to withstand high device voltages and the heat dissipation from multiple PAs in a single integrated circuit. Figure 4 shows an example of a highly integrated mmW AAS, with the backside shown to the left and the front side to the right. The frontside has 8x8 dual polarized antennas, requiring in total 128 transceivers. The backside shows four integrated circuits, each with 32 transceivers.

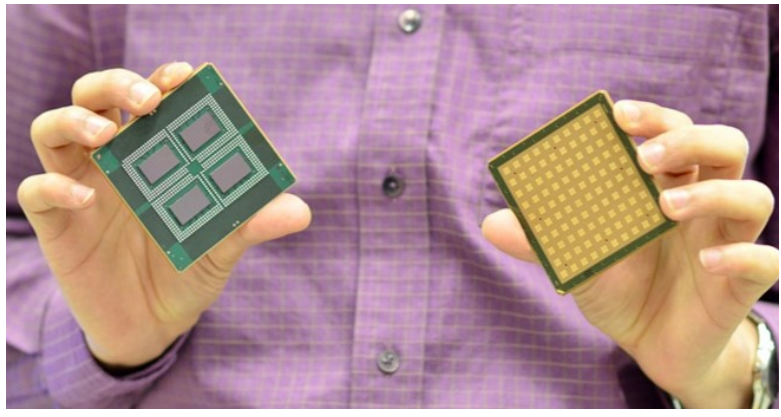

Figure 4: A highly integrated mmW AAS. Left: backside, four dies with 32 transceivers on each die. Right: frontside, 8x8 dual polarized mmW AAS, including an outer ring of dummy antennas. Picture from [12]. IBM/Ericsson silicon-based mmWave phased array antenna module operating at 28 GHz.

To summarize; AASs with beamforming are used to mitigate the effects of increased path loss and reduced output power at mmW. One major challenge then becomes how to accomplish the necessary integration of the PA into a CMOS chip containing multiple transceivers, due to the limited space in the AAS and the required reduced cost per transceiver. To meet targets in range and data rates of the transmission, the PA must be able to deliver high enough output power to the antenna, and it must perform the amplification in a power efficient way not to dissipate too much heat and to reduce power consumption in general. Furthermore, the PA must amplify the transmit signal while adding low level of distortion, also for wideband signals with high data rate modulation, which stands in stark contrast to the efficiency requirement, and it must allocate a relatively small die area. All things considered, the PA suitable for cellular mmW communication, highly integrated in a low-cost CMOS process, with high output power, high efficiency, wide bandwidth, low distortion, high reliability, and small size offers significant research challenges.

1.2 Outline

- **Chapter 1** introduces mmW communication, motivates why research in integrated mmW power amplifiers for wireless systems is needed, and organizes the dissertation.

- **Chapter 2** describes important transmitter architectures to place the

power amplifier in its context.

- **Chapter 3** defines, with examples, the most commonly used power amplifiers metrics.
- **Chapter 4** summarizes the most commonly referred power amplifier classes.
- **Chapter 5** presents, in brief, some theory about power amplifier linearity.
- **Chapter 6** analyses two efficient power amplifiers, the Doherty PA and the Chireix outphasing PA.
- **Chapter 7** presents the included papers with their scientific contribution and the author's contribution to the papers.
- **Chapter 8** concludes the dissertation with some suggestions for future work.
- **Paper I** presents a mmW variable gain PA driver and a PA in 28 nm FD-SOI CMOS, suitable for a highly integrated antenna array system for a 5G basestation.
- **Paper II** presents a mmW variable gain PA driver and PA, and a transmit/receive switch (TRX-switch) in 22 nm FD-SOI CMOS, suitable for a highly integrated antenna array system for a 5G basestation.
- **Paper III** presents a mmW transceiver with a Doherty amplifier in 22 nm FD-SOI CMOS, suitable for a highly integrated antenna array system for a 5G basestation.
- **Paper IV** presents analytical derivations of why a Doherty PA gains from dynamically altering the bias level depending on the input signal amplitude, so called adaptive bias.

Chapter 2

Transmitter Architectures

To place the PA in its context, some of the most well-known transmitter architectures are presented in this chapter. The task of the transmitter is to generate the analog signal to be radiated by the antenna, with a certain output power, carrier frequency, bandwidth, and modulation type. The output power should be large enough that the receiver can correctly decode the transmitted signal information, and if the targeted range of communication is increased more output power is required. The carrier frequency and bandwidth depends on in which part of the spectrum that the wireless communication should occur. The modulation type can be seen as a predetermined agreement between the transmitter and receiver on how the information should be coded into the analog signal. The bandwidth and modulation type will determine the maximum data-rate of the communication link.

2.1 Signal Fidelity and Constellation Diagram

In digital communication systems, where the analog signal carries digital information bits, the bits are typically represented by the phase and amplitude of the signal. A so-called 16-QAM constellation diagram, see Fig. 5, is used to visualise how the digital bits can be coded in the analog signal.

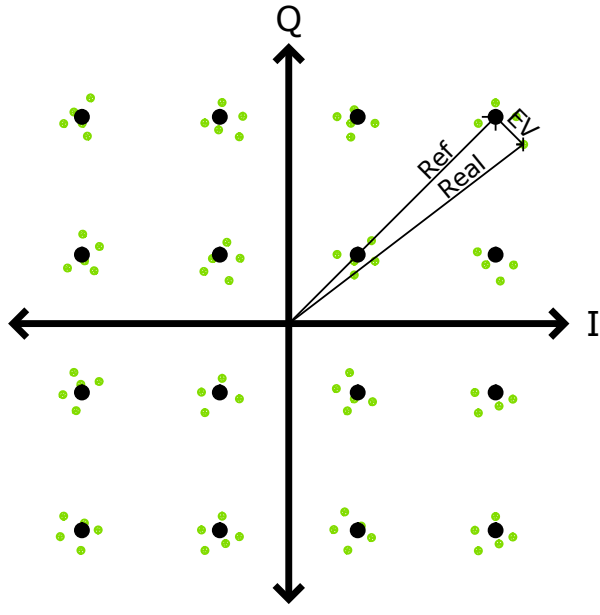

Figure 5: 16-QAM constellation diagram showning, I and Q vectors, Real, Ref and EV (Error Vector).

In QAM the signal is constructed from two carrier components, one called the (I)-component, which is $I \times \cos(\omega_c t)$, and one called the Q-component, which is $Q \times \sin(\omega_c t)$. The I and Q are amplitudes that can have both postive and negative values, the letters stand for in-phase and quadrature-phase which relates to the 90-degree phase shift between sinus and cosinus. Each of the I and Q components are then independently amplitude modulated. For the 16-QAM signal there are four possible values of I and Q, which gives in total 16 combinations as marked in the I,Q plane in Fig. 5. The distance from the origin represents the signal amplitude, and the phase is represented by the phase from the I-axis. Each black dot represents an ideal position, i.e. amplitude and phase of the total carrier signal, for transmission of a certain digital information symbol. In Fig. 5 there are in total 16 black dots, which each represent four digital bits, i.e. the information in a symbol. The green dots represent some actual, or non-ideal symbol transmissions due to transmitter imperfections. The Ref vector is the desired ideal transmission, the Real vector is the actual transmission, and the error vector (EV) is the vector between the actual transmission and the desired ideal point. To quantify how severe the deviations from the desired ideal points are, the average magnitude of the EV is compared with the average Reference vector magnitude and often reported in percent as the error vector magnitude (EVM):

$$EVM(\%) = \sqrt{\frac{\overline{P_{EV}}}{\overline{P_{Ref}}}} \cdot 100 \tag{4}$$

Where $\overline{P_{EV}}$ and $\overline{P_{Ref}}$ are defined as the average power for a number of symbols. As long as the EVM is low enough it will be possible to determine with high probability which symbols were transmitted, i.e. to decode the information correctly. The limit on how large EVM that a signal transmission can tolerate depends on how densely the constellation points are packed. The example shown in Fig. 5 has 16 constellation points, which requires an EVM below 12 %. But if more points are added, the points gets closer to each other, which requires lower EVM. Table 1 summarizes the EVM requirements for some QAM constellation sizes, and as can be seen the EVM requirements are much more stringent at 256-QAM than at 16-QAM [13].

Table 1: EVM Requirement for a selection of QAM signals

Modulation	**Required EVM**		**Bits/sym**
	[%]	[dB]	
16-QAM	12	-18.4	4
64-QAM	6	-24.4	6
256-QAM	3	-30.5	8

2.2 Cartesian Transmitters

The input signal to a Cartesian transmitter is represented in a Cartesian coordinate system and the baseband (BB) input signal consists of the two components I and Q (corresponding to x and y of a Cartesian coordinate system). Each component carries half the information, and consequently half the data-rate.

The central component of both the homodyne and heterodyne transmitter, which will be presented below, is the IQ-modulator (also known as image reject mixer or IQ-mixer), which comprises two mixers. One mixer produces the modulated I signal at the carrier frequency by multiplying the BB I component $BB_I(t)$ with $\cos(\omega_{LO}t)$ and the other mixer produces the modulated Q signal at the carrier frequency by multiplying the BB Q component ($BB_Q(t)$) with $\sin(\omega_{LO}t)$. The modulated I and Q signals from the two mixers are then combined into one signal by summation. The modulation of I and Q makes it possible to control both the amplitude and phase of the output signal, and thereby to represent any signal. Naturally, the goal is to control the signal such that it moves between the constellation points that represents the transmission bit sequence. An ideal IQ-modulator has identical gain in the BB I and Q paths and upconverts to carrier frequency with exactly 90° phase separation

between the cosine and sine functions. A single mixer operation is equivalent to a multiplication of the input baseband signal with the LO and will produce spectrum at both sides of the LO, i.e. at $\omega_{LO} \pm \omega_{BB}$ for a baseband input tone at ω_{BB}. The IQ-Modulator, however, can generate signal with different spectrum at the two sides of the LO, e.g. a tone at $\omega_{LO} + \omega_{BB}$ but no tone at $\omega_{LO} - \omega_{BB}$, called image rejection. How well the IQ-modulator suppresses the unwanted image depends on how accurately the mixers output signals are matched to the ideal 90-degrees phase difference, and also how accurately the amplitude of the two paths match. For small phase and amplitude deviations the image reject ratio (IRR) can be approximated as [14]:

$$IRR \approx \frac{4}{\varepsilon^2 + \theta^2} \tag{5}$$

where θ is the phase mismatch in radians and ε is the fractional gain mismatch. In practice, a typical design produces an IRR of about 25 - 35 dB [14]. Furthermore, ideally there should be no non-modulated LO signal present at the output of the IQ-modulator, i.e. no LO-leakage. In reality though, all IQ-modulators suffer from IQ-gain imbalance, IQ-phase mismatch, and LO-leakage. Gain and phase imbalance might arise from global and local process variations and the unavoidable deviation from a perfectly symmetric layout. The LO-leakage originates from DC offset(s) at the BB input of the mixer, which is upconverted to the LO frequency, and from direct leakage from the LO input to the output of the mixer. Gain imbalance will effectively make the I and Q components (vectors) have unequal length. Phase imbalance will make the I and Q vectors to deviate from a 90° phase relation in the IQ-plane. An LO leakage will add a DC shift to the constellation points. The effects on the symbols of a 64-QAM constellation diagram when introducing gain imbalance, phase imbalance, a combination of gain and phase imbalance, and an LO-leakage, and how the EVM varies over the IQ-plane are illustrated in Fig 6.

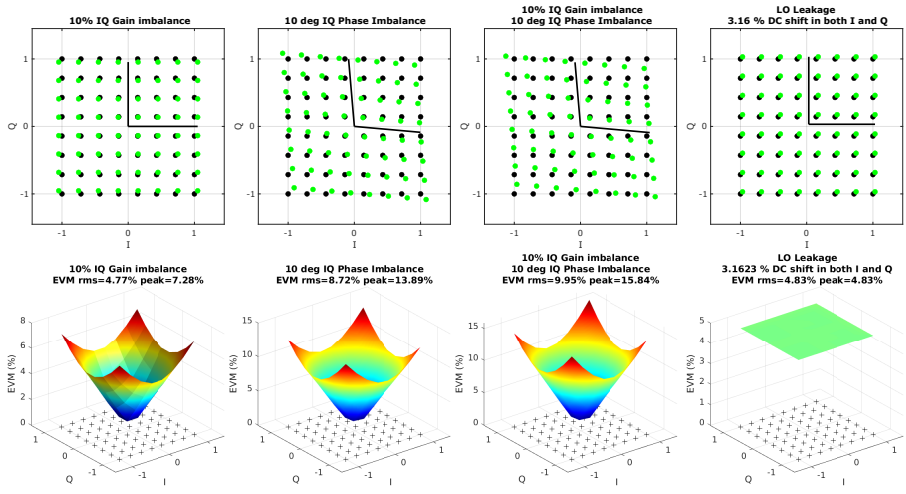

Figure 6: Upper: A 64-QAM constellation diagram showing green constellation points for a 10% gain imbalance, a 10° degree phase imbalance, a combined 10% gain imbalance with 10° degree phase imbalance, and a 5% DC shift caused by an LO-leakage. The black dots represent the ideal constellation points. Lower: Corresponding EVM contributions for the upper four plots.

Gain and phase imbalances will result in EVM that increases with the distance from origin. Table 2 summarizes the EVM contribution of various degrees of gain and phase imbalance.

Table 2: 64-QAM EVM contribution from gain and phase imbalance

Imbalance	**Gain EVM [%]**		**Phase EVM [%]**		**Gain & Phase EVM [%]**	
[%,deg]	rms	peak	rms	peak	rms	peak
0.1	0.05	0.08	0.09	0.13	0.10	0.15
1	0.5	0.8	0.9	1.3	1.0	1.5
10	4.8	7.3	8.7	13.9	10	15.8
20	9.1	14	17.4	28.9	19.7	32.6

LO-leakage produces a flat EVM, effecting all symbols equally and thereby making the rms and peak values the same. Table 3 summaries the EVM contribution of various degrees of LO-leakage for 16-QAM to 1024-QAM modulations. The small difference of EVM for the different constellation diagrams depends

on the small variation of average energy in the ideal symbols. The required EVM, on the other hand, is very different for 16-QAM and 1024-QAM.

Table 3: EVM contribution from LO-leakage

LO-leakage Below Max Ampli. [dB]	**16-QAM EVM [%]** rms/peak	**64-QAM EVM [%]** rms/peak	**256-QAM EVM [%]** rms/peak	**1024-QAM EVM [%]** rms/peak
-60	0.13	0.15	0.16	0.17
-50	0.42	0.5	0.51	0.53
-40	1.34	1.5	1.63	1.68
-30	4.24	4.8	5.14	5.31
-20	13.4	15.3	16.3	16.8

From Fig. 6 it is clear that the gain and phase errors between the I and Q signal components will impact the EVM of the transmitted signal. However, for such deterministic errors, as a constant gain and/or phase imbalance, the receiver could perform IQ-imbalance compensation algorithms as in [15], significantly reducing their EVM impact. On the other hand, even with ideal compensation of IQ-imbalance and LO-leakage in the receiver, the constellation points have been shifted from their intended positions in the transmission signal. This will affect how the signal gets distorted and reduce the useful signal power, as part of the transmitted power is allocated to transmit undesired signal components such as image and LO-leakage.

2.2.1 The Heterodyne Transmitter

The heterodyne (or superheterodyne) transmitter, see Fig. 7, uses multiple (at least two) stages for frequency upconversion from BB input signal to output carrier frequency. The analog input signals I and Q are created from digital input signals using digital to analog converters (DACs), and low pass filters (LPF) that suppresses repetitive spectrum due to limited sampling frequency in the DACs. An IQ-Modulator upconverts the baseband signals to an intermediate frequency (IF), which typically is band pass filtered (BPF) to suppress unwanted out-of-band signals. Then follows a mixer which will produce signal at both $\omega_{F2} + \omega_{F1}$ and $\omega_{F2} - \omega_{F1}$. The second BPF is needed to suppress the unwanted one of these two signals, the image signal. A variable gain amplifier is then used to adjust the signal amplitude such that the PA produces the desired output power level. A benefit with the heterodyne transmitter is that it allows for a simpler LO design as the challenging LO quadrature phase accuracy, needed in the IQ-Modulator, can occur at a lower frequency (IF). Another thing worth mentioning is that LO-leakage from the second mixer ends up out-of-band, which causes less problem for EVM, but might violate a possi-

ble stringent out-of-band emission requirement, or put increased requirements on BPF_2. Simply speaking, quite a large part of the design takes place at a lower frequency than the high output frequency. The drawback is the obvious problem with the needed image suppression after the second mixer, but also that the heterodyne transmitter requires two LO frequencies for upconversion and a multitude of analog subblocks.

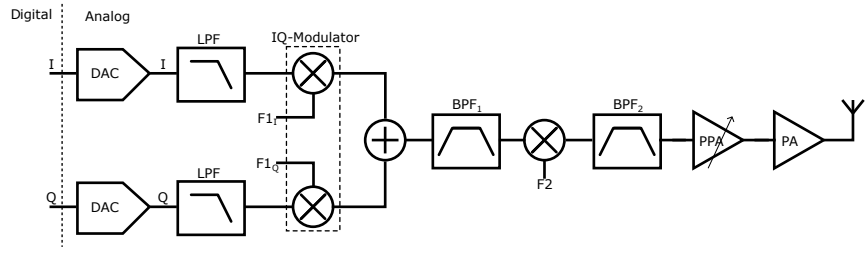

Figure 7: A heterodyne transmitter.

2.2.2 The Homodyne Transmitter

The homodyne transmitter, see Fig. 8, uses a single stage frequency upconversion from baseband to carrier frequency. As for the heterodyne, LPFs are used to suppress aliasing products from the input signal generated by the DACs. After frequency upconversion by the IQ-modulator, a variable gain amplifier adjusts the signal to the PA for appropriate output power level. The homodyne transmitter has a couple of benefits, but also disadvantages compared to the heterodyne. The obvious advantages are that the homodyne transmitter requires fewer sub-blocks and it does not require generation of two different LO frequencies for up-conversion. However, perhaps more importantly, it does not require any band pass filters since it does not produce an out-of-band image. On the downside, depending on the operating frequency and choice of process, it can be a much harder design task, especially the IQ modulator. To avoid a significant EVM contribution from the IQ-imbalance, which would effectively limit the use of high order modulation, see table 1 and table 2, the requirements on amplitude and phase match becomes very challenging at high frequencies. As an example, to achieve an EVM contribution of 1% at 30 GHz, and assuming perfect amplitude match, the phase mismatch should be below 1°, which corresponds to a time error below 0.1 ps between the LO I and Q signals, and since in reality a small amplitude mismatch is unavoidable the time mismatch must be even lower than that. To reduce the level of LO-leakage, which causes a DC offset in the constellation diagram, careful mixer design is needed. The most important is to avoid introducing DC offset at the BB input of the two mixers in the IQ-modulator, by using well-known design and layout

strategies to reduce the local mismatch, and to plan the layout such that the coupling from the LO input to the output is minimized, thereby reducing the direct LO-leakage. Even after careful design, the LO-leakage might still be too large for system specifications. Luckily there are various LO-leakage suppression techniques available, such as presented in [3], which uses programmable DC current sources that inject current into the Gilbert mixer, to cancel the effect of LO-leakage. Another weak spot of the homodyne transmitter is that since the PA operates at the same center frequency as the LO signal, used for frequency upconversion, pulling by the PA of the LO circuitry (oscillator or possibly by quadrature generation dividers) can become a challenging problem. This problem can, however, quite often be mitigated by careful planning of the LO frequency generation. It also helps that the signal is often wideband in mmW systems, as LO circuitry is more sensitive to narrowband pulling.

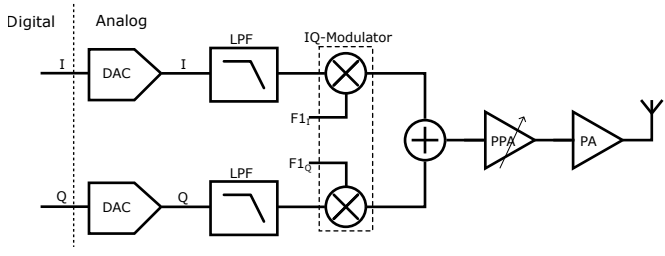

Figure 8: A homodyne transmitter.

2.3 The Polar Transmitter

This section presents a polar transmitter architecture along with some considerations. The input signal to a polar transmitter is represented in a polar coordinate system. Rather than Cartesian I and Q, the baseband input signal then consists of amplitude (A) and phase (θ), corresponding to r and ϕ of the polar coordinate system often used in mathematics. Figure 9 shows a polar transmitter architecture [16, 17]. In the digital domain the phase modulation signal θ is derived from the Cartesian IQ signals using the nonlinear arctan2 function, but since the input signal to the digital phase locked loop (DPLL) is a digital frequency control word, a conversion from phase modulation θ to frequency modulation is also performed (including phase unwrapping). The digital frequency modulation is then input to the DPLL, which generates a constant envelope phase modulated signal at the carrier frequency. The amplitude modulation is applied directly at the supply voltage of the PA. The main advantage of the polar transmitter is that since the input signal to the PA has a constant envelope, it becomes possible to use a highly efficient switched mode PA (SMPA). One major challenge with polar transmitters is the bandwidth ex-

pansion due to the nonlinear transfer from Cartesian to polar coordinates, the drawback being similar to that of the outphasing PA as described in chapter 6, both in text and with Eq. 68 - 71. The bandwidth of the amplitude and phase signals can be more than ten times the bandwidth of the signals in a Cartesian coordinate system, and it is most severe for the phase signal [18]. To increase the phase modulation bandwidth, the PLL may employ so-called two-point modulation. This means that the frequency control word in the digital domain is fed to the DPLL using two modulation paths, one slow and one fast, where the slow path controls the division ratio in the feedback frequency divider of the DPLL and the fast one controls capacitors in the resonance tank of the digitally controlled oscillator (DCO) to quickly steer the frequency. The bandwidth expansion is particularly troublesome for 5G mmW communication as the transmission bandwidths have increased significantly compared to 4G. Another drawback of the polar transmitter is the timing accuracy requirement between the amplitude and phase modulation, which becomes difficult for high modulation bandwidths. In a Cartesian transmitter the I and Q paths are identical parallel paths, which makes matching relatively easy, whereas in the polar transmitter the two modulation paths are completely different and still must be very accurately matched in time. A final remark about the polar transmitter is that it requires an effective and wide bandwidth supply modulator, typically a DC-DC converter, to control the PA output amplitude, and that the switching frequency of the DC-DC converter risks introducing additional supply noise degrading EVM and causing out-of-band emissions. EVM contributions and emissions will also originate from the mis-alignments, bandwidth limitations, and nonlinear behaviour of the PA at low supply voltages.

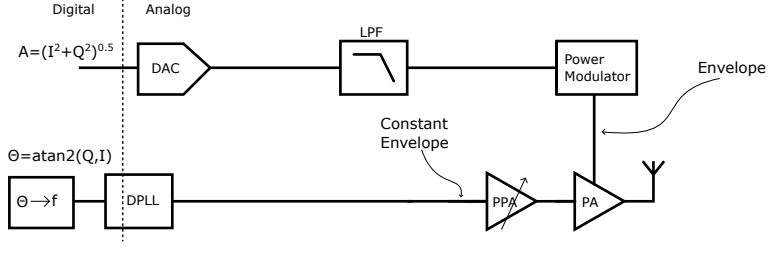

Figure 9: A transmitter using polar modulation.

Chapter 3

Power Amplifiers Metrics

This chapter presents a list with a brief explanation of the most commonly used PA metrics. Understanding them is vital when designing or evaluating a PA, transmitter, or a complete system for transmission and reception of wireless signals. The most commonly used unit for the metric is indicated within brackets. To facilitate the understanding of how the metrics are used, an example for most of the metrics indicating the performance of a mmW PA is provided in the list. All the examples are taken from the PA presented in [3].

- Power gain [dB] is the output power (P_{out}) [dBm] minus the input power (P_{in}) [dBm] of the PA. Typically it is desirable to have high gain such that a lower input signal is required.

$$Gain = P_{out} - P_{in} \tag{6}$$

  Example: *The Power gain of the PA at 26.5 GHz is 16 dB*

- The gain for small signals is referred to as small signal gain (SSG) [dB].
  Example: The SSG of the PA at 26.5 GHz is 16 dB

- Saturated output power (P_{sat}) [dBm], see Fig. 12, is the maximum output power that the PA can deliver. However, typically the PA cannot deliver a useful output power all the way up to P_{sat} due to nonlinearities.
  Example: P_{sat} *at 26.5 GHz is 19.4 dBm*

- Input referred 1 dB compression point (IP_{1dB}) [dBm], see Fig. 12, is the input power for which the gain has compressed by 1 dB compared to the SSG. Using a third order power series expansion to express the output signal as a function of the intput signal $v_{out}(t) = k_1 v_{in}(t) + k_2 v_{in}^2(t) + k_3 v_{in}^3(t)$, as proposed in Eq. 53 in chapter 5 and presuming that the PA has a compressive behaviour, then IP_{1dB} can be calculated as [19]:

$$IP_{1dB} = \sqrt{|\frac{4k_1}{3k_3}|}\sqrt{0.11} = IIP3 - 9.6\,dB \tag{7}$$

Example: IP_{1dB} *at 26.5 GHz is 2 dBm*

- Output referred 1 dB compression point (OP_{1dB}) [dBm], see Fig. 12, is the output power for which it has compressed by 1 dB compared to a linear gain of the input signal. Ideally the OP_{1dB} should be close to P_{sat}. According to definitions OP_{1dB} is equal to IP_{1dB} plus the SSG, minus 1 dB:

$$OP_{1dB} = IP_{1dB} + SSG - 1\,dB \tag{8}$$

  Example: OP_{1dB} *at 26.5 GHz is 17 dBm*

- Drain efficiency (DE) or (η) [%], is the output power in Watts divided by the DC power consumption of the PA in Watts and defined as:

$$\eta = \frac{P_{out}}{P_{DC}} \tag{9}$$

  The efficiency is signal dependent, and needs to be specified for a certain signal type and level.

  Example: *The DE at 26.5 GHz for a 13.3 dBm orthogonal frequency division multiplex (OFDM) signal with 1600 MHz bandwidth is 13.8 %*

- Power added efficiency (PAE) [%], is the output power in Watts minus the input power in Watts, divided by the DC power consumption of the PA in Watts, and consequently defined as:

$$PAE = \frac{P_{out} - P_{in}}{P_{DC}} \tag{10}$$

  By taking into account also the input power required, this provides a more complete performance measure than DE.

  Example: *The PAE at 26.5 GHz for a 18 dBm CW tone is 17.5 %*

- 1- or 3 dB small signal bandwidth (SSBW) [Hz], is the frequency bandwidth in which the SSG has decreased by less than 1- or 3 dB relative to the peak SSG.

  Example: *The 3 dB SSBW is 2.4 GHz*

- Fractional bandwidth [%], is the ratio between the absolute bandwidth and the center frequency.

  Example: *The fractional bandwidth when transmitting a 1.6 GHz wideband signal at 26.5 GHz is 6 %*

- Instantaneous bandwidth [Hz], is the bandwidth that can be used concurrently for transmission.

- Amplitude modulation to amplitude modulation (AM-AM), typically reported in dB as a function of input or output power in dBm, is a measure of how the signal amplification varies with the amplitude of the output signal. For low input signal levels the amplification is typically linear and the output amplitude follows the input signal amplitude with constant gain, but for large input signals the gain compresses. AM-AM shows how linear the amplification of an amplifier is, i.e. for which input and output signal levels it can be used. Fig. 10 is an example of an AM-AM plot.

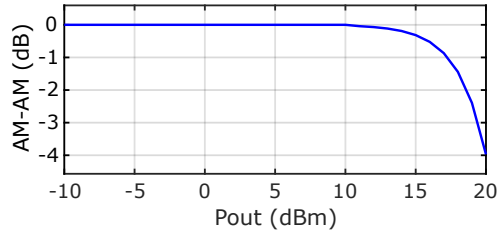

Figure 10: Example of an AM-AM plot.

- Amplitude modulation to phase modulation (AM-PM), typically reported in degrees as a function of input or output power in dBm, is a measure of how the output phase varies with the amplitude of the output signal. For low input signal levels the amplification is typically linear and the output phase follows the phase of the input signal with a constant phase shift. As the amplitude of the input signal increases, however, the PA starts to deviate from linear amplification and the output phase might might lead or lag compared to the phase for low signal levels. AM-PM shows how well an amplifier preserves the phase information, for various output signal levels, i.e. for which input and output signal levels it can be used. Fig. 11 is an example of an AM-PM plot.

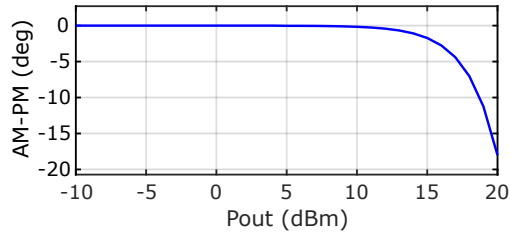

Figure 11: Example of an AM-PM plot.

- Second order output referred intercept point (OIP_2) [dBm], see Fig. 12, is the intersection of the fundamental and 2nd order intermodulation (IM_2) products when extrapolated with a slope of 1 dB / 1 dB of increased input power for the fundamental and 2 dB / 1 dB of increased input power for the IM_2 products.

- Third order output referred intercept point (OIP_3) [dBm], see Fig. 12, is the intersection of the fundamental and 3rd order intermodulation (IM_3) products when extrapolated with a slope of 1 dB / 1 dB of increased input power for the fundamental and 3 dB / 1 dB of increased input power for the IM_3 products.

- Main channel power [dBm], is the power that is transmitted inside the intended frequency channel as shown in Fig. 13.

  Example: *The main channel power at 26.5 GHz for a 16-QAM OFDM signal with 1600 MHz bandwidth is 13.3 dBm*

- Adjacent channel leakage ratio (ACLR) [dB], is the ratio between the transmitted power in the main channel and the power in one of the adjacent channels as shown in Fig. 13.

  Example: *ACLR at 26.5 GHz for a 13.3 dBm 16-QAM OFDM signal with 1600 MHz bandwidth is 27.5 dB*

- Error Vector Magnitude (EVM) [% or dB], is the rms deviation from the desired points in the constellation diagram compared with the desired points rms amplitude.

  Example: *The EVM at 26.5 GHz for a 13.3 dBm 16-QAM OFDM signal with 1600 MHz bandwidth is -19.8 dB*

Figure 12 (a) shows a subset of the frequency tones generated from a third order nonlinearity in a two tone test. The subset in the figure consists of the two fundamentals and 2nd and 3rd order intermodulation products. Appendix A

shows a complete list of the frequency response from a third order nonlinearity in a two tone test. Figure 12 (b) shows the power levels of fundamental, IM2, and IM3 products as a function of input tone power. At small signal levels the fundamental tones increase linearly with input level (1 dB per 1 dB), IM2 increases quadratically (2 dB per 1 dB), and IM3 increases cubically (3 dB per 1 dB). Figure 12 (b) offers a graphical visualization of the SSG, OP_{1dB}, OIP_3, OIP_2, IP_{1dB}, IIP_3, IIP_2, and P_{sat}.

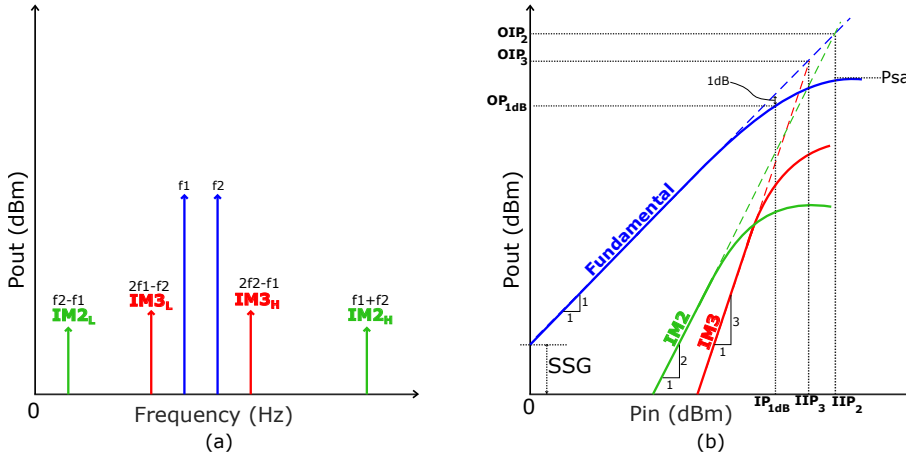

Figure 12: (a): Frequency spectrum from a two tone test producing 2^{nd} and 3^{rd} order intermodulation products. (b) Pin/Pout of fundamental, IM2, and IM3, which in (a) is f1 & f2, IM2_L & IM2_H, and IM3_L & IM3_H, respectively

Figure 13 shows an example of the power spectral density (PSD) of the input and output signals to/from a PA, simulated in a 22 nm FD-SOI CMOS process.

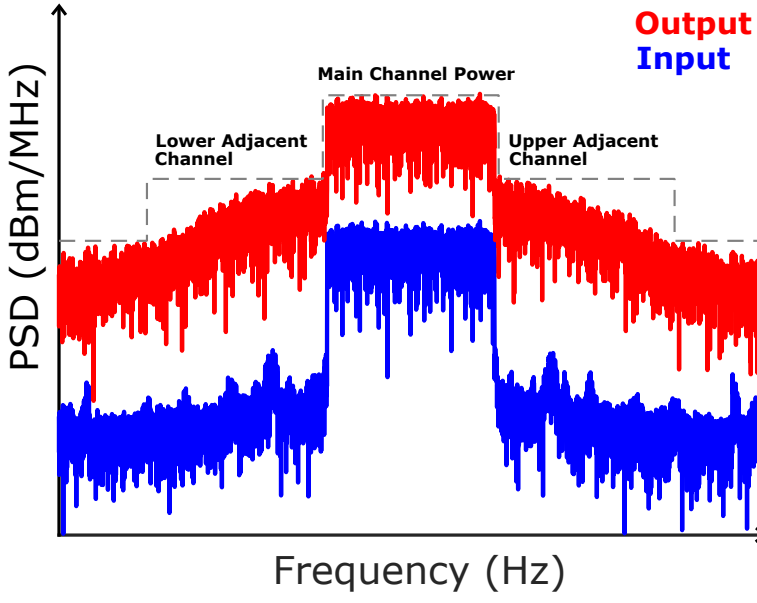

Figure 13: Example of a modulated input signal that has been amplified by a nonlinear PA. Due to third order nonlinearities power has leaked into the adjacent channels.

Chapter 4

Power Amplifier Classes

Power amplifiers classification depends on their bias, harmonic termination, and the operation mode of the transistor. In this chapter, class A to class F are discussed. More PA classes exist, but for brevity, they are left out from the discussion here. In the first three classes (A - C), the transistor operates in active mode as a voltage controlled current source, and these are referred to as transconductance mode PA classes. For the last three classes (D - F), the transistor operates in triode or cut-off region and acts as a voltage controlled switch, which are referred to as switched mode power amplifier (SMPA) classes.

4.1 Transconductance PAs

In transconductance mode PAs, which are also called linear PAs, the transistor operates as a voltage controlled current source. For this the transistor needs to be biased so it remains in the active region for all input signal levels were the transistor is turned on, i.e. the voltage across the transistor, between drain and source, must not drop below the pinch-off voltage $\mathrm{V_{ds}} > \mathrm{V_{pinch-off}}$. Such a bias provides possibility for good linearity as the output signal is a direct representation of the input signal, at least for some portion of the cycle, but unfortunately it also results in high voltage and large current through the transistor at the same time. The power dissipated in the transistor is directly proportional to the drain-to-source voltage multiplied by the drain-to-source current, and the average power dissipation over a signal period T becomes:

$$P_{diss} = \frac{1}{T}\int_0^T v_{DS(t)} i_{DS(t)}\, dt \tag{11}$$

Which implies that to minimize the dissipated power in the transistor, the voltage across, and the current through, the transistor should be separated in time. Separation in time of the two quantities, voltage and current, is controlled by the so-called conduction angle of the PA. Generally speaking, reduced conduction angle reduces losses in the transistor and increases the efficiency of the PA. Nevertheless, as the conduction angle decreases, the part of the cycle for

which the output signal, to the first approximation, is a linear representation of the input signal, decreases, which naturally reduces the linearity of the PA, and hence the trade-off between efficiency and linearity becomes clear. However, this is a bit simplified as, theoretically, perfect linearity can be achieved not only for conduction all the time, but also for conduction of exactly half the time, i.e. for a conduction angle of π. Furthermore, the transistors also have a non-linear transfer from input voltage to output current, i.e. a signal level dependent transconductance that limits the PA linearity. There are also other non-linear impedances in real transistors.

A simplified schematic of a transconductance mode PA operating at radio frequency (RF) is depicted in Fig. 14.

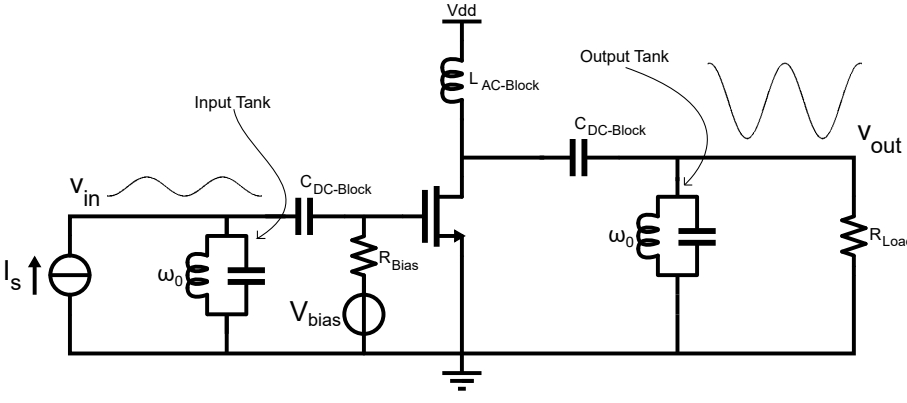

Figure 14: A general schematic of a single stage, single ended, single transistor PA

The PA has four basic parts, the input resonance tank, the amplifying transistor, the bias networks, and the output resonance tank. The input resonance tank, together with the input parasitic capacitances, normally dominated by the gate to source capacitor C_{gs} of the transistor, is designed to resonate at the carrier frequency. The impedance at parallel resonance, set by the quality factor of the resonance tank and the size of the inductor, becomes $R_p = Q\omega_0 L$, where $Q = Q_L//Q_C$. Furthermore, the quality factor also sets the bandwidth of the voltage signal at the gate of the input transistor. The input and output bias networks $L_{AC-Block}$, $C_{DC-Block}$ & R_{Bias} provide the transistor with DC power from the supply, biases the input gate, and prevents DC power from being dissipated in the load. Finally, the output resonance tank resonates with the parasitic capacitances at the output, presenting a resistive load R_{load} at the resonance frequency, and here it is assumed that all parasitic output capacitances have been absorbed into the output resonance tank. The losses of

the unloaded output resonance tank produces an unwanted equivalent parasitic resistance, reducing the output power and efficiency, typically dominated by the limited Q-value of the inductor.

4.1.1 Class A

Figure 15 shows drain voltage and current for a class A biased PA. In class A operation the bias is chosen such that the transistor always conducts, i.e. $I_D > 0$ at all times, resulting in a conduction angle equal to 2π. Class A is sometimes referred to as the linear mode, and although class A is the most linear mode of operation, PAs biased in class A are not perfectly linear, since they will still suffer from weak non-linearities in transistor transconductance (g_m) and other impedances. Since g_m of a transistor increases with the bias level, class A PAs also have the highest gain. The downside is that class A PAs have the lowest efficiency as the transistor constantly dissipates power.

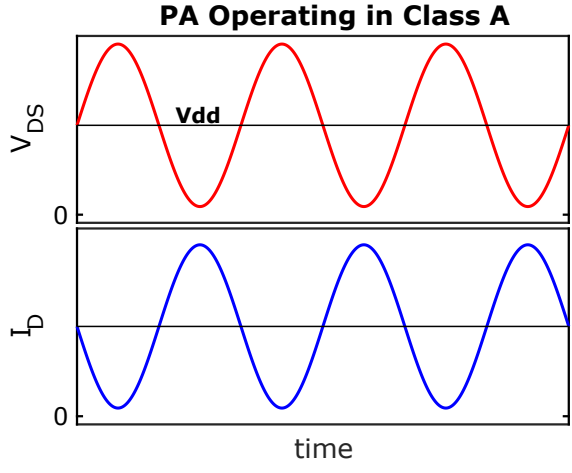

Figure 15: Idealized drain voltage and current for a class A biased PA

If biased in class A and driven by a continuous wave (CW) sinusoidal of angular frequency ω, assuming an ideal transistor with zero pinch-off voltage, the PA in Fig. 14 maximum output voltage (V_{out}) is:

$$v_{out} = V_{dd} \sin(\omega t) \tag{12}$$

which gives the maximum delivered average output power into R_{Load}:

$$P_{R_{Load}} = \frac{V_{dd}^2}{2R_{Load}} \tag{13}$$

The lowest possible bias current, that still keeps the PA in class A, is equal to the peak signal current, which gives $\mathrm{I_{DC}} = \frac{\mathrm{V_{dd}}}{\mathrm{R_{Load}}}$. The minimum DC power then becomes:

$$P_{DC} = \frac{V_{dd}^2}{R_{Load}} \tag{14}$$

which gives the theoretical maximum drain efficiency η for class A:

$$\eta = \frac{P_{R_{Load}}}{P_{DC}} = \frac{\frac{V_{dd}^2}{2R_{Load}}}{\frac{V_{dd}^2}{R_{Load}}} = \frac{1}{2} = 50\,\% \tag{15}$$

which is a rather low value for the theoretical maximum achievable efficiency. Another weak spot for the transconductance PAs is the relatively large stress that the transistors are exposed to. This is visible from the low normalized power output capability, which for the class A amplifier becomes:

$$P_N = \frac{P_{out}}{v_{DS,max} i_{D,max}} = \frac{V_{dd}^2/(2R_{Load})}{2V_{dd} \cdot 2V_{dd}/R_{Load}} = \frac{1}{8} \tag{16}$$

In conclusion the class A amplifier provides high linearity and gain at the cost of efficiency and device stress.

4.1.2 Class B

Figure 16 shows drain voltage and current for a class B biased PA. In class B the bias is chosen such that the drain current is equal to zero exactly 50 % of the time. The transistor thus conducts during half of the cycle, resulting in a conduction angle equal to π. This is clearly a deviation from linear operation and therefore the class B amplifier requires a high-Q output resonance tank to filter out harmonics at the output. Another popular class B implementation, not in detail discussed here, is to use a push-pull stage to avoid the need for a high-Q output tank. Ideally, however, a class B PA is completely linear, except for the generation of harmonics, if the transistor turns on/off perfectly with constant transconductance when on, which unfortunately is impossible. It is important to point out that true class B operation does not exist in reality, since it in principle is impossible to bias a PA to conduct exactly 50 % of the time. Having the bias slightly too high results in class AB and having it slightly too low gives class C. Regardless, class B is still important to analyse from an efficiency perspective, as it marks the border when class AB goes into class C.

As for class A, the maximum power delivered into $\mathrm{R_{Load}}$ for the class B is:

$$P_{R_{Load}} = \frac{V_{dd}^2}{2R_{Load}} \tag{17}$$

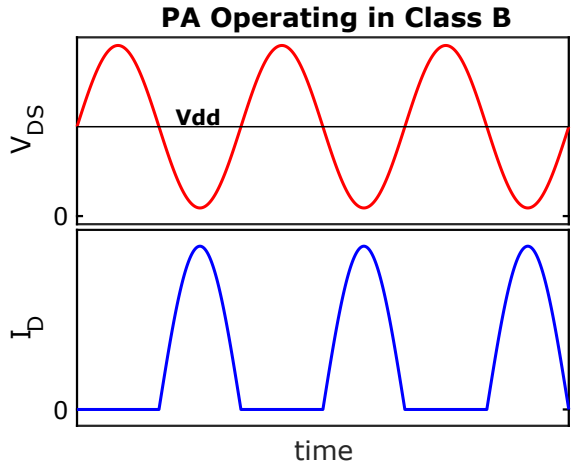

Figure 16: Idealized drain voltage and current for a class B biased PA

The DC power is found by first integrating the drain current, which is zero for half the cycle, and a rectified sinusoidal with a peak value of $2V_{dd}/R_{Load}$ at maximum power delivered for the other half of the cycle, and then dividing the integrated current by the period time T to get the DC current drawn, which is then multiplied with V_{dd} to get the DC power from the supply:

$$P_{DC} = V_{dd}\frac{2V_{dd}}{TR_{Load}}\left[\int_0^{T/2} 0\,dt + \int_{T/2}^{T} sin\,\omega t\,dt\right] = \frac{2V_{dd}^2}{\pi R_{Load}} \tag{18}$$

Which gives the theoretical maximum drain efficiency η for class B:

$$\eta = \frac{P_{R_{Load}}}{P_{DC}} = \frac{\frac{V_{dd}^2}{2R_{Load}}}{\frac{2V_{dd}^2}{\pi R_{Load}}} = \frac{\pi}{4} \approx 78.5\,\% \tag{19}$$

Class B has the same normalized output power capability as class A, with $P_N = 1/8$. In conclusion class B, compared to class A, has improved efficiency at the cost of linearity and gain, but with about the same device stress.

4.1.3 Class AB

Class AB is, as indicated by the name, biased between class A and B, with a conduction angle somewhere between that of a class A and class B, i.e. $\pi < 2\Phi < 2\pi$. The resulting theoretical peak efficiency is between that of class A and B, i.e. $50\,\% < \eta < 78.5\,\%$, and the linearity performance is also somewhere between that of a class A and a class B biased PA.

4.1.4 Class C

Figure 17 shows drain voltage and current for a class C biased PA. The gate bias voltage has been further reduced compared to class B, allowing conduction for an even smaller portion of the cycle. For the class C amplifier the conduction angle, shown in Fig. 17 as 2Φ, is below π, which results in conduction of less than 50 % of the time. As for the class B amplifier, even an ideal implementation of the class C amplifier needs a high-Q output resonance tank to filter out unwanted harmonics. As for class B a push-pull implementation for the class C exists, but contrary to the class B push-pull implementation the class C push-pull still needs a high-Q output resonance tank to suppress unwanted harmonics.

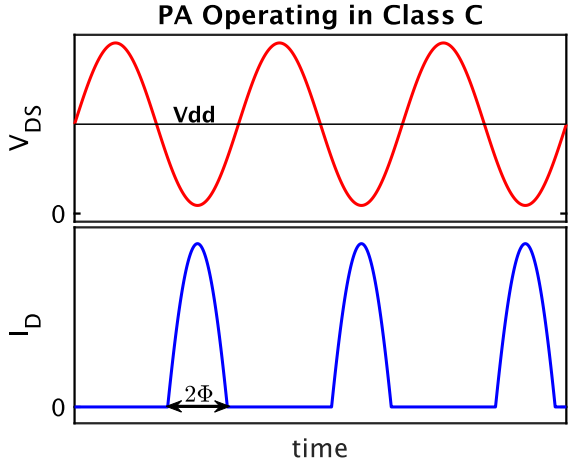

Figure 17: Idealized drain voltage and current for a class C biased PA

Defining the conduction angle as 2Φ from Fig. 17 and using the result in [14] for a class C PA, the maximum efficiency can be expressed as a function of Φ, and is found to be:

$$\eta = \frac{2\Phi - sin\,2\Phi}{4(sin\,\Phi - \Phi cos\,\Phi)} \tag{20}$$

Using $\Phi = \pi$ and $\Phi = \pi/2$ as a sanity check gives the expected result of $\eta = 1/2$ and $\eta = \pi/4$ for the maximum efficiency of the class A and B, respectively. As the conduction angle 2Φ approaches zero the maximum efficiency increases towards 100 %. Concurrently, however, the delivered output power also goes towards zero, since the power at the fundamental frequency approaches zero for the shorter and shorter drain current pulses, which for limited peak

current values contain less and energy and pushes the delivered power to higher and higher harmonics. The normalized output power handling capability then also goes to zero as P_{out} goes to zero while $i_{D,max}$ and $v_{D,max}$ remain. Even though listed among the transconductance (linear) PAs, the class C is far from linear. It exhibits low gain, ideally zero, for low input signal levels, and gain expansion as the input signal increases, which is contrary to the class A,B, and AB for which the gain compresses, as input signal level increases. For small signal levels, however, the class B amplifier can also exhibit gain expansion as the transistor turns on.

The normalized output power handling capability of a class C amplifier becomes a function of the conduction angle. For high conduction angles it will asymptotically approach the performance of the class B amplifier of $P_N = 1/8$.

4.1.5 Transconductance mode Back-Off Output Power, Gain, and Efficiency

An output power versus input power comparison for ideal class A, AB, B, and C is shown in Fig. 18, with the corresponding gains shown in Fig. 19. The power in the fundamental harmonic is found by squaring the current and multiplying with R_{Load}.

$$P_{fund} = \frac{I_{fund}^2 R_{Load}}{2} = (\frac{I_{max}}{2\pi}\frac{2\Phi - sin\,2\Phi}{1 - cos\,\Phi})^2 \frac{R_{Load}}{2} \tag{21}$$

In this comparison all the amplifiers have ideal transistors with the same gm and the supply voltage is set to 5 V such that a maximum of 24 dBm can be delivered into a 50 Ω load. P_{fund} was therefore limited to 24 dBm. The class A and B amplifiers are assumed to have a constant bias, and constant conduction angle of $2\Phi = 2\pi$ and $2\Phi = \pi$, respectively. For output power levels below 10 dBm the class AB amplifier is biased in class A and has a conduction angle of 2π, but for output power levels above 10 dBm, the conduction angle asymptotically approaches class B operation with a conduction angle of π. Since the class B has 6 dB lower gain than class A, class AB becomes fundamentally nonlinear since its conduction angle changes with the output power level. The bias for the class C amplifier is chosen such that when it delivers the peak output power of 24 dBm it has a conduction angle of 0.8π. When biased in class C the gain is even lower than for class B.

The calculated drain efficiencies for class A, B, AB, and C transconductance mode PAs in the previous sections are the theoretical maximum values, where the transconductance is constant when the transistor is on, all output voltage harmonics are filtered out, and the output voltage amplitude equals the supply voltage so that the transistor has zero voltage drop at the signal peaks. Using the same ideal conditions but reducing the input signal amplitude, reduces the output power, and, as might be expected, results in a reduced efficiency. This

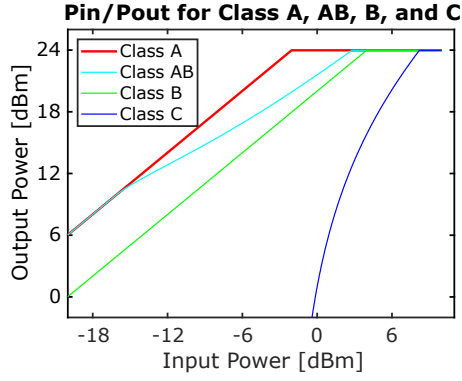

Figure 18: Output power vs. input power for class A, AB, B, and C

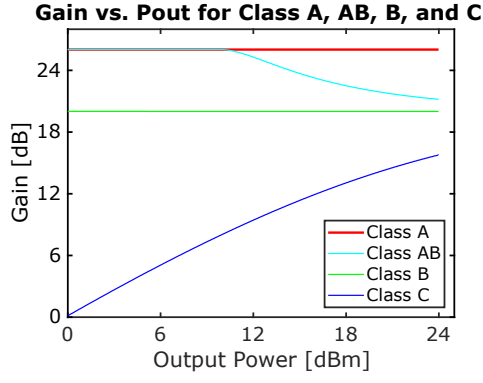

Figure 19: Gain vs. output power for class A, AB, B, and C

has the simple explanation that the output signal amplitude no longer equals the supply voltage, and the difference voltage will appear as a voltage drop over the transistor, increasing the fraction of the supply power dissipated as heat in the transistor. A comparison of the efficiency between the class A, AB, B, and C, with the same conditions as in the output power versus input power comparison above, for reduced output power levels, is shown in Fig. 20. Class A has the peak efficiency of 50 %, which reduces linearly when the output power is reduced. Class B has the peak efficiency of 78.5 %, which reduces with the square root of the output power. For high output power levels, above 10 dBm, the efficiency of the class AB is similar to class B, but below 10 dBm the efficiency of the class AB goes down linearly when the output power is reduced, since it is then biased in class A. The class C amplifier is

also assumed to have a constant bias, and in this example it has a conduction angle $2\Phi = 0.8\pi$ at the maximum output signal, then as the output signal level decreases the conduction angle gradually decreases.

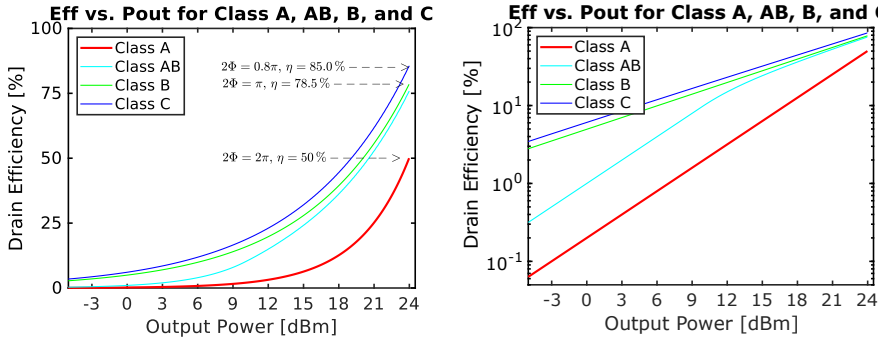

Figure 20: Efficiency for class A, B, and C plotted v.s. output power. For class C the max output power has a conduction angle of $2\Phi = 0.8\pi$. Left: linear scale on y-axis. Right: logarithmic scale on y-axis.

4.2 Switched Mode PAs

For SMPAs the transistor operates in triode and cut-off regions and acts as a voltage controlled switch. Theoretically, i.e. for an ideal switch, the voltage and current are completely separated in time, when there is voltage across the switch there is no current and vice versa, resulting in zero power dissipation in the switch. However, in reality, complete separation in time is impossible and switch losses will occur. The losses in an SMPA originates from three main sources, firstly power losses in the effective on resistance R_{on} of the transistor when conducting, i.e. when in triode region, secondly losses related to charging/discharging the output capacitance when moving between the cut-off and triode regions, and finally losses related to charging/discharging the input capacitance, dominated by C_{gs}. These losses can be summarized as:

$$P_{loss} = P_{R_{on}} + P_{C_{out}} + P_{C_{in}} = \frac{1}{T}\int_0^T i_S^2(t) R_{on}\, dt + f_0 \frac{1}{2}(C_{out}V_{C_{out}}^2 + C_{in}V_{in}^2) \tag{22}$$

Switched mode operation is ensured by applying a large enough input signal to compress/saturate the transistor, and in this context, large enough is in relation to the chosen bias point. One disadvantage with all SMPAs, since they operate in a digital on/off manner, is that all amplitude information from the input signal is lost, but nevertheless they can still be used to transmit amplitude modulated signals through various techniques, such as pulse width modulation (PWM) or supply voltage modulation. For the SMPA to work well, i.e. for the switching operation to be fast enough, it has to operate significantly below f_T of the transistor [14].

4.2.1 Basic RF SMPA

When analysing the basic RF SMPA driving a broadband resistive load as depicted in Fig. 21 it is clear that, since the switch is considered ideal, all DC power is converted to RF power. Even so, as it converts power from DC not only to the fundamental harmonic, but also to higher harmonics, its efficiency will still be below 100 %. Derivation of the analytical expression for the efficiency of a basic SMPA that drives a broadband resistive load, is readily available from the detailed analysis in [20] and here the most important steps are reproduced. Current and voltages across the ideal switch are shown in Fig. 22, and defined as before, with the conduction angle 2Φ indicated. The output power at the fundamental harmonic as a function of 2Φ is found to be:

$$P_{fund-R_{Load}} = V_{dd} I_{dc} \frac{2sin^2\,\Phi}{\Phi(\pi - \Phi)} \tag{23}$$

and since $V_{dd}I_{dc} = P_{DC}$ this naturally gives the efficiency:

$$\eta = \frac{2sin^2\,\Phi}{\Phi(\pi - \Phi)} \tag{24}$$

which has a peak efficiency of $8/\pi^2 \approx 81\,\%$ at $\Phi = \pi/2$, equivalent to a symmetrical square wave at the output, basically providing peak power at the fundamental frequency.

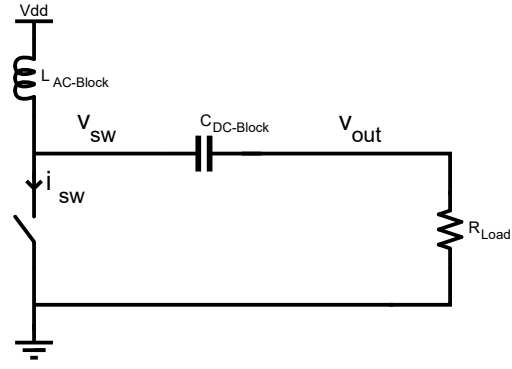

Figure 21: Schematic of a basic SMPA

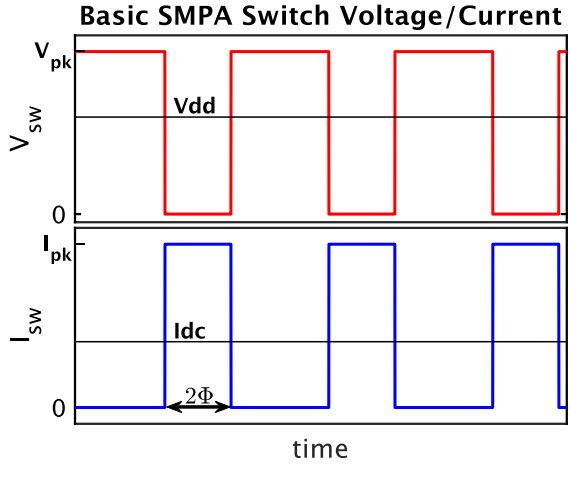

Figure 22: Ideal switch voltage and current for a basic SMPA

An implemented switch will have an on resistance, $R_{on} > 0$, which will degrade the performance of the SMPA. Figure 23 shows the switch voltage and

current for the basic SMPA from Fig. 21, but with an on resistance of the switch that is not insignificant compared to the load resistance. The voltage at the switch V_{SW} alternates between the two extreme states $V_{SW_{high}}$ and $V_{SW_{low}}$. To simplify the following analysis the conduction angle is fixed at $2\Phi = \pi$. Due to the presence of the DC Block capacitor, all DC current has to flow through the switch. Since the off resistance of the switch R_{off} is considered infinite in this analysis, the DC current has to flow for half the period i.e. when the switch voltage is at $V_{SW_{low}}$. This leads to a relation between the DC current and $V_{SW_{low}}$, which can be expressed as:

$$I_{dc} = \frac{V_{SW_{low}}}{2R_{on}} \tag{25}$$

The current that flows from the AC-Block inductor goes into the load for the half period that the switch is open, and since no DC can flow through the DC-Block capacitor, the current through the load has to be reversed for the other half. This makes it possible to conclude that the peak-to-peak voltage at the switch and at the output is:

$$V_{pp} = 2I_{dc}R_{Load} \tag{26}$$

Due to symmetry reasons, meaning that $2V_{dd} - V_{SW_{high}} = V_{SW_{low}}$ one can also conclude that:

$$V_{dd} - V_{SW_{low}} = \frac{V_{pp}}{2} = I_{dc}R_{Load} \tag{27}$$

Combining Eq. 25 and Eq. 27 makes it possible to define $V_{SW_{low}}$ from known quantities:

$$V_{SW_{low}} = \frac{2V_{dd}R_{on}}{R_{load} + 2R_{on}} \tag{28}$$

Using the result of Eq. 28 in Eq. 25 the DC current I_{dc} is found to be:

$$I_{dc} = \frac{2V_{dd}R_{on}}{R_{load} + 2R_{on}} \frac{1}{2R_{on}} = \frac{V_{dd}}{R_{Load} + 2R_{on}} \tag{29}$$

Using the result of Eq. 29 and 26 the amplitude is found to be:

$$V_{pp}/2 = \frac{2V_{dd}R_{Load}}{2(R_{Load} + 2R_{on})} = V_{dd}\frac{R_{Load}}{R_{Load} + 2R_{on}} \tag{30}$$

Recalling that the fundamental harmonic has an amplitude that is $4/\pi$ larger than the amplitude of the square wave, the output power for the fundamental can be determined as:

$$P_{fund-R_{Load}-R_{on}} = \left(\frac{4}{\pi}\right)^2 \left(\frac{V_{dd}}{\sqrt{2}}\right)^2 \frac{1}{R_{Load}} \left(\frac{R_{Load}}{R_{Load} + 2R_{on}}\right)^2 \tag{31}$$

Which has been rearranged to show the output power without $R_{on} > 0$ and an adjusting factor of $R_{Load}/(R_{Load} + 2R_{on})$ for the amplitude. However, the analysis presented above could be repeated for any conduction angle to find the output power as a function of Φ, R_{Load}, and R_{on}. Unfortunately, simply combining the result in Eq. 23 and Eq. 30 will not give the correct result since the factor $2R_{on}$ in the denominator will depend on the conduction angle and must be modified accordingly, which for brevity is left out here.

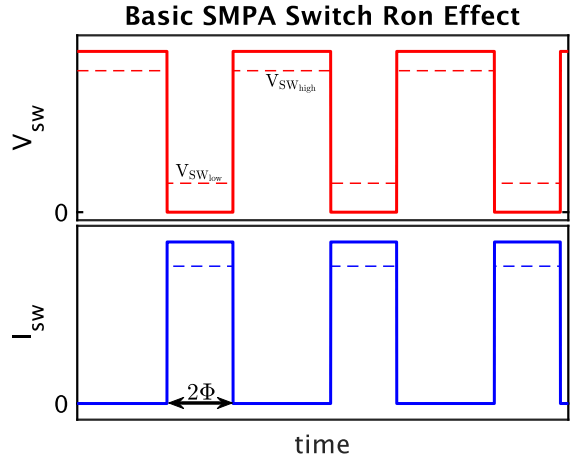

Figure 23: Idealized switch voltage and current for a basic SMPA, but with an on resistance in the switch

4.2.2 Class D

The class D aim is to solve the unwanted effect that power is also transferred to higher harmonics. When operating in voltage mode, it can be implemented as a push-pull structure as shown in Fig. 24. Figure 25 shows its drain voltage and currents. At the drain of the transistors, the class D PA has a square wave voltage waveform. This implies that the high-Q output resonance tank from Fig. 14 no longer is present. Instead an output matching network, acting as a filter, is placed between the transistor drain and the load, ensuring that only the wanted fundamental reaches the load [21], and not the odd order harmonics that composes the square-wave. The output power is found to be:

$$P_{R_{Load}} = \left(\frac{4}{\pi}\right)^2 \left(\frac{V_{dd}}{2}\right)^2 \frac{1}{2R_{Load}} \tag{32}$$

The $4/\pi$ is because the fundamental component of the Fourier expansion

of the square wave is $4/\pi$ times larger than the amplitude of the square wave. Having an identical expression for the DC power consumption, the efficiency becomes 100 %, which is expected, since in this example there is no overlap of voltage and current in the ideal switch, and only the fundamental frequency reaches the load. Another advantage of the PA operating in class D mode is its high normalized power output capability.

$$P_N = \frac{P_{out}}{v_{DS,max} i_{D,max}} = \frac{\frac{2V_{dd}^2}{\pi^2}/R_{Load}}{V_{dd}\frac{4V_{dd}}{2\pi}/R_{Load}} = \frac{1}{\pi} \approx 0.32 \tag{33}$$

However, since the push-pull class D PA uses two transistors, the normalized power output capability per transistor becomes:

$$P_N/Nbr_{transistors} = \frac{1}{2\pi} \approx 0.16 \tag{34}$$

As with all SMPAs the input amplitude information is lost when being subject to the switching mechanism. In conclusion, the class D amplifier has high efficiency and low device stress, but with extremely poor linearity. While this, at least to some extent, seems promising, the class D has some major drawbacks when operating at higher frequencies. Perhaps the main drawback is that the parasitic capacitance at the drain of the transistor is not absorbed into a resonance tank and thus has to be charged and discharged by the transistor every cycle [22], and as mentioned before, the limited frequency of operation as the transistors have to operate at frequencies well below f_T.

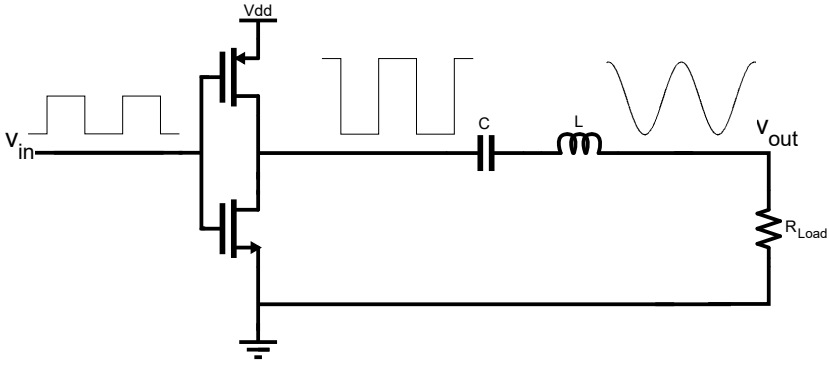

Figure 24: Omitting input resonance tank, a single ended, push-pull class D implementation of a PA.

Adding an effective switch resistance R_{on} somewhat changes the analysis. The high-Q resonance tank between the load and output stage forces the current through each transistor to become a rectified sinusoid. The drain-to-source

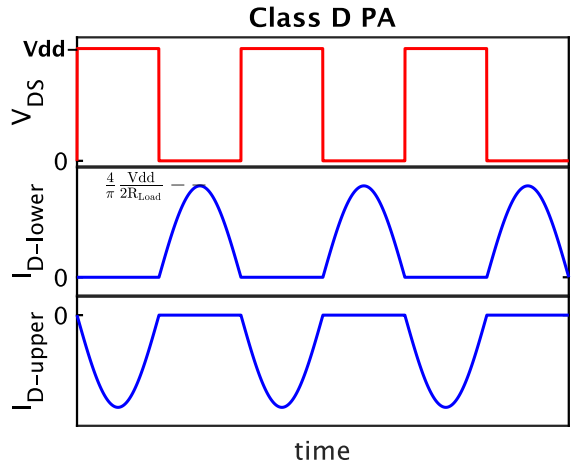

Figure 25: Idealized drain voltage and current for a class D biased PA.

voltage and current are adjusted according to Fig. 26. The voltage at the drain of each of the switches becomes a square wave with an overlaid opposite sign sinusoid. We perform a similar analysis as in section 4.2.1 with the target to identify the amplitude A_I of the rectified sinusoidal current through the bottom-switch. As before the DC-current has to flow through the switch during the half of the period that the bottom-switch is conducting. This results in

$$I_{DC} = A_I \left[\int_0^{T/2} 0\, dt + \int_{T/2}^{T} sin\, \omega t\, dt \right] = A_I \left[\frac{-cos\, \omega t}{\omega} \right]_{T/2}^{T} = \frac{A_I}{\pi} \tag{35}$$

A_I is the fundamental current amplitude in the load, which gives:

$$V_{out} = A_I R_{Load} \tag{36}$$

Using Eq. 36 and the fact that the fundamental voltage component, $A_I R_{on}$, present at transistor drains also will transfer to the output, gives:

$$A_I R_{Load} = \frac{4V_{dd}}{2\pi} - A_I R_{on} \tag{37}$$

Simplifying and solving for A_I gives:

$$A_I = \frac{4}{\pi} \frac{V_{dd}}{2R_{Load}} \frac{R_{Load}}{R_{Load} + R_{on}} \tag{38}$$

Where R_{Load} has been inserted in both the numerator and denominator so that the unspoiled fundamental load current $4V_{dd}/(2\pi R_{Load})$ becomes visible and the adjusting factor $R_{Load}/(R_{Load} + R_{on})$ is clear. This makes it possible to express the voltage amplitude of the opposite sign sinusoidal at the drains, which becomes:

$$A_V = A_I R_{on} = \frac{4V_{dd}}{2\pi} \frac{R_{on}}{R_{Load} + R_{on}} \tag{39}$$

From Eq. 38 the output power becomes:

$$P_{R_{Load}} = \left(\frac{4}{\pi}\right)^2 \left(\frac{V_{dd}}{2}\right)^2 \frac{1}{2R_{Load}} \left(\frac{R_{Load}}{R_{Load} + R_{on}}\right)^2 \tag{40}$$

The expression has been rearranged so that the adjusting factor falls out nicely. Comparing the adjusting factor for the output power of the class D PA when adding $R_{on} > 0$ with the corresponding expression for the basic RF SMPA, a factor of two in front of R_{on} differs. A simple explanation for this is that the basic RF SMPA only has one switch, whereas the push-pull structure of the class D PA has has two transistors, which makes the effective on resistance appear as half. Finally an expression for the efficiency is found by combining Eq. 35 for the DC current and Eq. 40 for the output power.

$$\eta_{R_{on}} = \frac{\left(\frac{4}{\pi}\right)^2 \left(\frac{V_{dd}}{2}\right)^2 \frac{1}{2R_{Load}} \left(\frac{R_{Load}}{R_{Load}+R_{on}}\right)^2}{\frac{V_{dd}\frac{4}{\pi}\frac{V_{dd}}{2R_{Load}}\frac{R_{Load}}{R_{Load}+R_{on}}}{\pi}} \tag{41}$$

which simplifies to:

$$\eta_{R_{on}} = \frac{R_{Load} + R_{on}}{R_{Load}} \tag{42}$$

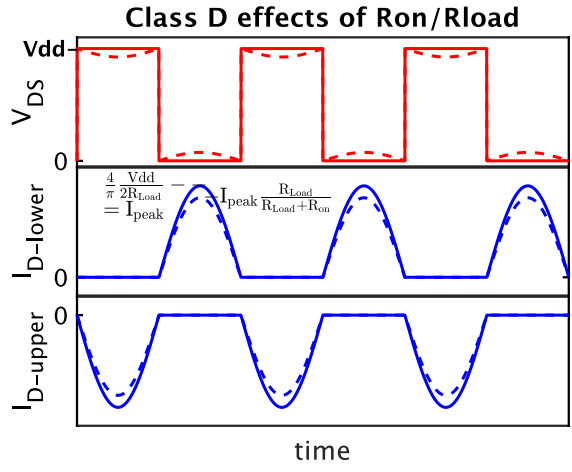

Figure 26: Ideal drain voltage and current for a class D PA (solid line), and when modelling non-zero on resistance, for both switches (dashed line)

4.2.3 Class D^{-1}

To overcome the problem with charging and discharging of the switching transistor output capacitances at higher frequencies, the inverse class D, or class D^{-1}, was first demonstrated at RF frequencies in 2001 by Kobayashi [23]. The class D^{-1} is best understood as a current mode class D amplifier, where the voltage and current waveforms are interchanged, i.e. the voltage across each transistor is now a half rectified sine wave and the current is a square wave. Figure 27 shows the schematic of the class D^{-1} amplifier and Fig. 28 shows its drain voltage and current. The parasitic drain capacitances are now absorbed into the parallel resonance tank, which due to its resonance creates zero voltage over the transistors at the time of the switching [23]. This creates the very desirable situation of zero voltage switching (ZVS), which greatly reduces losses as the drain parasitic capacitances no longer must be charged and discharged by the transistors at every cycle [22].

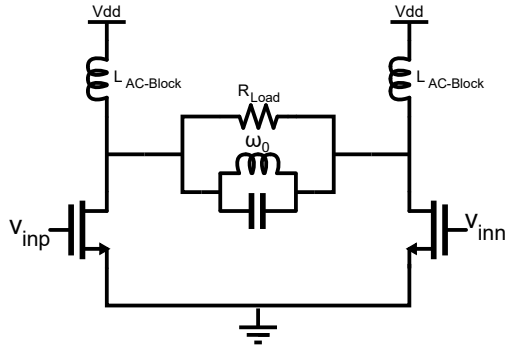

Figure 27: Schematic design of inverse class D implementation of a PA.

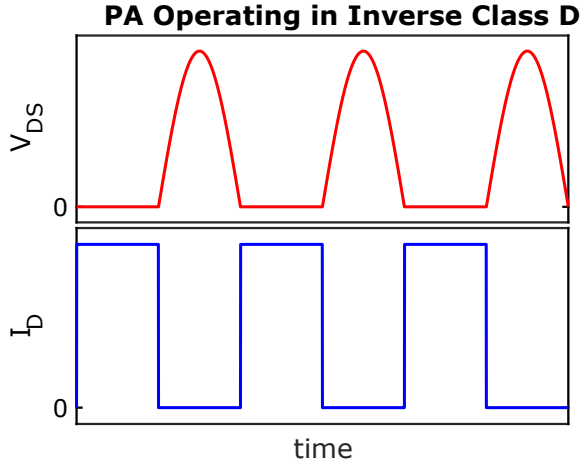

Figure 28: Idealized drain voltage and current for one of the two transistors in an inverse class D biased PA.

4.2.4 Class E

Perhaps the most troublesome design constraint for SMPAs is the limited frequency range of operation for which the switch on/off transition time is sufficiently fast, so that the conduction losses are small enough to achieve competitive efficiency. To mitigate this problem, the class E amplifier, invented in 1975 by Sokal and Sokal [24], uses a higher order output matching network, that ideally creates not only the ZVS as explained for the inverse class D, but also a zero derivative of the voltage, at the time that the switching occurs. ZVS is needed not to waste power by charging and discharging the capacitance at the output of the transistor acting as the switch. The criteria for the zero derivative of the voltage across the switch for the output matching network is called zero voltage derivative switching (ZVDS), and since $\mathrm{i} = \mathrm{C}_1 \frac{\mathrm{dV}}{\mathrm{dt}}$ it is required to avoid current through the switch, when going from the non-conducting to the conducting state [25]. Unfortunately, as the switch in the class E amplifier turns off, the current is close to its peak value, degrading the overall efficiency. Figure 29 shows an implementation of the class E SMPA. How to dimension the components in the output matching network is found in [14] and for convenience repeated here.

$$L_1 = \frac{QR_{Load}}{\omega} \tag{43}$$

$$C_1 = \frac{1}{\omega R_{Load}(\pi^2/4+1)(\pi/2)} \approx \frac{1}{5.447\omega R_{Load}} \tag{44}$$

$$C_2 \approx C_1 \frac{5.447}{Q}(1 + \frac{1.42}{Q - 2.08}) \tag{45}$$

The Q-value is determined by the bandwidth requirement of the system and L_1 and C_2 are tuned at the fundamental frequency and effectively prevents harmonic currents from reaching the load. Exact calculation of the transistor drain voltage and current is demanding, but Fig. 30 shows the typical corresponding voltage and current waveforms obtained from [14].

For ideal switching operation, a successful implementation of the output matching network guarantees a 100 % efficiency, but the whole idea of the class E amplifier is to get high efficiency even with switches operating at frequencies where they are far from ideal. This has proven to be a difficult task due to for instance the aforementioned high switch current at turn off. For completeness the output power of a class E PA is found to be [14]

$$P_{R_{Load}} = \frac{V_{dd}^2}{R_{Load}} \frac{2}{1+\pi^2/4} \tag{46}$$

The class E PA is rather sensitive to device stress due the the very high peak voltages and currents. A detailed analysis shows that $\mathrm{V_{DS_{pk}}} \approx 3.6\mathrm{V_{dd}}$ and $\mathrm{I_{DS_{pk}}} \approx 1.7\mathrm{V_{dd}}/\mathrm{R_{Load}}$ [14], which gives the normalized power output capability

$$P_N = \frac{P_{out}}{v_{DS,max} i_{D,max}} \approx \frac{\frac{V_{dd}^2}{R_{Load}} \frac{2}{1+\pi^2/4}}{3.6 V_{dd} \frac{1.7 V_{dd}}{/R_{Load}}} \approx 0.094 \tag{47}$$

which is even lower than the class A and class B PAs, and much lower than the other SMPAs.

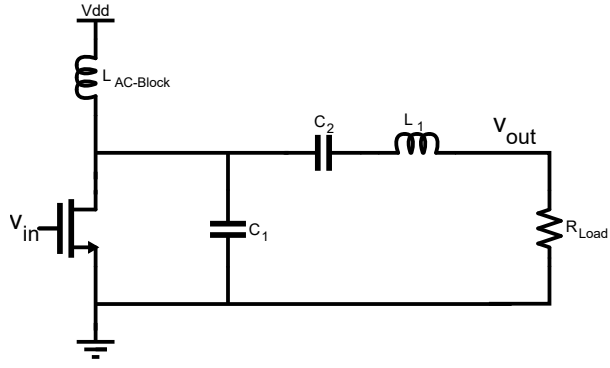

Figure 29: Schematic of the class E amplifier.

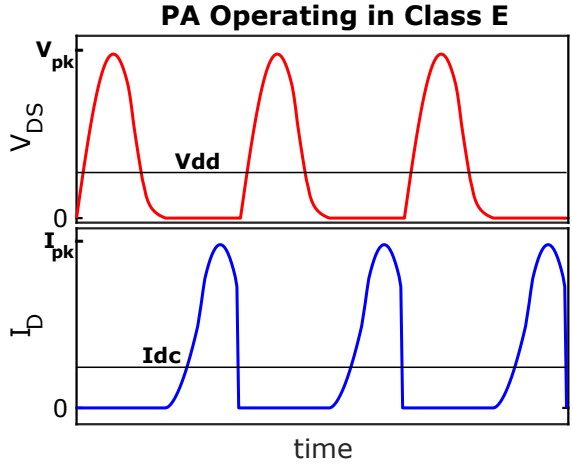

Figure 30: Drain voltage and current for a PA operating in class E.

4.2.5 Class F

Analogous to the class E, the class F PA exploits further the possibilities in shaping the output voltages in a desirable fashion. The schematic of a class F amplifier, depicted in Fig. 31 has a transmission line placed between the tuned load and the transistor. With the exception of the fundamental frequency, the output resonance tank produces a low impedance at all harmonics. Recall that a $\lambda/4$ transmission line works as an impedance inverter, effectively by a rotation of π in the Smith chart. At the fundamental frequency the impedance seen from the drain is simply R_{load} since the $\lambda/4$ transmission line is perfectly matched. For even harmonics $2N\omega_0$ the $\lambda/4$ transmission line appears as a $2N\lambda/4 = N\lambda/2$ transmission line, which does not invert the load impedance and makes the impedance seen from the drain at even harmonics appear as a short. Contrary, for odd harmonics the $\lambda/4$ transmission line still appears as an impedance inverter, inverting the short presented by the output resonance tank to an open circuit, as seen from the drain. The transistor is driven hard enough to behave as a switch that in turn drives a load that is a short circuit for even harmonics and an open for odd, which leads to a square wave voltage appearing at the drain of the transistor. The drain voltage and current can be seen in Fig. 32. There is no voltage to current overlap and no harmonic energy is dissipated in the load, so the efficiency will be 100 %, under the assumption of a lossless output matching network and an ideal switch. However, one disadvantage of the class F power amplifier is that it does not fulfill the ZVS requirement, which is highly disadvantageous when operating at higher frequencies where the switch is far from ideal. But the output matching network shapes the drain voltage better than for class D, and it does not exhibit as high peak voltages as the class E.

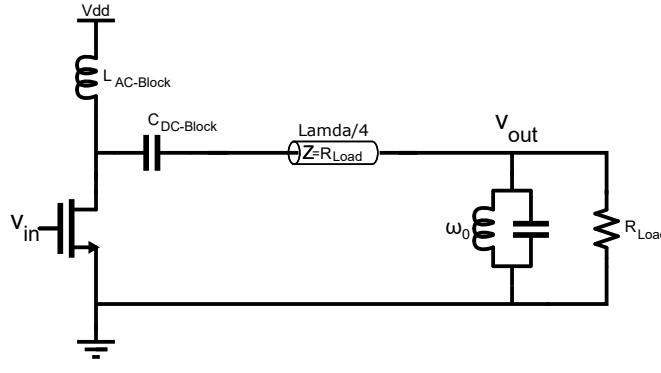

Figure 31: Schematic of the class F amplifier.

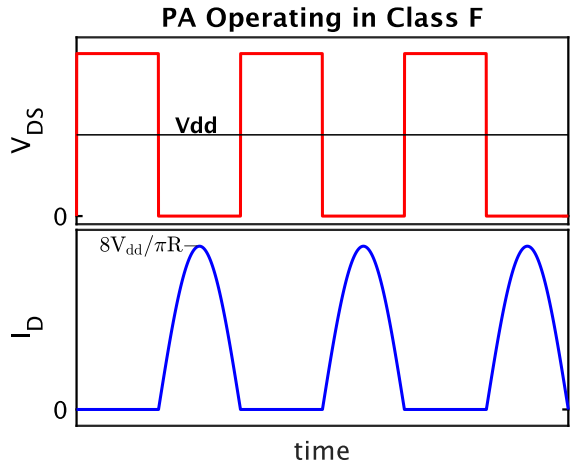

Figure 32: Drain voltage and current for a PA operating in class F.

Repeated from [14], the output power and normalized power handling capability of class F are:

$$P_{R_{Load}} = \frac{(4/\pi V_{dd})^2}{2R_{Load}} \tag{48}$$

$$P_N = \frac{P_{out}}{v_{DS,max} i_{D,max}} = \frac{\frac{(4/\pi V_{dd})^2}{2R_{Load}}}{2V_{dd}(\frac{8}{\pi}\frac{V_{dd}}{R_{Load}})} = \frac{1}{2\pi} \approx 0.16 \tag{49}$$

As for the class D amplifier, the output power of the class F is higher than $V_{dd}^2/2R_{Load}$, since the fundamental component of the Fourier expansion of the square wave is $4/\pi$ times larger than the amplitude of the square wave.

4.2.6 Class F^{-1}

Similar to the class D and inverse class D, the inverse class F amplifier has interchanged the waveforms of the voltage and current compared to class F, with the goal to achieve ZVS that enables efficient operation at higher frequencies. The interchange of current and voltage waveforms in the inverse class F amplifier is achieved by creating an open circuit for the even harmonics and a short circuit for the odd. A transmission line based implementation of an inverse class F amplifier is depicted in Fig. 33, and a lumped component based version is shown in Fig. 34. The corresponding voltage and current waveforms are shown in Fig. 35.

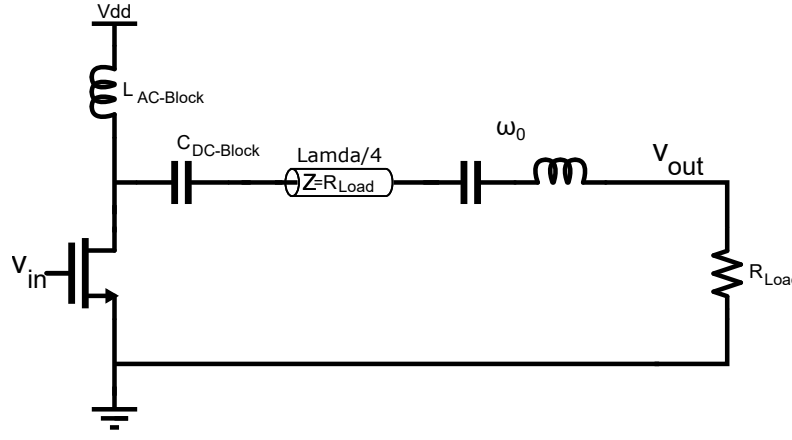

Figure 33: Schematic of a transmission line based inverse class F amplifier.

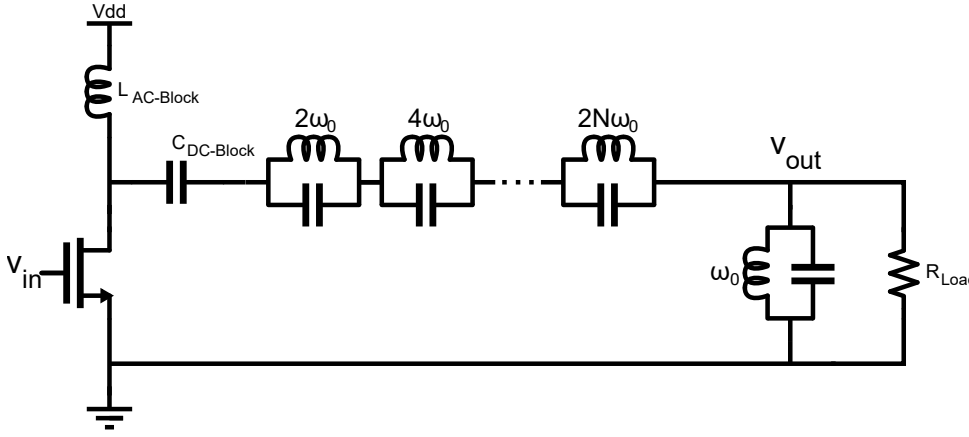

Figure 34: Schematic of lumped component based inverse class F amplifier.

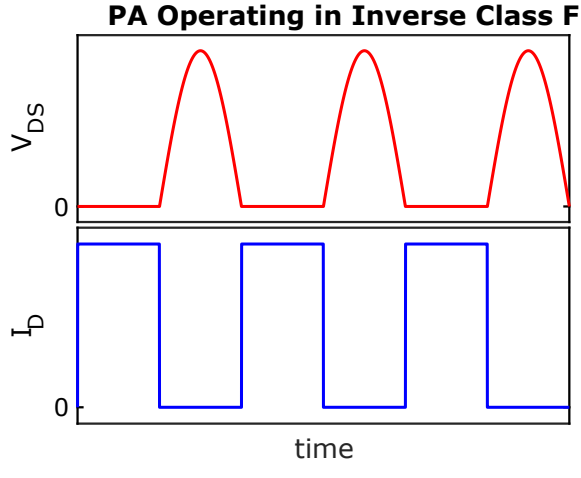

Figure 35: Drain voltage and current for a PA operating in inverse class F.

4.3 PA Classes Comparison Table

Table 4 presents a comparison of the four most important performance aspects Linearity, Efficiency, Gain, and Output Power for the different PA classes that were analysed in this chapter.

Table 4: Comparison of key parameters for analysed PA Classes

PA Class	**Linearity**	**Max Efficiency** [%]	**Gain**	**Output Power** P_N[%]
A	Best	50	Best	12.5
AB	Good	50-78.5	Good	12.5
B	Ok	78.5	Ok	12.5
C	Poor	78.5-100	Poor	< 12.5
D	Poor	100	Poor	16.0
$\mathbf{D}^{-1}$	Poor	100	Poor	16.0
E	Poor	100	Poor	9.4
F	Poor	100	Ok	16.0
$\mathbf{F}^{-1}$	Poor	100	Ok	16.0

Chapter 5

Power Amplifier Linearity

One of the key parameters of amplifiers in general is the linearity, and especially so for power amplifiers, since they must handle large signal excursions, which inevitably compromise signal fidelity by creating distortion. Linearity determines how well the input signal is reproduced at the output. Apart from scaling the signal magnitude and adding a fixed time delay, a PA would be free of distortion if the output signal is identical to the input signal [26]. This means that without distortion the amplifier copies its input to the output, and the only thing the amplifier does, is to scale up the signal level and add a constant time delay [27]. In addition to the wanted linear amplification of the input signal, all PAs will also, to some degree, distort the signal through their nonlinearities. The nature of the nonlinearities determines how the signal is distorted, and can thereby greatly affect the output signal characteristics. Distortion not only affects the fundamental frequency, it also moves energy in the frequency domain by producing higher order harmonics and intermodulation products, or spectral regrowth of a modulated signal. This chapter covers basic theory of nonlinearities and memories in electronic systems, their respective sources, how to improve the linearity, and finally how to reduce the effects of nonlinearities.

5.1 Time Invariant and Time Variant Systems

Electronic systems can be either time-invariant or time-variant. Time-invariant means that the system does not change with time, such as an amplifier. A time-variant system on the other hand changes its transfer function with respect to time. An example of a time variant system is a mixer. Both time-invariant and time variant systems can be linear and nonlinear. A linear time-invariant (LTI) system does not produce any new frequency components in the output signal that were not present in the input signal, whereas a linear time-variant (LTV) system does. The remainder of this chapter does not deal with time-variant systems, and all systems (amplifiers) are considered time-invariant from now on.

5.2 Classification of Time Invariant Electrical Systems

All systems can be classified into four types, depending of linearity and memory of the system. In table 5 below the four types are shown together with an example of components that would comprise such a system [27].

Table 5: Classification of Electrical Systems

	Memoryless	**With Memory**
Linear	**Instantaneous Linear** Linear Resistance	**Stationary Linear** Linear Capacitance/Inductance
Nonlinear	**Instantaneous Nonlinear** Nonlinear Resistance	**Stationary Nonlinear** Nonlinear Capacitance/Inductance or Nonlinear Resistance and Linear Capacitance/Inductance

5.2.1 Linear Without Memory -Instantaneous Linear System

A system is linear if the output signal is proportional to the input signal, regardless of the amplitude of the input signal. If not, then the system is nonlinear. Expressed mathematically a linear system must fulfil equation 50 and 51 below [28].

$$f(x+y) = f(x) + f(y) \tag{50}$$

$$f(\alpha x) = \alpha f(x) \tag{51}$$

Thus, the time domain output voltage of a linear electronic system can be expressed as in equation 52, where k_1 is the proportionality, or gain of the system. Such a system is referred to as a memoryless linear system, or instantaneous linear system.

$$v_{out}(t) = k_1 v_{in}(t) \tag{52}$$

Practical systems considered linear will still have a range of input signals where they behave linear. Outside this range the system will behave nonlinearly.

5.2.2 Nonlinear Without Memory -Instantaneous Nonlinear System

For nonlinear systems the proportionality is broken and the transfer function has to be expressed by a nonlinear function. The by far most common is to use a polynomial expression, which is an effective way to describe nonlinear systems. It allows the designer to understand the origin of, and to model, the most important wanted and unwanted functions of a PA. Consider a third order Taylor power series expansion representing a nonlinear time domain, memoryless model of a PA:

$$v_{out}(t) = k_1 v_{in}(t) + k_2 v_{in}^2(t) + k_3 v_{in}^3(t) \tag{53}$$

As for the linear system, the wanted linear amplification is the coefficient k_1, but k_2 and k_3 determine to what extent the system produces second and third order distortion, respectively. Naturally, more higher order terms in the Taylor series expansion can be added to increase the accuracy of the transfer function of the system. Conversely to the linear system, the nonlinear system can be seen as linear for an infinitesimal small input signal, as the higher order nonlinear terms go faster towards zero than the linear term. This phenomenon is frequently used when performing a small signal excitation/simulation of a nonlinear circuit[1]. An electrical system that is described by Eq. 53 is called instantaneous and nonlinear. An important way to characterize a nonlinear system is to measure its response to harmonic stimuli. A two tone input signal, to a third order nonlinear system, produces an output signal containing various frequencies. In Appendix A the frequency spectrum response from a two tone test of such a system is enclosed.

5.2.3 Linear With Memory -Stationary Linear System

If the output of the system not only depends on the instantaneous value of the input signal, but also on previous values, then the system has memory. The Volterra-Wiener approach makes it possible to describe both linear and nonlinear stationary systems using multi variable convolution integrals. To model a linear system with memory, we start with equation 52 representing the memoryless linear system, but now modified to equation 54 below [29] and [30].

$$v_{out}(t) = \int_{-\infty}^{\infty} h_1(\tau_1) v_{in}(t - \tau_1) \cdot d\tau_1 \tag{54}$$

[1]Small signal simulation, or AC simulation, can be seen as a linearisation around an operating point of a nonlinear circuit. The small signal simulation is completely linear, and thus the output from such a simulation only contains the same frequency as the input signal.

The first term inside the convolution integral, h_1, is the impulse response or Volterra kernel of the system, which for causality has to be zero for $\tau_1 < 0$. This reduces the lower integration boundary to zero instead of minus infinity.

$$v_{out}(t) = \int_0^\infty h_1(\tau_1) v_{in}(t-\tau_1) \cdot d\tau_1 \tag{55}$$

Many methods exist to estimate the Volterra kernel coefficients [31], but unfortunately this is rather complicated and goes beyond the scope of this work. However, for an instantaneous system, the Volterra kernel $\mathrm{h_1}$ becomes $\mathrm{h_1}(\tau) = \mathrm{k_1} \cdot \delta(\tau)$, where $\mathrm{k_1}$ is a real valued constant, which then reduces the convolution integral to $\mathrm{v_{out}(t) = k_1 v_{in}(t)}$.

5.2.4 Nonlinear With Memory -Stationary Nonlinear System

Similar to the linear stationary system the Volterra-Wiener approach is used to model a nonlinear system with memory, but now adding the nonlinear parts which leads to a multidimensional convolution. From [27], [31], and [32] the ready made result is acquired, but modified here to only cover causal systems.

$$\begin{aligned} v_{out}(t) = & \int_0^\infty h_1(\tau_1) v_{in}(t-\tau_1) \cdot d\tau_1 \\ & + \int_0^\infty \int_0^\infty h_2(\tau_1,\tau_2) v_{in}(t-\tau_1) v_{in}(t-\tau_2) \cdot d\tau_1 d\tau_2 \\ & + \int_0^\infty \int_0^\infty \int_0^\infty h_3(\tau_1,\tau_2,\tau_3) v_{in}(t-\tau_1) v_{in}(t-\tau_2) v_{in}(t-\tau_3) \cdot d\tau_1 d\tau_2 d\tau_3 \end{aligned} \tag{56}$$

The first convolution term in the series is the linear response, which was the complete representation of the linear stationary system. Then follows the terms that model the second and third order nonlinear effects. The use of such a model can be understood by showing its output from a known input signal, for instance a single or a two tone test. Unfortunately, due to the rather lengthy calculations this is left out here, but the interested reader can follow the steps in [27], and also find other relevant sources. However, what can be said about the output of a two tone stimuli of a nonlinear system with memory, is that the frequency response, and particularly the intermodulation products, will show a dependency of the tone separation, i.e. the bandwidth of the input signal.

5.3 Sources of Nonlinearities and Memories in Electronic Systems

In reality all systems, and thus all PAs, are both nonlinear and with memory. Below follows a brief explanation of the main sources of nonlinearties and

memories in PAs.

5.3.1 Sources of Nonlinearities

The dominating source of nonlinearites is the active components, here assumed to be CMOS transistors. A CMOS transistor has many sources that contribute to its nonlinearity. Both the wanted transconductance and the unwanted parasitics are affected. For most analog electronic circuits, the two dominant sources of nonlinearty, both originating from the CMOS transistor, are the nonlinear transconductance g_m and the nonlinear output conductance g_d [33], but for PAs it is typically the nonlinear transconductance that will limit the linearity. The nonlinear output conductance has a relatively larger impact on the distortion of the output voltage when the transistor drives a high impedance load, which is normally not the situation under which the transistors in a PA operate. However, for large output signals near compression, the nonlinear behaviour of the output conductance will start to affect the overall linearity.

Nonlinear Transconductance

The underlying physical effects in the semiconductor, that often dominate the nonlinearity of the transconductance of the high frequency PA, are two short channel effects, reduced effective mobility and velocity saturation, which are further described in [34]. Furthermore, as discussed in chapter 4, PAs can operate under a wide range of bias levels and input signal magnitudes. This will cause the CMOS transistor(s) in the PA to operate in weak to moderate to strong inversion of the channel. The models describing the transistor's I-V characteristics is generally accurate for strong and weak inversion, but not as well defined for moderate inversion. To overcome this problem, smoothing functions are used to interpolate the transition between the weak and strong inversion regions, [34] describes this in more detail.

Figure 36 shows the drain-to-source current and its first, second, and third derivatives as a function of V_{gs} for a constant V_{ds}. The first derivative is the wanted linear transconductance term g_m or g_{m1}. The second and third derivative, g_{m2} and g_{m3}, determine how much second and third order distortion that is produced by the transistors. Notice how g_{m3} changes sign and crosses zero as V_{gs} increases. This phenomenon can be exploited either by targeting the bias level that minimizes g_{m3} or by having two parallel transistors (or complete amplifiers) with different operating points, one with positive g_{m3} and one with negative g_{m3} as in [35] and more recent [36]. Unfortunately this technique seems to struggle with linearity when exposed to different signal levels [20]

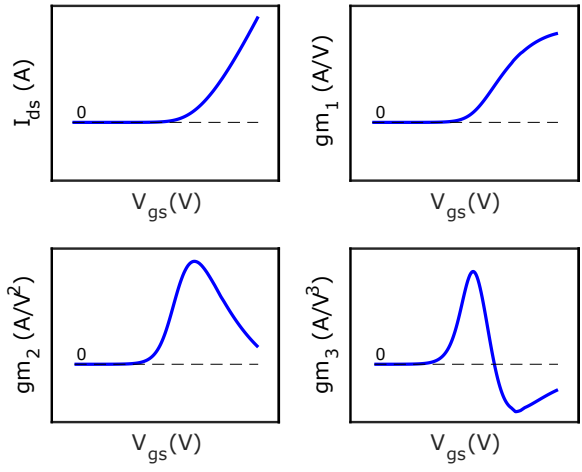

Figure 36: I_{ds} and its first, second and third derivative, i.e. g_{m1}, g_{m2}, and g_{m3} plotted vs. V_{gs}, for a short channel low-V_{th} NMOS transistor.

Nonlinear Output Conductance

As mentioned above, the nonlinear output conductance g_{ds} that can be a major contributor to the distortion for short channel transistors driving a high impedance load, is predominantly explained by drain induced barrier lowering (DIBL) which is explained in [34]. Effects of a nonlinear output conductance become less prominent if the PA drives a low impedance, since the output voltage (and current) will then be less affected by changes in transistor output conductance. One might say that the nonlinear output conductance is hidden by a linear low impedance loading at the output. Nevertheless, for large output signals, where the PA operates close to compression and the transistor enters the triode region, g_{ds} becomes much higher and changes much more rapidly (w.r.t. V_{ds}) than in the saturation region, and it will then affect the overall linearity.

Nonlinear Parasitic Capacitances

Another important source of distortion is nonlinear parasitic capacitances in the transistors, especially the gate to source capacitance C_{gs}. A clear example of this is when C_{gs} is the dominating capacitance in an LC resonance tank at the input of the PA. It can then significantly distort both amplitude and phase of the input gate signal voltage, since the amplitude of the input signal will

modulate C_{gs}. C_{gs} is defined as the ratio of the change in charges at the gate to a small change in voltage at the source, for constant gate, drain, and body voltages [34]. Without going into details, the changes of charges at the gate is very much dependent on the region of operation of the transistor. Holding V_{ds} constant and gradually increasing V_{gs}, the transistor will go through all regions, i.e. accumulation, depletion, weak inversion, moderate inversion, strong inversion saturation, and strong inversion nonsaturation [34]. Figure 37 shows extracted results of a DC simulation for C_{gs} when gradually increasing V_{gs}. For a constant bias level, one might think that C_{gs} should stay constant, but a large amplitude modulated signal on top of the bias level will significantly modulate the value of C_{gs}.

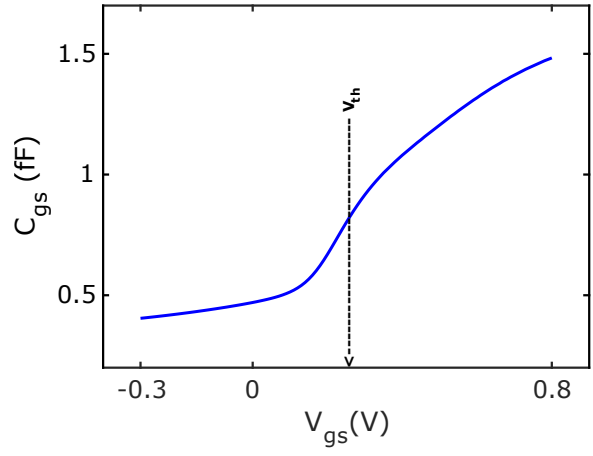

Figure 37: Simulated C_{gs} plotted vs. V_{gs} for a small short-channel low-V_{th} NMOS transistor ($W/L = 2.4\,\mu m/20\,nm$) with $V_{ds} = 0.8\,V$

Passive Components

Passive components, i.e. resistors, inductors, and capacitors, such as polysilicon resistors and metal inductors and capacitors, are inherently very linear. However, due to for instance temperature effects, such as self-heating, they can still exhibit nonlinear behaviour, but their contribution to the distortion is likely to have a very limited effect on the overall performance and therefore they are typically treated as linear.

5.3.2 Sources of Memories

Means of remembering previous states, i.e. memories, in electronic circuits can be either electrical, thermal or mechanical[2]. Normally, it is only the first two that are relevant for PAs and therefore description of mechanical memories is disregarded here.

Electrical Memories

The energy stored in the electric and magnetic fields in capacitors and inductors, respectively, acts as a memory in an electronic system. The voltage over a capacitor and the current through an inductor depends on all previous values of the current and voltage, respectively. This is expressed in equation 57 and 58 below.

$$v_C(t) = \frac{1}{C}\int_{-\infty}^{t} i(t') \cdot dt' \tag{57}$$

$$i_L(t) = \frac{1}{L}\int_{-\infty}^{t} v(t') \cdot dt' \tag{58}$$

Capacitors and inductors store information about previous states and together with resistors they produce different time constants that determine how long time it takes for the circuit to forget old states. Naturally these time constants have a large impact on the memory effects, and thereby the distortion of the circuit. Moreover, they play a key role when attempting to linearise a PA. More of this in section 5.4 and 5.5.

Actual implemented PAs are not comprised of a single nonlinear mechanism as modelled by equation 53. Instead they contain multiple interconnected or cascaded nonlinear processes. This means that they can produce combined nonlinear distortion of different order than the individual nonlinearities. Concentrating on third order intermodulation distortion, IM_3 components can be produced not only from k_3, but also by cascading quadratic nonlinear effects (k_2) e.g. a baseband IM_2 component from a first quadratic nonlinearity mixing with the RF signal in a second. Since memory effects are defined as bandwidth dependent nonlinear effects [27], it becomes clear that frequency dependent node impedances, at the baseband, the fundamental frequency, and the second harmonic are the reason for how electrical memory effects arise, as the PA contains multiple nonlinear mechanisms. The non-constant node impedance naturally has both real and imaginary parts, which links back to equation 57

[2] Also worth mentioning, even though perhaps irrelevant for PAs, is that in a crystal oscillator, memory is stored in the kinetic and potential energy in a vibrating piezoelectric crystal. The piezoelectric crystal then links the mechanical and electrical systems. Furthermore, the mechanical system can be represented by the use of electrical components, thereby giving the vibrating crystal an equivalent electrical model as in [5]

and 58, and also explains how a frequency dependent node impedance effectively creates a memory of prior signals of the PA. Of the three node impedance frequencies of interest, the baseband (envelope), the fundamental frequency, and the second harmonic it is undoubtedly the impedance at the baseband frequency that is the most problematic to keep constant [27]. The baseband frequency band goes from DC to the maximum modulation frequency, which gives a fractional bandwidth equal to two. At the fundamental and second harmonic the fractional bandwidth is much lower and the node impedances can then be held much more constant. Narrowing down the memory effects contribution even further, it is bias impedances at the baseband frequencies that is the origin of the majority of the electrical memory effects [27]. An exception would be if harmonic traps are used, since harmonic traps are narrowband and can create large impedance variations at for instance the second harmonic, which may thus result in increased memory effects.

To summarize, the different frequency components originate from combinations of different distortion orders, that arrive at the same frequency. One example of this is seen in table 10 in Appendix A. When combining bandwidth dependent nonlinear effects, memory effects with multiple time constants may result. Figure 38, reproduced from [37], shows a vector representation of the constituents of IM_3 under the influence of memory effects. For instance the phasor of the 2^{nd}-order envelope contribution will change magnitude and phase with input signal tone separation, depending on baseband impedance variations with frequency, causing memory effects.

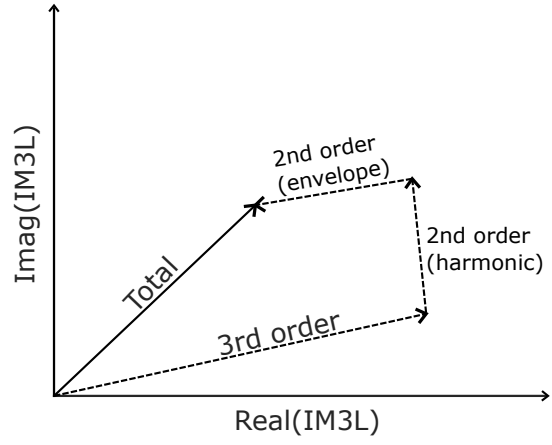

Figure 38: Illustrative constituents of complex IM_3 under the influence of memory effects arising from cascaded second order nonlinearities.

Thermal Memories

The mechanism through which thermal memories interact with the electrical system is that the behaviour of electrical components, primarily the transistors, depend on the temperature. As the PA dissipates power in the form of heat, due to its non-ideal efficiency, it heats itself and the surroundings. Thermal energy is stored in the mass, for instance in the silicon substrate close to the PA transistors. The temperature, and the distribution of it, then represents a memory of preceding power dissipation. Similar to electrical memories, time constants for how fast the temperature rises and cools off exist. Time constants, for different parts of the chip, depend on material constants such as heat capacity and conductivity, transistor layout on the silicon die, the die thickness, the package, the PCB, and the heat sink. The power dissipation in the transistor is simply the drain to source voltage multiplied with the drain to source current: $P_{diss} = V_{ds} \cdot I_{ds}$. Both physical quantities (voltage and current) represent the fundamental signal, and since they are multiplied with each other, the frequency spectrum of the dissipated power follows a second order distortion and thus includes DC, baseband, and second harmonics [27] and [38]. Heating of the silicon surface reacts faster, due to the presence of the self heating transistor at surface, than deeper in the silicon substrate. This makes thermal memory effects more prominent due to self heating than due to heating from nearby components [38]. In addition, bias circuits, built to compensate for thermal changes, for instance constant gm biasing, cannot react fast enough to compensate for the fastest temperature changes at the surface due to self heating. Obvious, since the transistor of the bias circuit will have a different junction temperature than that of the transistors in the PA. These effects can have bandwidths up to 1 MHz range [38]. In addition to the relatively fast memory effects due to self heating, the PA will suffer from memory effects with much longer time constants that may deteriorate the performance. For time division duplex (TDD) systems, where the PA turns on and off between transmit and receive slots, these memory effects can be particularly troublesome.

5.4 Improving Linearty

The easiest way to improve linearity is to reduce the output power, i.e. to operate at back-off. This technique works fairly well with most PAs, but it is easiest understood for typical class A transconductance PAs, where the transconductance will be more constant as the signal current is reduced compared to the bias, and the signal voltage will be further from the maximum where the transistor enters the triode region. However, operating a PA at back-off has a severe impact on efficiency, which is illustrated in Fig. 20. This leads us to the efficiency-linearity trade-off that the PA designer faces and that has resulted in the development of various linearity enhancements techniques, of which the

most well-known are presented below.

5.4.1 Linearity Enhancements Techniques

Polar Loop Feedback

The polar loop feedback linearisation technique is used to linearise polar modulated transmitters, which operate with two input signals to the PA. One input signal represents the phase information and the other represents the amplitude. The idea of linearisation through polar loop feedback is to sense the signal at the output of the PA, downconvert it and then compare it with the input signal. The comparison has to be done both for amplitude and phase, and each of these results in a signal representing the error. The error signals, from the amplitude and phase comparisons, are then used to compensate the amplitude and phase of the input signal to the PA. A polar loop feedback system is shown in Fig. 39. The loop gain and bandwidth of the negative feedback loop will determine how much the distortion can be suppressed and for how wide signal bandwidths. Unfortunately, the polar loop feedback technique has many disadvantages, such as problems measuring phase differences and correcting them at high modulation frequencies and long loop delays in downconversion, causing low stable bandwidth. With current techniques, as presented here, the polar loop is not suitable for high bandwidth systems such as 5G. With respect to linearisation of memory effects, the polar loop feedback technique should work well, as long as the bandwidth of the memory effects are below the bandwidth of the feedback loop. This is likely to be true for at least thermal memory effects. The relatively shorter time constants for the electrical memory effects are more probable to be outside the feedback loop bandwidth and will thereby not be corrected [20] and [27]. For PAs suffering from significant distortion due to memory effects it could even mean that the correction signal in some cases could deteriorate the linearity.

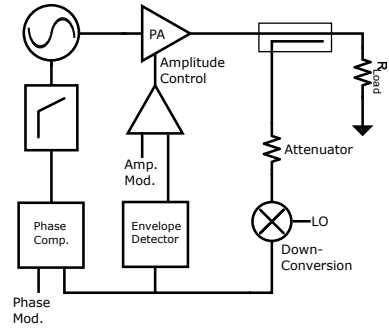

Figure 39: A polar loop negative feedback PA linearisation system.

Cartesian Loop Feedback

The Cartesian feedback loop linearisation technique is used to linearise Cartesian modulated transmitters (IQ-transmitters), which operate by upconverting a baseband signal represented in IQ-domain to carrier frequency through an IQ-modulator (IQ-mixer). The Cartesian loop feedback senses the output after the PA and demodulates it through another IQ-mixer to I and Q baseband signals. The downconverted I and Q baseband signals are then compared with the input I and Q signals and the residuals are amplified and used as input signals to the PA. A Cartesian loop feedback system is shown in Fig. 40. As for the polar loop, the loop gain and bandwidth of the negative feedback loop will determine how much the distortion can be suppressed and for how wide bandwidths. The Cartesian loop feedback technique also suffers from long delays in downconversion causing low bandwidth, which also makes it unsuitable to linearise high bandwidth modulations. Similar to the polar loop, linearisation of memory effects should work well as long as the loop bandwidth is sufficient, [20] and [27].

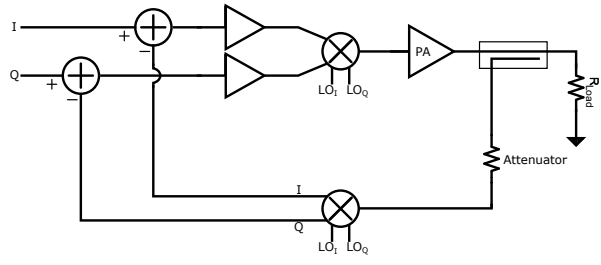

Figure 40: A Cartesian feedback PA linearisation system.

Feedforward

The feedforward linearisation technique, depicted in Fig. 41, adds the linearisation signal at the output of the amplifier, instead of at the input as feedback does. The input signal is fed to a main amplifier, whose task is to deliver the majority of the output power. An attenuated version of the output signal from the main amplifier is then compared with a delayed version of the input signal. The result from the comparison, which ideally only contains a replica of the distortion added by the main amplifier, but inverted, is then amplified by an auxiliary amplifier. Finally a delayed version of the output signal from the main amplifier is combined with the output from the auxiliary amplifier. Ideally this should produce a perfectly linear output signal. The feedforward linearisation technique is rather simple to understand and seems very attractive at first. It clearly has a major advantage over linearisation techniques based on negative

feedback, which is its ability to handle larger signal bandwidths, even though the bandwidth will still be limited by the bandwidth of combiners and phase shifters (time delays). Nevertheless, the feedforward suffers from some severe drawbacks. The first, and perhaps an obvious one, is that since the linearisation process happens at the output, where the signal is the strongest, the auxiliary amplifier also has to handle large signals, which naturally increases its power consumption. Secondly, the process of combining signals, and especially high power signals, is costly in terms of loss. Moreover, signal leaking back from the main amplifier to the auxiliary amplifier due to limited isolation in the combiner risks degrading the linearity of the auxiliary amplifier and thereby introducing distortion. Even though feedforward linerisation does not have a negative feedback loop it still senses the distortion added by the main amplifier, which has to be assumed to dominate the nonlinearities, which makes it possible for the feedforward to track and cancel memory effects of the main amplifier. However, the efficiency of the cancellation will be limited by the frequency response of the auxiliary amplifier and passive components, [20] and [27].

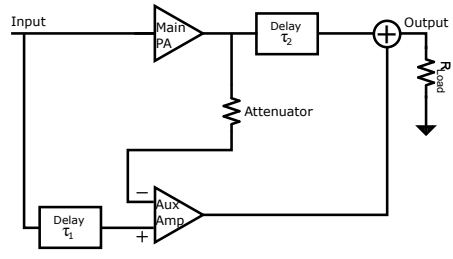

Figure 41: A feedforward error correction technique.

Other Linearisation Methods

In addition to the above mentioned linearization methods many other techniques exist. It is worth mentioning a few of them, and briefly describe how they work.

Adaptive Bias

The adaptive bias technique adjusts the bias level of the PA with the level of the input signal. It is practically demonstrated in [2–4, 39–49] and also analytically investigated in [4]. As shown in Fig. 36, the transconductance g_m or (g_{m1}) reduces with increasing input signal V_{GS} above a certain input signal level. To counteract the transistors compressive behaviour for large input/output signals, the bias signal can be gradually increased. Figure 42 shows measured

output power and the gate bias voltage of the CS input transistors of a mmW power amplifier versus input signal level, with and and without the use of adaptive bias, and Fig. 43 shows DE and AM-AM for the same measurement. Both Fig. 42 and 43 are reproduced from [2]. The results clearly show that an increase of the CS bias voltage with signal level linearises the amplitude of the output signal, as the adaptive bias with only minor gain expansion significantly increases the 1 dB compression point. Furthermore, since the circuit has a limited maximum input signal level, the saturated output power is also increased.

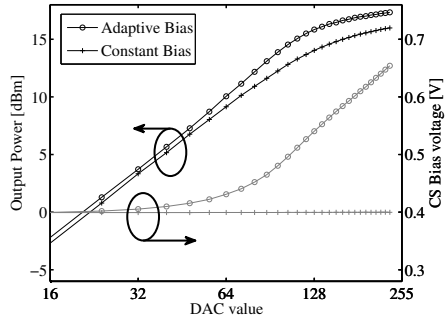

Figure 42: Output power and input CS transistor bias voltage as a function of a digital input CW signal at 27 GHz, reproduced from [2].

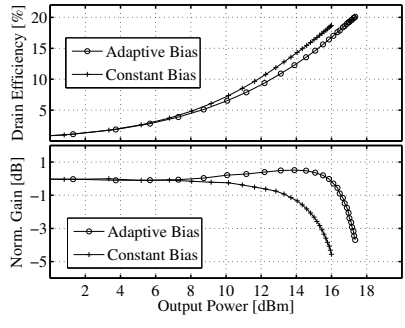

Figure 43: Drain efficiency and normalized gain vs. output power for CW signal at 27 GHz, reproduced from [2].

As shown, for a CW signal, adaptive bias can successfully reduce AM-AM, but the effects on AM-PM are not investigated here. Moreover, a major chal-

lenge with adaptive bias is to well enough follow the envelope of a high bandwidth modulated signal. If that cannot be done, the improvements brought by the adaptive bias will be reduced and it will also add memory effects to the PA, since the bias level of the PA will then depend on the envelope of previous input signals. In [3,4] this is analyzed and successfully demonstrated, however, from the perspective of deploying the adaptive bias in the auxiliary path of a Doherty amplifier.

Derivative Superposition

Derivative superposition, also called G_{m3} cancelling, is a technique that uses parallel structures, transistors or whole amplifiers, biased differently so that the G_{m3} products cancel due to having opposite signs. A thorough analytical paper describing the theory behind the technique is [50], and two examples of successful circuit implementations can be found in [35] and [36].

5.5 Reducing Effects of Nonlinearty by Predistortion

Distortion is a deterministic error, which, if known, makes it possible to compensate for it by adjusting the input signal. Adjusting the input signal so that it compensates for nonlinearities is called predistortion, which can be realized using both analog and digital techniques. The basic idea is to predistort the input signal with the inverse non-linearity of the PA as depicted in Fig. 44.

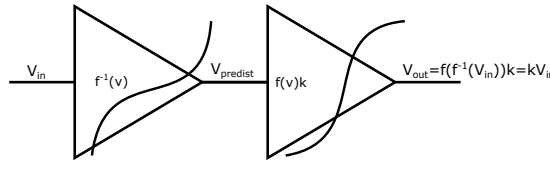

Figure 44: A nonlinear predistorted PA.

5.5.1 Analog Predistortion

Analog predistortion takes the linear input signal, either in base band or in RF domain, and predistorts it using analog techniques. Naturally this requires knowledge of the non-linearity of the PA. Analog predistortion should preferably compensate for both AM-AM and AM-PM of the PA, which can be very challenging. A very simple example on how to realize this would be to use a nonlinear resistor, implemented by a diode, in combination with a linear capacitor to correct for both AM-AM and AM-PM [20]. Another challenge is to adjust the predistortion, or compensation, with respect to memory effects of

the PA. This requires that the analog predistortion also has memory, so that its predistortion can be adjusted based on previous input signals.

Modern analog predistorters use polynomial based solutions operating in the base band or RF domain. This means that a polynomial representing the desired predistortion non-linear characteristic, as in equation 53, is implemented in the analog domain. The analog implementation of the polynomial is normally based on Gilbert cells[3] to realize the multiplication and scaling of the signal.

Since the predistortion occurs in the analog domain, the base band filter (or the anti aliasing filter) after the digitial to analog converter (DAC), can filter out frequencies above the channel bandwidth, since the necessary bandwidth expansion occurs afterwards in the analog predistorter circuit. This also means that the DAC only has to operate at the fundamental bandwidth of the modulated signal. Furthermore, the analog predistion can be implemented in open or closed loop mode. The closed loop mode uses an observation receiver that updates the predistortion coefficients based on the output signal, whereas the open loop mode purely operates in a feed forward manner. Programmable time delays, used to compensate for memory effects are implemented as LC delay elements and delayed versions of the input signal are fed to parallel unique polynomials [51]. The combined outputs from the various delay taps then produce the predistorted input signal to the PA. It is fairly easy to create short time delays, but long time delays will consume very large die area, making them effectively impractical to implement. There are many published works that report successful memory effects cancellation using analog predistortion with memory taps, such as [52], [53], and [54]. A high level block schematic of a transmitter using a closed loop RF domain analog predistortion is shown in Fig. 45. Figure 46 shows an implementation of compensation of memory effects using m different delay elements, each path generating a nonlinear transformation of the input signal. An effective implementation would require programmability of time delays and polynomial coefficients.

[3] A Gilbert cell is basically a multiplier that can also be used as a variable gain amplifier or a mixer

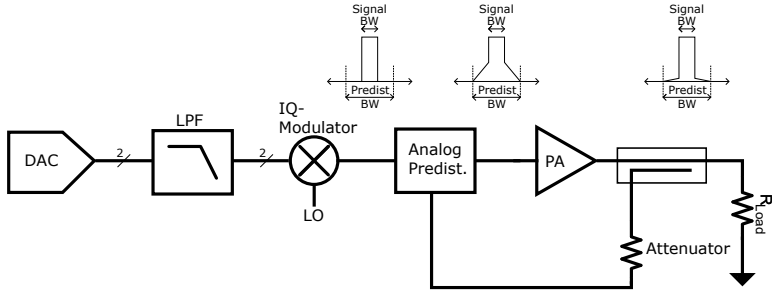

Figure 45: A high level block schematic of a closed loop RF domain analog predistorted transmitter.

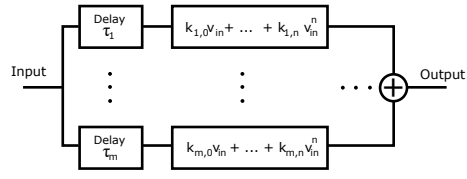

Figure 46: Illustration of implementation of memory taps in an analog predistorter.

5.5.2 Digital Predistortion

Digital predistortion (DPD) predistorts the input signal to the PA already in the digital domain. A DPD can operate in open loop mode, where the DPD has predetermined fixed nonlinearity, or in closed loop mode, where the DPD measures the quality of the PA output signal and updates the predistortion coefficients accordingly. A DPD operating in closed loop mode is shown in Fig. 47 and a DPD operating in open loop mode is shown in 48. The most common way to implement a DPD is, similarly to the analog predistortion, to use a polynomial expression of the nonlinearity. The order of the polynomial will determine the order of nonlinearity that can be suppressed. Generally speaking, higher order polynomial DPDs are more complex, but can also suppress nonlinearities more effectively. In addition, the DPD also has memory taps so that it can reduce the impact of memory effects in a PA, similar to how it is done in an analog predistorter. The memory taps are used to store information about prior input signal values to be able to adjust the input signal depending on the current state of the PA. The sampling frequency of the DPD will determine the highest possible bandwidth that can be accurately reproduced

according to the Nyquist criterion[4]. A so-called in-band DPD operates only on the same frequency bandwidth, or channel, as that of the transmitted signal, and can then improve in-channel distortion (EVM), but has little effect on frequencies outside the channel (ACLR). To significantly reduce ACLR the DPD must operate with at least three times the bandwidth of the channel. Another challenge with DPDs that operate with high polynomial order and bandwidth is that the DAC has to produce the high bandwidth input signal and that the succeeding baseband filter has to be made with higher bandwidth, not to suppress the added predistortion. In addition, the clocking of the DPD has to be increased with the same rate as the sampling frequency. DPDs, which lately have become the dominant method to cope with nonlinear PAs, are proven very effective in reducing distortion. Nevertheless, they require difficult tradeoffs, such as sampling rate, polynomial order, number of memory taps, DAC bandwidth, and analog filter bandwidth, in order not to consume more power than improving the linearity by simply spending more power in the PA.

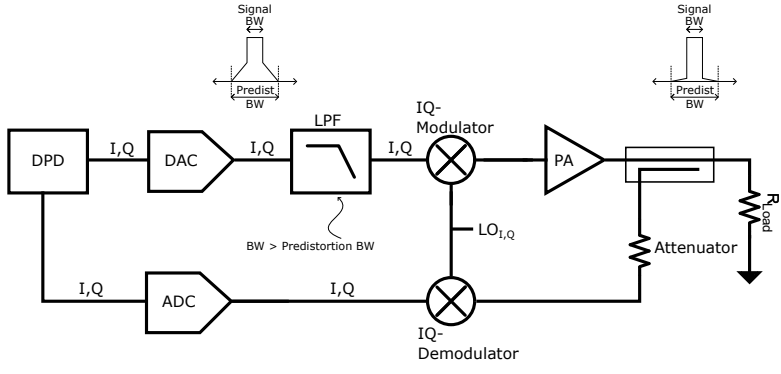

Figure 47: A DPD operating in closed loop mode.

[4]The Nyquist criterion states that to accurately represent a frequency, and to avoid folding, the sampling rate must be at least twice that of the maximum frequency in the signal that is sampled [55]

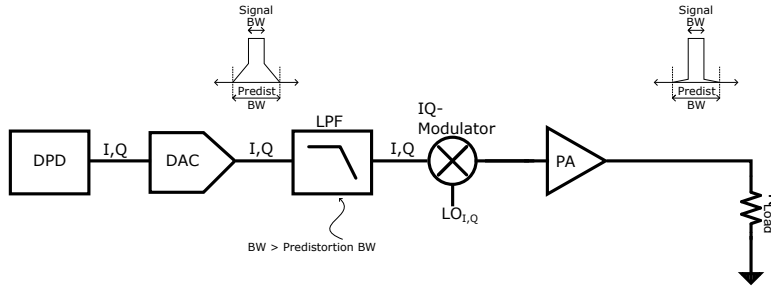

Figure 48: A DPD operating in open loop mode.

5.6 Simulation Example of Memory Effects in an idealized CMOS mmW Power Amplifier

With the target of investigating memory effects of a PA with an example, a simulation test bench for a PA operating at 27 GHz was set up in a 22 nm FD-SOI CMOS process design kit, depicted in figure 49 a-b. Following the definition of memory effects, i.e. bandwidth dependent nonlinear effects, and since memory effects arise due to non-constant node impedances at the baseband, fundamental, and second harmonic, the test bench uses close to ideal biasing of the CS and CG transistors, but with a baseband frequency dependent node impedance for the supply voltage. Furthermore, the output resonance tank has a rather low Q-value, due to the 50 ohm load resistor, which makes the node impedance stay virtually constant over the signal bandwidth at the fundamental frequency. At the supply node a parasitic resonance tank is formed by the parasitic inductance and the decoupling capacitor, which is depicted in Fig. 49 b. This parasitic resonance tank is close to impossible to avoid and becomes increasingly troublesome for high signal bandwidths. Nonetheless, the effects of a supply resonance can be highly mitigated by minimizing the inductance, increasing the capacitance, and finally adding a damping resistance to reduce the maximum impedance. However, the details in designing a supply network suitable for a high bandwidth PA are left out from this example.

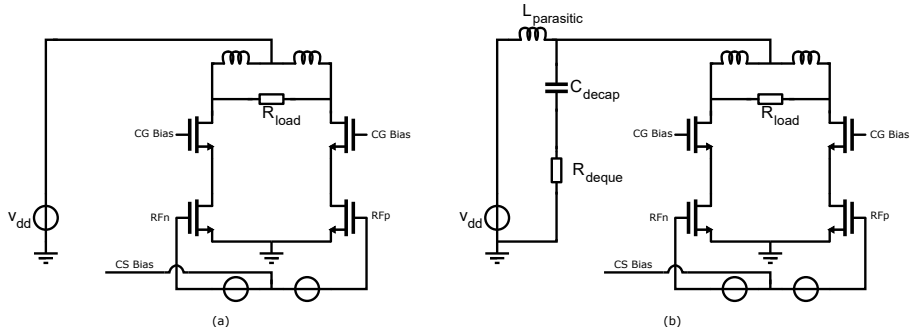

Figure 49: Simulation test bench of a differential CMOS PA. In (a) the bias networks are ideal and in (b) a parasitic resonance is present at the PA supply node.

Parts of the analysis in [27] is repeated to demonstrate a method on how to simulate and detect memory effects in a PA. The aim is to detect the presence of memory effects by their impact on third order intermodulation products (IM3). The first step is to perform a standard two-tone test on the power amplifier test bench in Fig. 49 a. However, even for a memoryless polynomial nonlinearity as described by Eq. 53, some higher order distortion also arrive at the same frequency as the IM3 products. Of these, the strongest contributions are typically from the lowest odd-order nonlinearity, i.e. the fifth order distortion. How much fifth order contributions that end up at the same frequencies as the third order intermodulations products, can be estimated by identifying the amplitude of the fifth order intermodulation products (IM5). From a fifth order polynomial stimulated by two input tones, it is clear that the coefficients of the fifth order distortion at IM3 frequencies are equal to 25/8, and 5/8 at IM5 frequencies [27]. This means that the fifth order contribution is five times as high at the frequency of IM3 as its IM5 contribution, and hence can be removed by subtracting five times the IM5 producs from the IM3. After this removal the remaining distortion products that fall on same frequency as the IM3 products are nearly proportional to the cube of the input amplitude. This gives a proposed normalized IM3 product as in equation 59, which for a memoryless circuit that is also free of distortion products of higher order than five, should be completely flat with respect to input amplitude.

$$IM3_{Norm} = \frac{IM3 - 5 \cdot IM5}{A^3{}_{in}} \tag{59}$$

Figure 50 shows the IM3 and normalized IM3 products from a two-tone test of the circuit in Fig. 49 a. Even after removal of fifth order distortion the

normalized IM3 products deviate from a flat, amplitude independent response, due to higher order distortion. Anyhow, removal of fifth order distortion still increased the input amplitude range by more than half a decade, for which the normalized IM3 is flat.

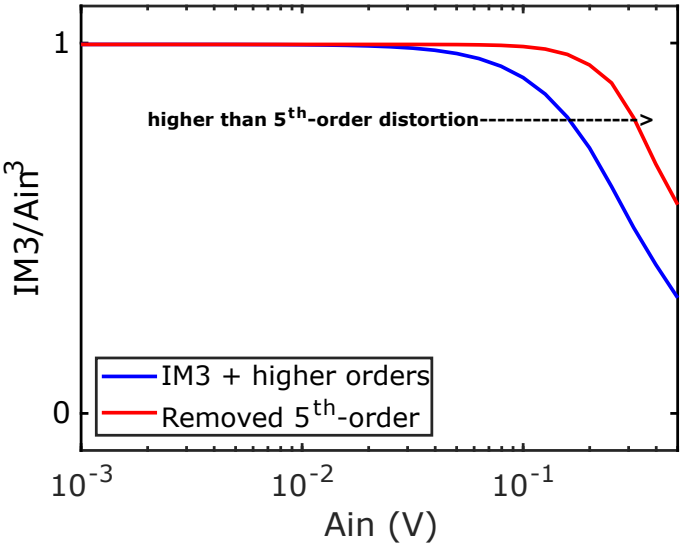

Figure 50: IM3 and normalized IM3 from a two tone test with low tone separation frequency and scaled to 1 for low levels.

The test-bench is then modified to Fig. 49 b, which suffers from a 40 MHz baseband resonance at the PA supply node that should produce memory effects, since it gives rise to a strong frequency dependent node impedance at base-band frequency. The supply node impedance is shown in Fig. 51. The base-band resonance is at 40 MHz and with a Q-value of 4. A two-tone test is then carried out, for different input amplitudes as well as for different tone separations, and the normalized IM3 products are plotted in an amplitude and tone separation plane. Without the presence of memory effects the expectation would be a flat plane, but with a high amplitude bending due to the higher order effects seen in Fig. 50. The result, however, shown in Fig. 52, clearly shows the presence of the strong memory effect, with a peak at 40 MHz tone separation caused by the supply parasitic resonance.

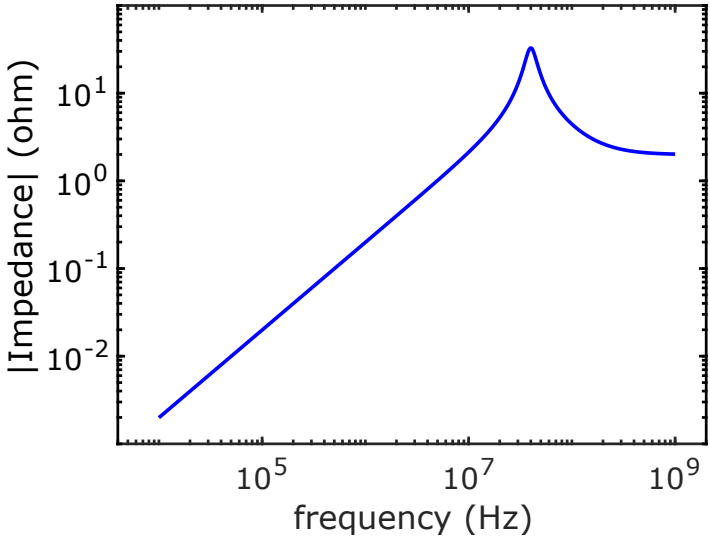

Figure 51: Impedance of the PA supply node due to the parasitic resonance at 40 MHz with a Q-value of 4.

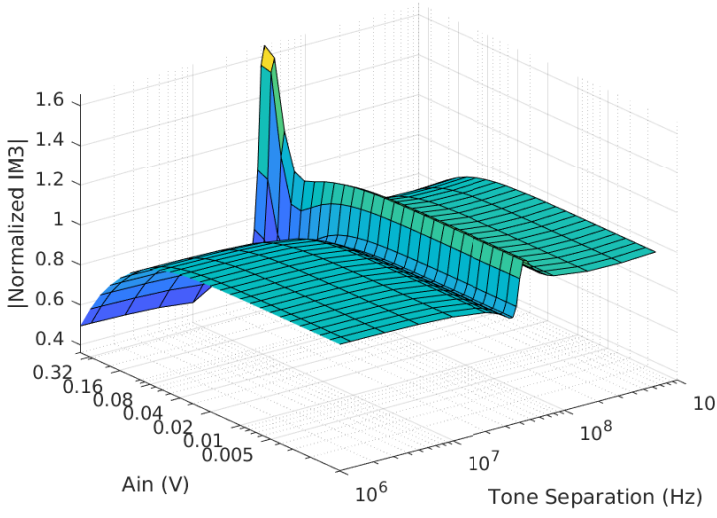

Figure 52: Normalized IM3 plotted versus input signal amplitude and tone separation. The presence of a memory effect can be seen for 40 MHz tone separation.

To identify how severe the problem is, the PA is simulated using a 100 MHz bandwidth OFDM-signal at 27 GHz, and its AM-AM and AM-PM are plotted in Fig. 53, and the ACPR for the two cases is shown in Fig. 54. Both plots reveal that the memory effects cause distortion only for high output power levels but with virtually no deterioration for low levels. It is beyond the scope of this work to identify exactly why that is, but a simple explanation would be that the PA in the test bench is biased rather high, and for very low output powers it operates in class A, making the supply current constant regardless of the input signal. This prevents the PA from injecting signal current at the supply node with the parasitic supply resonance. Furthermore, the PA is rather insensitive to supply voltage as long as it PA does not compress in output voltage, which it only does for high output power levels. A more general explanation would be that the memory effect is a combination of nonlinear effects with the bandwidth dependent node impedance. In this case the nonlinear effect is the rectifying of input signal voltage to baseband supply current by even order nonlinearities, that together with other nonlinearities, for baseband supply voltage to output RF amplitude and phase transfer, become a rather high order nonlinearity, which thus grows rapidly with the amplitude.

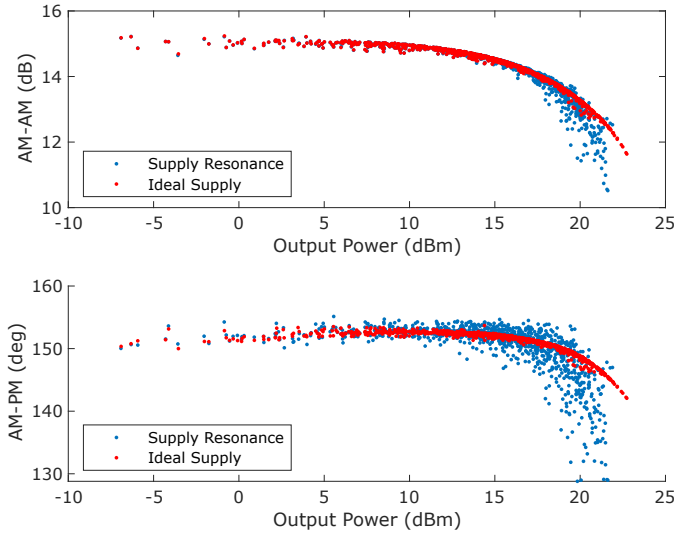

Figure 53: Simulation of PA with/without memory effects with 100 MHz BW OFDM signal at 27 GHz. Memory effects are significant at high signal levels.

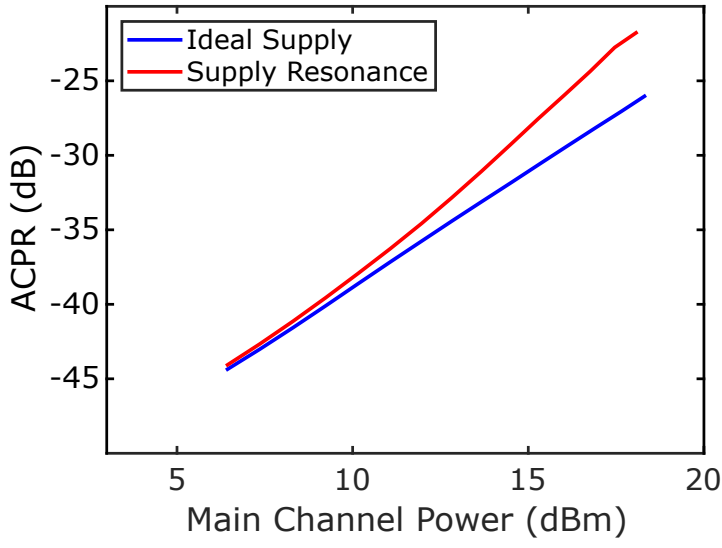

Figure 54: Simulated power sweep with/without memory effects with 100 MHz BW OFDM signal at 27 GHz. Memory effects degrade ACPR at high power levels.

Chapter 6

Efficient Power Amplifiers

The efficiency versus linearity trade-off is central in the PA classes and linearity discussions in chapters 4 and 5. An important aspect is how the efficiency depends on the output signal level. This chapter deals with the most popular PA architectures that aim at increasing the efficiency for a wider range of output signal levels. To show the range of signal levels needed to transmit modern wireless signals, the chapter is started by a section on peak-to-average ratio (PAR).

6.1 Peak to Average Ratio

In modern wireless systems, such as 4G and 5G cellular, OFDM signals are used with fundamentally high PAR value. An OFDM signal is constructed from multiple subcarriers, separated in frequency domain by the subcarrier spacing. Each subcarrier has a PAR value that is determined by its modulation, for instance 4-QAM, 16-QAM, or 64-QAM with increasing PAR value for higher orders of modulation, due to larger relative amplitude variation of the different constellation points. Table 6 summarizes the PAR value for different orders of quadrature amplitude modulations, and as can be seen the PAR value increases with the modulation order, but for very high modulation orders the PAR value becomes roughly constant[5].

However, when the subcarriers are added together in a real time domain OFDM signal, the sinusoidal like voltage signals of the subcarriers add together almost randomly, at least if there are many enough of them, which there definitely are in cellular systems. At some time instants they add constructively and create high peaks, and at other times they add destructively and create low valleys in the signal, resulting in a high PAR value regardless of the underlying subcarrier PAR. When combining many uncorrelated subcarriers, the instantaneous amplitude will follow a Rayleigh distribution and the likelihood P_0 that the signal amplitude, at any given time, is above the a certain PAR threshold can be expressed by Eq. 60 [56].

[5]The PAR for the symbols will asymptotically approach about 4.8 dB for $\infty - \text{QAM}$

Table 6: PAR comparison symbol points of QAM modulations

Modulation	Symbol Points PAR [dB]
4-QAM	0.0
16-QAM	2.6
64-QAM	3.7
256-QAM	4.2
1024-QAM	4.5
4096-QAM	4.6

$$P_0 = e^{-PAR^2} \tag{60}$$

which is plotted in Fig. 55 to illustrate the probability for different amplitudes.

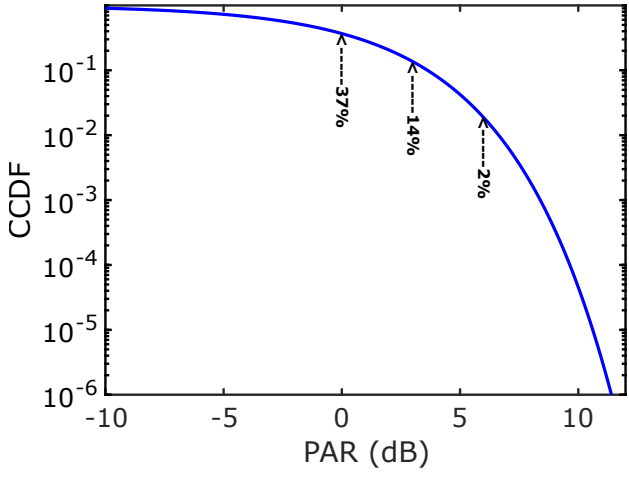

Figure 55: The complementary cumulative distribution function shows the probability that the output power is above a certain level relative to the average value. The probability that the power is above the average level, the average level +3 dB and the average level +6 dB is 37%, 14%, and 2%, respectively.

This leads to the conclusion that for the PA to be efficient when amplifying an OFDM signal, or another high PAR signal, it must be so for signals with

a wide range of power levels. A precise measure of the efficiency of a PA is the average efficiency for a signal with a given signal level distribution. The PAR value of an OFDM signal gives a rather good estimate of its signal level distribution. To further simplify the measure of the PA efficiency, the efficiency at average output power often gives a good approximation of the efficiency that the PA will have when amplifying the real modulated signal. But as Fig. 55 indicates there is no limit on how high the signal peaks can be[6], which means that a maximum signal level that the PA ideally should amplify linearly, has to be decided as the "peak" value of the PA. The PA peak level can be chosen differently, and the lower the choice is, the greater the impact on distortion it will have due to increased probability for clipping of signal peaks. On the other hand, choosing a lower PA peak value will increase the average efficiency, since the PA will operate closer to its peak value when operating at average output power. The efficiency at average output power is often referred to as the back-off efficiency. There are certain characteristics that will distort this measure, for instance that the slope of how the efficiency depends on signal level is not constant, and the use of crest factor reduction, a signal level limiting method. Nevertheless, if the efficiency when amplifying the real signal is not known, the back-off efficiency is an often accurate and well-used measure of the efficiency of a PA used for amplifying high PAR signals.

6.2 Doherty Power Amplifier

The Doherty amplifier was invented in 1936 by Willliam H. Doherty [57] and close to hundred years later it is still used in modern electronics and is by far the most popular efficient PA. The Doherty amplifier depicted in Fig. 56 consists of three essential parts, the input network, the two parallel amplifying stages (main and auxiliary[7]), and the output network. Starting in the middle, the purpose of using two amplifiers, main and auxiliary, instead of a single, is to make it possible to turn off the auxiliary amplifier for low to medium signal levels, whereas the main amplifier always operates. This is achieved by applying different bias levels to the two amplifiers. Typically the main amplifier is biased in class AB and the auxiliary amplifier is biased in class C[8]. Recalling that a class C amplifier is biased below the threshold voltage, it will essentially be turned off for low to medium signal levels, whereas the

[6] In reality the absolute maximum peak will be N times the amplitude of the subcarriers, where N is the number of subcarriers. However, for thousands of subcarriers this is so unlikely to happen that it can be completely disregarded.

[7] Quite often the main and auxiliary amplifiers are referred to as carrier and peaking amplifiers

[8] For a PA to be biased in class C it should have a conduction angle between 0 and 180°. In the Doherty amplifier, for low and medium signal levels, when the class C amplifier is turned off, the conduction angle for the class C amplifier is zero. Then as the envelope of the input signal goes up the conduction angle increases, but never reaches 180°.

class AB amplifier is turned on for all input signals. The output network has two functions, firstly it combines the output power from the two amplifiers, and secondly it provides an impedance inversion so that the auxiliary amplifier can perform desired load modulation of the main amplifier when it turns on. The idea of the load modulation is that the main amplifier should drive a high impedance for low to medium signal levels, and as the main amplifier starts to compress in voltage, the auxiliary amplifier turns on and the impedance inverting load modulation actively reduces the impedance at the output of the main amplifier. The input network should distribute the input power to both main and auxiliary amplifiers, with phases so that the signals from the main and auxiliary amplifiers combine constructively at the output.

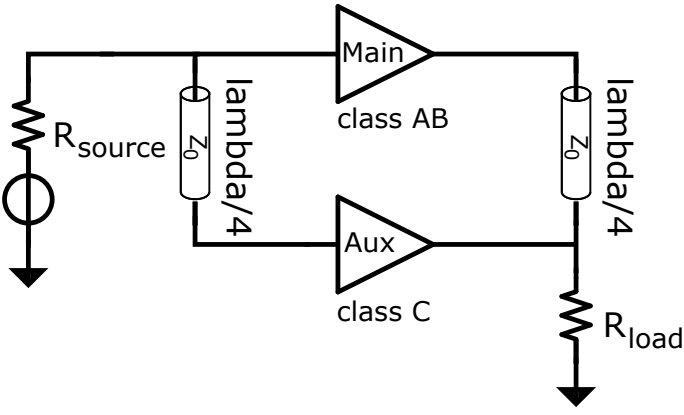

Figure 56: A current combined Doherty PA

6.2.1 Load Modulation

Load modulation is to change, or modulate, the load impedance of an amplifier in response to the modulation of the input signal. To explain the concept we start by going back to Ohms law and the definition of resistance[9] R=V/I. Figure 57 shows how load modulation can increase or decrease the load resistance depending on the sign of a controlled current source, connected to the same load. In the example, current I_2 is controlled by I_1 and scaled with a positive constant c. In Fig. 57 a, the voltage across the load becomes $V = I_1(1+c)R_{load}$, which after division with I_1 gives the load resistance $R_1 = (1 + c)R_{load}$ seen from the source I_1. In Fig. 57 b, the voltage across the load becomes $V = I_1(1-c)R_{load}$, which after division with I_1 gives the load resistance $R_1 = (1-c)R_{load}$.

[9]For simplicity we limit the below reasoning to the real part of the impedance

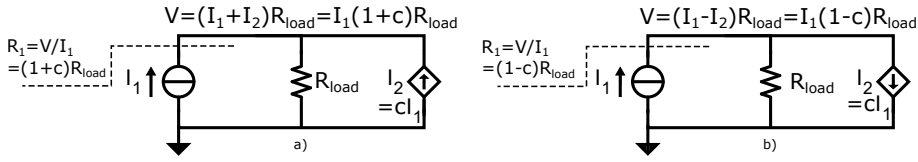

Figure 57: a) Load modulation increasing load resistance and b) load modulation reducing load resistance.

6.2.2 Load Modulation in a Doherty Power Ampifier

For the Doherty amplifier the current source I_1 can represent the main amplifier and the controlled current source I_2 the auxiliary amplifier. As mentioned before, the aim is to reduce the impedance at the main amplifier when the auxiliary amplifier turns on. What should be done is thus more similar to figure 57 b than a. However, simply using the auxiliary amplifier to pull out current from the load with the purpose of reducing the voltage at the main amplifier would ruin the output power. Introducing an impedance inverting network, between the main amplifier and the load, however, will make it possible to add current from the auxiliary amplifier into the load while also reducing the load impedance as seen from the main amplifier. A $\lambda/4$-transmission line functions as an impedance inverter and the impedance at the main amplifier can be expressed as

$$Z_{main} = \frac{Z_{\lambda/4}^2}{Z_{load}} \tag{61}$$

where Z_{main} is the impedance at the main amplifier, $Z_{\lambda/4}$ is the characteristic impedance of the $\lambda/4$-transmission line, and Z_{load} is the load impedance. For instance, if the impedance at the load is $25\,\Omega$, then a $50\,\Omega$-$\lambda/4$-transmission line will invert the $25\,\Omega$ up to $100\,\Omega$, and contrary, a $100\,\Omega$ load will be inverted down to $25\,\Omega$ at the main amplifier by the same $50\,\Omega$-$\lambda/4$-transmission line. This makes it possible to control the impedance level at the main amplifier by controlling the impedance level at the load side of the $\lambda/4$-transmission line. But the transmission line also adds a delay corresponding to a 90° phase shift, due to the time it takes for the wave to travel through it. So for the impedance inversion to work as intended, for the signal currents to combine in phase at the output, a 90° phase shift must be added to the signal from the auxiliary amplifier. This phase delay in the signal path of the auxiliary amplifier is typically provided at the amplifier input. At the input of the main and auxiliary amplifiers the impedance can typically be assumed to be matched, and the phase shift can then be achieved by a $\lambda/4$-transmission line. Redrawing Fig.

57 a so it resembles the complete Doherty amplifier in Fig. 56 is shown in Fig. 58.

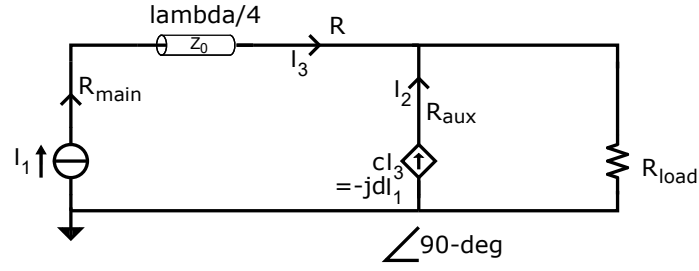

Figure 58: The output interconnection of a Doherty amplifier. Main and auxiliary amplifiers have been replaced by ideal current sources. Input signal to the auxiliary amplifier has been delayed by 90 degrees and a $\lambda/4$-transmission line, with impedance Z_0, connects the two current sources.

As before the main current is labelled I_1 and the auxiliary current I_2. The relations between the three currents in Fig. 58 are defined in equation 62.

$$I_2 = cI_3 = -jdI_1 \tag{62}$$

Using the schematic in Fig. 58 and the relationships in Eq. 62, it is possible do derive expressions for the impedance levels at the main amplifier (R_{main}), the auxiliary amplifier (R_{aux}), and the load modulated impedance seen by current I_3 (R). Derivations can be found in [4] and the main results are repeated here for convenience.

$$R = R_{load}(1 + \frac{d}{\frac{Z_0}{R_{load}} - d}) \tag{63}$$

$$R_{aux} = R_{load}(1 + \frac{d}{\frac{Z_0}{R_{load}} - d})/\frac{d}{\frac{Z_0}{R_{load}} - d} \tag{64}$$

$$R_{main} = \frac{{Z_0}^2}{R_{load}(1 + \frac{d}{\frac{Z_0}{R_{load}} - d})} \tag{65}$$

We now assume a Doherty amplifier where the auxiliary amplifier is biased to turn on at 6 dB back-off from maximum input signal level. Furthermore, for ideal Doherty operation, current I_1 must increase linearly with the input signal, and current I_2 must be zero below the 6 dB back-off level after which it increases linearly, as shown in Fig. 59. Since d is the ratio of the two currents

the impedances R_{main}, R_{aux}, and R from Eq. 63, 64, and 65 can be plotted and the results are shown in Fig. 60 [4].

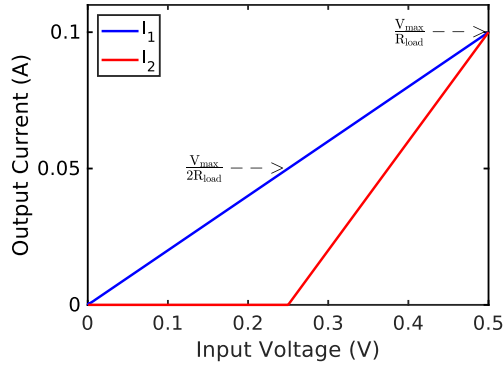

Figure 59: Output current of main amplifier I_1 should ideally be linear for all input signals, and output current from auxiliary amplifier I_2 should increase linearly for input signals above 6 dB back-off.

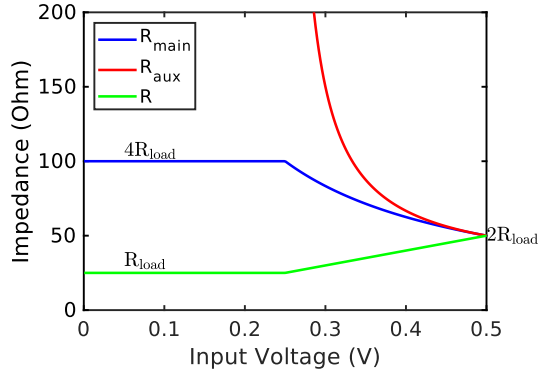

Figure 60: Impedance R, which load modulates R_{main} increases linearly for input signals above back-off, and in turn R_{main} reduces as 1/R. For this ideal case R_{aux} is infinite below back-off.

Since we know the main and auxiliary amplifier output currents for ideal Doherty operation and their load impedances as a function of the input signal level, we can calculate the output power of the main, auxiliary, and complete Doherty amplifier shown in Fig. 61.

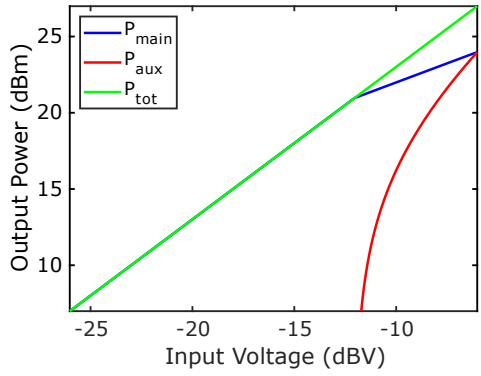

Figure 61: Output power when the output currents of main and auxiliary amplifiers follow the ideal linear output in Fig. 59. Output power for main amplifier is linear up to back-off, after which it increases proportional to $\sqrt{v_{in}}$. Output power from auxiliary amplifier exactly fills up the gap to make the combined output power increase linearly.

The output voltage levels of the main and auxiliary amplifiers can also be calculated from output current and load impedance, shown in Fig. 62.

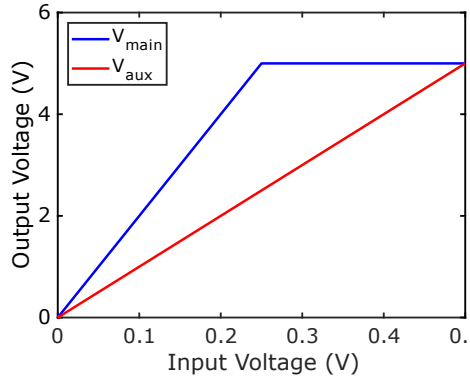

Figure 62: The output voltage of the main amplifier increases linearly up to the maximum output voltage, after which it stays constant due to load modulation. The output voltage of the auxiliary amplifier increases linearly as it is a linear representation of the output power.

6.2.3 Conduction angle of Auxiliary Amplifier

To provide definitions needed in the following section, Fig. 63 shows the drain current (I_{DS}), gate-to-source voltage (V_{GS}), maximum drain current ($\mathrm{I_{max}}$), threshold voltage ($\mathrm{V_{th}}$), overdrive voltage ($\mathrm{v_{od}}$), max input voltage ($\mathrm{v_{in\,max}}$), and conduction angle (2Φ) for the auxiliary amplifier.

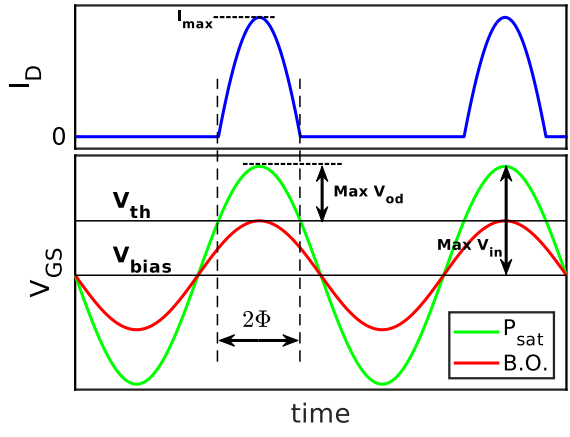

Figure 63: Threshold voltage, maximum drain current, conduction angle, maximum overdrive voltage, and maximum input voltage for the auxiliary amplifier and its transistor. The bias voltage in the figure is chosen such that the maximum conduction angle is $2\Phi = 2\pi/3$

For an amplifier modelled with an ideal voltage controlled current source, similar to the small signal model for a transistor, its $\mathrm{I_{max}}$ will be the product of the maximum overdrive voltage of the input signal and the transconductance $\mathrm{g_m}$. This will be true for all transconductance amplifiers. The bias level for the auxiliary amplifier lies beneath the threshold voltage, controlling the onset of current conduction, thereby providing us with an equation for $\mathrm{I_{max}}$.

$$I_{max} = g_m(v_{in_{max}} - v_{in_{6dB_{BO}}}) = g_m v_{od_{max}} \tag{66}$$

where $\mathrm{v_{od}}$ is the effective overdrive voltage. At max input signal, the conduction angle in this example is chosen to $2\pi/3$, and the auxiliary amplifier should then drive the load with full fundamental current, which will set a value for the required transconductance $\mathrm{g_m}$ and the maximum drain current $\mathrm{I_{max}}$.

6.2.4 Efficiency of Ideal Doherty Amplifier

The efficiency of a Doherty amplifier is the combined efficiency of the main and auxiliary amplifiers. The auxiliary amplifier is turned off below 6dB back-off[10] and the efficiency then becomes that of the main amplifier, which here is assumed to operate in class B, loaded for saturation at 6 dB back-off. Above back-off, things become a lot more complicated, and the main amplifier then operates at peak class B efficiency of ideally $\pi/4 = 78.5\%$, since it is load modulated by the auxiliary to have maximum and constant output voltage. The auxiliary amplifier operates in class C and its efficiency and output power is determined by its conduction angle and output voltage. Using Eq. 20 gives us the efficiency at max output power for a certain conduction angle. For conduction angle $2\pi/3$, which is set to occur at the auxiliary amplifier maximum output voltage swing, the efficiency becomes 89.7%. The efficiency of the auxiliary amplifier reduces linearly from the peak level, scaled as the ratio of the output signal voltage level and the supply voltage, when backing down the output power from the auxiliary PA:

$$\eta = \frac{2\Phi - sin\,2\Phi}{4(sin\,\Phi - \Phi cos\,\Phi)} \frac{v_{aux}}{v_{dd}} \tag{67}$$

For ideal Doherty operation the fundamental current output of th auxiliary amplifier should increase linearly above bock-off. However, as the conduction angle is signal level dependent, this cannot happen even for a transistor with ideal g_m. From [4] it is clear that the conduction angle depends on the input voltage v_{in}, and it is zero for $v_{in} < v_{in_{6dB_{BO}}}$ and for $v_{in} \geq v_{in_{6dB_{BO}}}$ it is $2\Phi = 2 \cdot \arccos(v_{in_{6dB_{BO}}}/v_{in})$. The conclusion is that an ideal Doherty PA, where the auxiliary amplifier starts to conduct exactly when the main amplifier goes into voltage compression cannot be linear.

But we limit the analysis to the case where the auxiliary amplifier is linear above back-off, namely a somewhat theoretical case were we can set conduction angle freely, so that it is zero below back-off and equal to $2\pi/3$ from back-off to max output power. Since its output voltage and conduction angle are known, it is now possible to use equation 67 to estimate the efficiency of the auxiliary amplifier, which finally gives us the efficiency of the ideal Doherty amplifier, see Fig.64. The constant conduction angle can be achieved with adaptive bias of the auxiliary amplifier, although the abrupt transition from 0 to $2\pi/3$ would cause bias signal bandwidth issues for modulated signals.

[10] Here the dB back-off relates both to output power and input power.

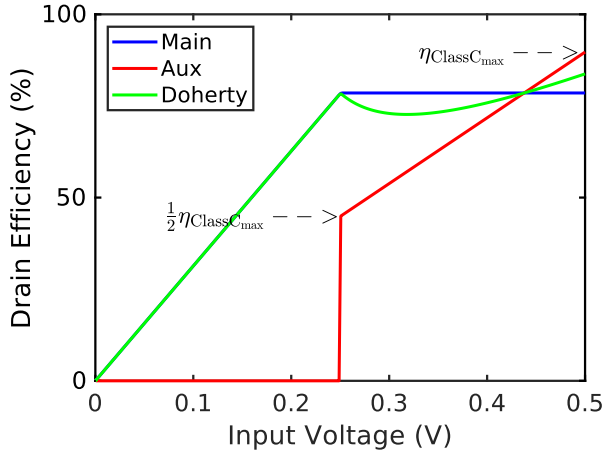

Figure 64: Efficiency of an ideal Doherty amplifier consisting of a class B main amplifier combined with an auxiliary amplifier biased in class C with a constant conduction angle of $2\pi/3$ above back-off.

6.3 Outphasing Power Amplifier

In 1935, a novel outphasing modulation technique was presented by Chireix [58], but later the work has been more known as outphasing power amplifier. The Chireix outphasing PA, depicted in Fig. 65, consists of two amplifiers with constant input and output voltage amplitudes. The amplifiers can be highly effective since they operate with constant amplitude, with no need to bother about back-off efficiency or even linearity, as long as they produce a constant output voltage amplitude. A prerequisite for the outphasing PA, as presented by Chireix, is that the transistors operate as voltage sources. For most cases this is not true and transistors are typically modelled as controlled current sources. However, since they are operating with a constant and large output amplitude, the transistor can be deep in saturation, and the approximation to model them as voltage sources then becomes acceptable [59].

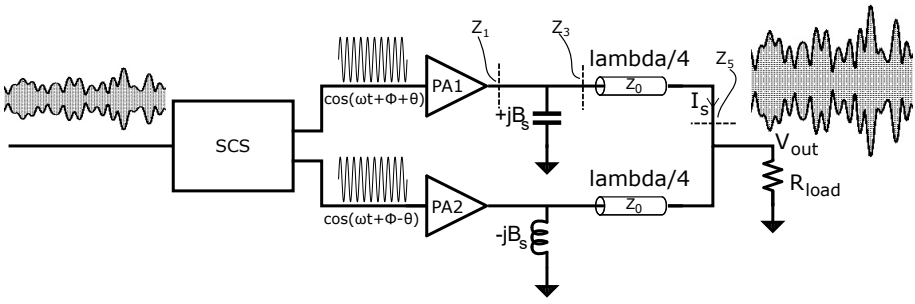

Figure 65: A Chireix outphasing PA

Instead of a single RF input signal, an outphasing PA requires two, and it therefore has a preceding circuit that produces the two input signals. This circuit is called a signal component separator (SCS), and its task is to transform the amplitude and phase modulated input signal into two constant amplitude signals representing the information. The two input signals have a common mode phase and a differential phase. The common mode phase follows the phase of the RF input signal and the differential phase represents the amplitude information. The differential phase between the two PAs is called the outphasing angle. The SCS is effectively performing a transformation from Cartesian to polar coordinates, i.e. from IQ-signals to amplitude and phase, with the addition that the amplitude is expressed as a phase difference, the outphasing angle, $2 \cdot \theta$. A major drawback of the outphasing PA is that the coordinate transformation is a nonlinear process which creates bandwidth expansion. The transformation can be expressed as in Eq. 68-71. The amplitude information that should be represented by the outphasing angle is:

$$A(t) = \sqrt{I(t)^2 + Q(t)^2} \tag{68}$$

with the outphasing angle θ given by

$$\theta = arccos(\frac{A(t)}{A_{max}}) \tag{69}$$

The common mode phase that represents the phase information of the modulated input signal is

$$\phi(t) = \arctan \frac{Q(t)}{I(t)}, I > 0 \tag{70}$$

$$\phi(t) = \arctan \frac{Q(t)}{I(t)} + \pi, I < 0 \tag{71}$$

As can be seem both the amplitude and phase coordinate transformations are nonlinear functions and give rise to bandwidth expansion, but it is the common mode phase for close to origin passings in the constellation diagram that becomes most troublesome. There are many techniques related to mitigating this problem, such as introducing forbidden zones close to the origin, but the bandwidth expansion problem is still highly affecting whether a polar PA is an attractive solution or not. The SCS can be performed digitally, prior to the DAC in the transmitter chain, where two transmitter chains are then used to produce the input signals to the two amplifiers in the outphasing PA. Another option is to perform the SCS in the RF domain prior to the outphasing PA. Both techniques have their advantages and disadvantages. Previously it was considered that the best choice was to perform the signal separation in the digital domain due to the nonlinear signal processing that is required. However, it would require that the two separate transmitter chains have very wideband DACs and signal paths, to handle the bandwidth expanded signals after the coordinate transformation [60]. RF domain SCS operates directly on the modulated RF signal and leaves the transmitter chain essentially unaffected, effectively making the outphasing PA plug and play. Other advantages of the RF domain SCS are reported, in more recent research, such as low complexity, low cost, and low power consumption [61]. Even though a lot of research work has been performed for outphasing PAs it is not as frequently used in real implementations as the Doherty PA [59]. However, recent advances in digital circuitry has made the implementation of signal separation in the digital domain less complex [59], and recent simpler and more linear RF-domain SCS also enable the use of outphasing PAs [62]. In addition, increased integration level makes it more likely to accept that the PA is not single-input single-output, since it will have to be designed together with the transceiver circuit anyway. Furthermore, increasing thermal problems in highly integrated solutions pushes requirements for power amplifier efficiency. All this points in a direction towards more outphasing PAs in real products. However, the fundamental bandwidth expansion problem will become harder and harder to overcome as the signal bandwidth increases rapidly in modern wireless systems.

6.3.1 Analysis of Simple Outphasing System

A thorough analysis of the efficiency at back-off is found in [63] and here some of the important results are repeated. We start with a simple outphasing system as depicted in Fig. 66, with the sub-PA output signal phasor relationships shown in Fig. 67.

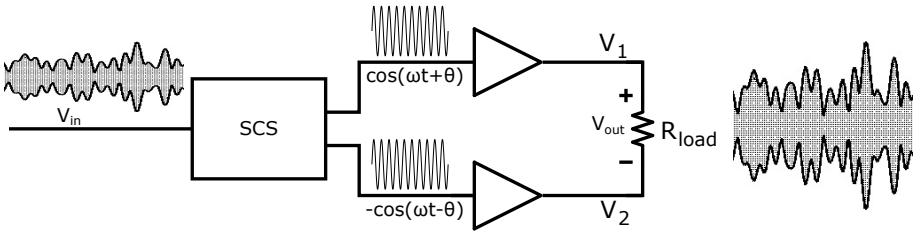

Figure 66: An outphasing system with a floating load.

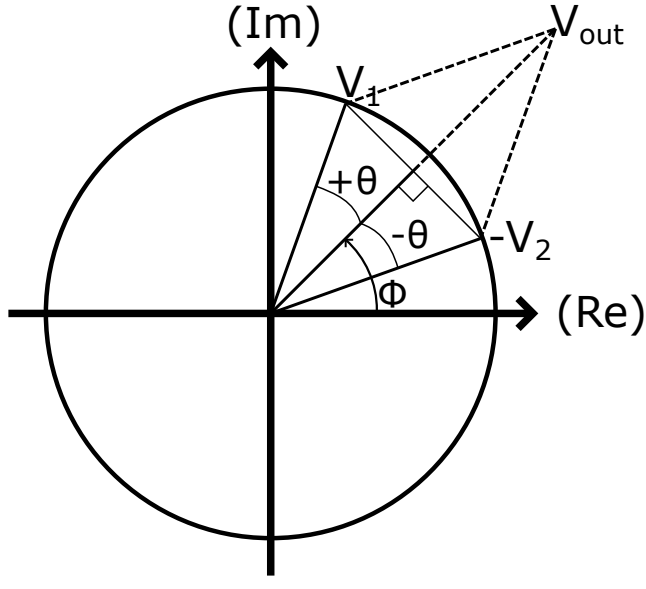

Figure 67: Output signal phasors. The outphasing angle θ controls output amplitude. The common mode phase ϕ represents the phase modulation.

The output voltages from the two PAs are expressed in complex phasor form, from which the output voltage, current and impedance as a function of the outphasing angle can be calculated.

$$V_1 = \frac{V_{outPeak}}{2}\left(cos(\phi+\theta) + jsin(\phi+\theta)\right) \tag{72}$$

$$V_2 = -\frac{V_{outPeak}}{2}\left(cos(\phi-\theta) + jsin(\phi-\theta)\right) \tag{73}$$

$$V_{out} = V_1 - V_2 = V_{outPeak} cos(\theta) \cdot (cos(\phi) + jsin(\phi)) \tag{74}$$

and with the output current

$$|I_{out}| = \frac{V_{outPeak} cos(\theta)}{R_{load}} \tag{75}$$

Set $\phi = 0$ to simplify the load impedance calculations below, will not affect the result. The load impedance of the amplifiers become

$$Z_{1Load} = \frac{\frac{V_{outPeak}}{2}}{\frac{V_{outPeak}}{R_{load}}} \frac{cos(\theta) + jsin(\theta)}{cos(\theta)} = \frac{R_{load}}{2} (1 + jtan(\theta)) \tag{76}$$

$$Z_{2Load} = \frac{-\frac{V_{outPeak}}{2}}{-\frac{V_{outPeak}}{R_{load}}} \frac{cos(-\theta) + jsin(-\theta)}{cos(\theta)} = \frac{R_{load}}{2} (1 - jtan(\theta)) \tag{77}$$

Clearly the complex part of the load impedance is affected by the outphasing angle θ. From now on we limit the analysis to the case where the two PAs are implemented using class B amplifiers that are driven at their peak amplitude. The DC current consumption of a class B PA is proportional to the output current from it. As we increase the outphasing angle θ from 0 radians were the peak output current occurs, the output current of the two class B PAs are reduced. Both sub-PAs will then reduce their DC current linearly with the output current, but the output power follows the square of the output current. This gives an efficiency that is proportional to the output current, or voltage, reaching the maximum class B efficiency at the maximum output level:

$$\eta = \frac{\pi}{4} \frac{V_{out}}{V_{max}} \tag{78}$$

Equation 78 should be interpreted as that even though the two sub-PAs operate at peak voltage swing the delivered real part of their complex output current is reduced for reduced output voltages, which makes the efficiency go down as much as for an individual class B PA when operating with reduced output voltage swing. Effectively what has been done is to achieve a load modulation that increases only the reactive part of the load impedance for back-off outphasing angles, which can be seen in Eq. 76 and 77.

6.3.2 Analysis of Chireix Outphasing System

The output voltage, and thereby the output power, of the outphasing PA is controlled by the outphasing angle, and again we repeat some of the most important steps in [63] to understand how the Chireix outphasing PA works.

Like in the simple outphasing system above, the load impedance of the two sub-PAs depend on the outphasing angle. Figure 65 shows the Chireix outphasing PA subject for the analysis below. The impedance Z_5 at the output of the upper $\lambda/4$ transmission line is

$$Z_5 = \frac{V_{out}}{I_5} = \frac{2R_{load}V_{out}}{V_{outPeak}}\left(cos(\theta) - jsin(\theta)\right) \tag{79}$$

which is then transformed by the $\lambda/4$ transmission line to $Z_3 = Z_0^2/Z_5$. The goal is now to explain why we should add the suspectances $-jB_s$ and $+jB_s$ at the output of the two sub-PAs. We therefore use the admittance $Y_3 = 1/Z_3$, which becomes

$$Y_3 = \frac{Z_5}{Z_0^2} = \frac{2R_{load}V_{out}}{Z_0^2 V_{outPeak}}\left(cos(\theta) - jsin(\theta)\right) \tag{80}$$

By symmetry the admittance becomes identical for the equivalent node in the lower branch, but with opposite sign of the complex part. Without adding $+jB_s$ and $-jB_s$ it is clear that the two power amplifiers would each have to drive a complex impedance. However, by choosing $+jB_s$ that is equal to, but with opposite sign, to the imaginary part of Y_3 we can create a real valued admittance, at least for some output amplitude (outphasing angle). Also by symmetry, the complex part of the admittance of the lower branch has the opposite sign, and therefore PA 2 will see a real admittance when instead adding the suseptance $-jB_s$. The admittance as seen by PA1 is [63]:

$$Y_1 = G_1 - jB_1 \tag{81}$$

with

$$G_1 = \frac{2R_{load}}{Z_0^2}\left(\frac{V_{out}}{V_{outPeak}}\right)^2 \tag{82}$$

and

$$B_1 = \frac{2R_{load}}{Z_0^2}\left(\frac{V_{out}}{V_{outPeak}}\left(1 - \left(\frac{V_{out}}{V_{outPeak}}\right)^2\right)^{0.5} - B_s'\right) \tag{83}$$

where R_{load} is the load resistance, Z_0 the characteristic impedance of the transmission lines, V_{out} the output voltage, $V_{outPeak}$ the maximum output voltage, and $B_s' = B_s Z_0^2/(2R_{load})$ is the normalized shunt susceptance. By normalizing, a simple expression falls out for when the susceptive part of Y_1 is zero.

$$B_s' = sin(\theta)cos(\theta) = \frac{V_{out}}{V_{outPeak}}\left(1 - \left(\frac{Vout}{V_{outPeak}}\right)^2\right)^{0.5} \tag{84}$$

Plotting B'_s versus normalized output voltage reveals that the suseptive part does not become zero for a single output voltage, instead it is achieved for two output voltages[11], except when choosing $B'_s = 0.5$

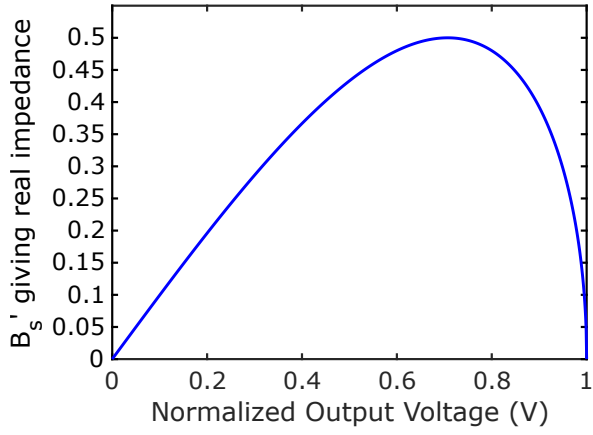

Figure 68: B'_s value giving real admittance at output of the sub-PAs

From Eq. 18 the DC power of a class B PA is found, and in this context it becomes

$$P_{dc_{subPA}} = \frac{2}{\pi} V_{dd} I_{out} = \frac{2}{\pi} V_{dd}^2 |Y_1| \tag{85}$$

and since the output power of course is

$$P_{out} = \frac{V_{out}^2}{2R_{load}} \tag{86}$$

the efficiency of the Chireix system becomes

$$\eta = \frac{P_{out}}{2P_{dc_{subPA}}} = \frac{\frac{V_{out}^2}{2R_{load}}}{\frac{4}{\pi} V_{dd}^2 |Y_1|} \tag{87}$$

The DC power consumption and the efficiency of the Chireix outphasing system is plotted in Fig. 69 and 70 below. For $B'_s = 0$ no susceptances are added and the system becomes equivalent to the simple outphasing system in Fig. 66, where efficiency is the same as for a class B PA. As B'_s is increased

[11] This result uses the derived result in Raabs excellent paper from 1985, but is in contrast with the conclusion in the paper

the amplitude levels are shifted for which the DC current, and thereby the DC power, is minimized, and the efficiency is peaking.

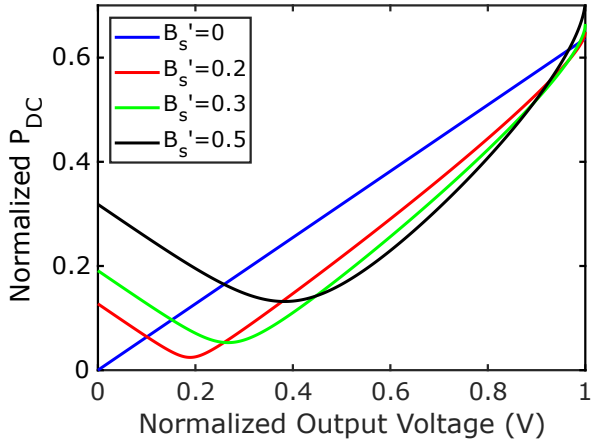

Figure 69: DC power for different B_s' values

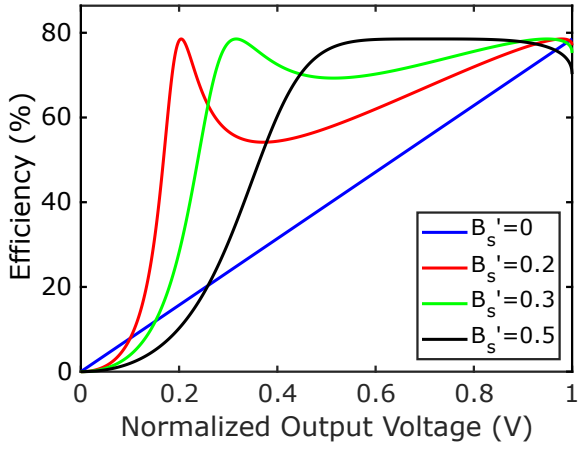

Figure 70: Efficiency of the Chireix outphasing system for different B_s' values

When the sub-PAs drive a real-valued load the efficiency of the Chireix outphasing system becomes equal to the peak efficiency of the class B PA.

Except for $B'_s = 0.5$, this always happens for two different output power levels. In the output power region between the two peaks the efficiency drops. The larger the complex part of the admittance becomes, the deeper the dip. At high output voltages the efficiency starts to go down, except for $B'_s = 0$, which is natural since at peak output voltage the ideal value of the added suceptances should be equal to zero. The nearly flat region for $B'_s = 0.5$ can be explained by that for a wide range of output voltages there is only a small deviation from the optimal B'_s in Fig. 68.

Chapter 7

Summary of Included Papers with Scientific Contribution

This chapter provides a summary of the papers that constitutes this dissertation. For each paper, a brief technical summary of the paper is given together its scientific contribution to the field and the details about the authors contribution to the paper.

7.1 Paper I: A 26 GHz 22.2 dBm Variable Gain Power Amplifier in 28 nm FD-SOI CMOS for 5G Antenna Arrays

7.1.1 Overview

Paper I presents an mmW power amplifier and pre-power amplifier, with aim for integration in an antenna array system. The circuit is designed in a CMOS process suitable for high level integration together with a large digital design located on the same chip. The process choice puts significant reliability challenges on the power amplifier design, and to increase the output power, while still reducing the maximum needed supply voltage, the circuit utilizes a "two way" output combiner prior to the load.

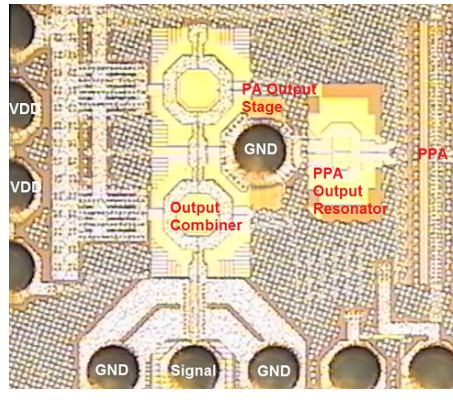

Figure 71: Die photo of the PPA, PA, and output combiner of paper I.

7.1.2 Scientific Contribution

- The design pushed the boundary for mmW power amplifier designs in 28 nm FD-SOI CMOS, demonstrating state-of-the-art performance for both output power and 1 dB compression point.

7.1.3 Comparison With State-of-the-art

A performance summary and comparison with state-of-the-art of the most relevant works at the time of the publication of paper I can be seen in table 7. The PA and PPA reached the highest output power, 1 dB output compression point, and Gain, while occupying the smallest die area. Furthermore, it was the only work that was also providing possibility for high accuracy gain control.

Table 7: Paper I Performance Summary and Comparison [1]

Parameter	This Work		RFIC'17 [64]	RFIC'17 [65]	ISSCC'16 [66]		IMS'16 [67]	
Tech. [nm]	28 SOI CMOS		28 Bulk CMOS	28 Bulk CMOS	28 Bulk CMOS		28 Bulk CMOS	
Freq. [GHz]	26		27	32	30		28	
Pwr.Comb.	2		2	2	1		1	
Gain Control	5 bits, 19dB		None	None	None		None	
No. of Stages	2		2	2	2		1	
Vdd [V]	1.5	1.8	1	1	1	1.15	1.1	2.2
P_{sat} [dBm]	20.6	22.2	18.1	19.8	14	15.3	14.8	19.8
P_{1dB} [dBm]	18.8	20.7	16.8	16	13.2	14.3	14.0	18.6
PAE_{max} [%]	22.6(1)	21.3(1)	41.5	21	35.5	36.6	36.5	43.3
PAE_{1dB} [%]	16.6(1)	14.6(1)	37.6	12.8	34.3	35.8	35.2	41.4
PAE_{max} [%]	20(2)	19.6(2)	41.5	21	35.5	36.6	36.5	43.3
PAE_{1dB} [%]	14.5(2)	13.5(2)	37.6	12.8	34.3	35.8	35.2	41.4
Gain [dB]	33.1(3)	33.9(3)	20.5	22	15.7	16.3	10.0	13.6
Area [mm^2]	0.144		0.361	0.59	0.16		0.28	

(1) Not including PPA power consumption.
(2) Including PPA power consumption.
(3) Simulated value since input signal is generated internally. Also used for calculating PAE values.

7.1.4 My Contribution

Together with the second author, I'm the creator of the complete chosen circuit architecture and circuit analysis, simulations, and layout. The third author did the design and layout of the pre-power amplifier. I did the lab measurement and manuscript writing. During the whole process the last author provided valuable feedback and support, but with emphasis on finalizing the manuscript.

7.2 Paper II: A 27 GHz Adaptive Bias Variable Gain Power Amplifier and T/R Switch in 22 nm FD-SOI CMOS for 5G Antenna Arrays

7.2.1 Overview

Paper II presents an mmW TRX-switch, power amplifier, and pre-power amplifier, targeted for integration in an antenna array system. As for paper I it was designed in a CMOS process suitable for high level integration with a large digital design on the same chip, and with similar reliability obstacles, however slightly more stringent. To linearise the output signal the PA bias is adjusted based on the input signal level, i.e. it uses adaptive bias. To support TDD mode operation, a TRX-switch was added, which for reliability uses always-on switches in TX-mode and always-off switches in RX-mode. The TRX-switch also provides a downward 1:2 impedance transformation in TX-mode to boost reachable output power, and in RX-mode it provides an upward impedance transformation of 2:1 for optimal noise figure.

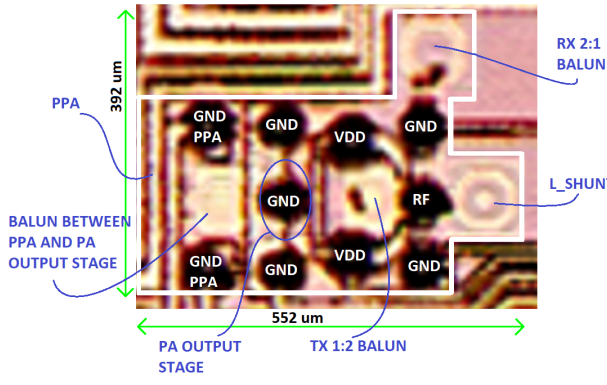

Figure 72: Die photo of the pre-power amplifier, the power amplifier and the TRX-switch.

7.2.2 Scientific Contribution

- Demonstration of a reliable integrated TRX-switch with low loss in both TX- and RX-mode for mmW frequencies.

- Successful demonstration of the first highly reconfigurable adaptive bias circuit to linearise the power amplifier. The adaptive bias improves the output referred 1 dB compression point by 3 dB and the saturated output power by 1.4 dB.

7.2.3 Comparison With State-of-the-art

A performance summary and comparison with state-of-the-art of the most relevant works at the time of the publication of paper II can be seen in table 8. As can be seen, the PA and PPA perform well in line with state-of-the-art and provide excellent gain control and have the smallest die area. The 1 dB output compression point is only 0.9 dB below the saturated ouput power as a result of the linearisation by the adaptive bias.

Table 8: Paper II Performance Summary and Comparison [2]

Parameter	**This Work**	APMC'18 [1]	RFIC'18 [68]	RFIC'17 [64]	RFIC'17 [65]
Technology [nm]	**22FDSOI CMOS**	28FDSOI CMOS	28FDSOI CMOS	28Bulk CMOS	28Bulk CMOS
Frequency [GHz]	**27**	26	31	27	32
Pwr.Combining	**1**	2	1	2	2
Gain Control [dB]	**28.8**	19	10	None	None
No. of Stages	**2**	2	2	2	2
Vdd [V]	**1.2**	1.5	0.7/1.98	1	1
P_{sat} [dBm]	**17.4**	20.6	17.9	18.1	19.8
P_{1dB} [dBm]	**16.5**	18.8	11.6	16.8	16
PAE_{max} [%]	**19.5**$^{(1)}$	22.6$^{(1)}$	25.5	41.5	21
PAE_{1dB} [%]	**17.3**$^{(1)}$	16.6$^{(1)}$	10	37.6	12.8
Power Gain [dB]	**34**$^{(2)}$	33.1$^{(2)}$	32.6	20.5	22
Area [mm^2]	**0.129**	0.144	0.508	0.361	0.59

$^{(1)}$Calculated from measured drain efficiency and simulated power gain. Not including PPA power consumption.
$^{(2)}$Simulated value since input signal is generated internally.

7.2.4 My Contribution

I'm the creator of the complete chosen circuit architecture and circuit analysis, simulations, and layout, including the idea of using the adaptive bias concept. The second author designed the static biasing, used for both PA and PPA. The third author did the design and layout of the pre-power amplifier. I did the lab measurement and manuscript writing. During the whole process the second and last author provided valuable feedback and support, but with emphasis on finalizing the manuscript.

7.3 Paper III: Efficient Wideband mmW Transceiver Front-End for 5G Base Stations in 22 nm FD-SOI CMOS

7.3.1 Overview

Paper III presents an mmW transceiver front-end including a TRX-switch. The transmitter has two paths, one providing input signal to the main amplifier and one to the auxiliary amplifier of a Doherty PA. Each path is equipped with a double balanced IQ-mixer for frequency upconversion and a pre-power amplifier with 16 gain steps. The required 90 degrees phase shift between the input of the main and auxiliary amplifier of the Doherty PA is achieved by rearranging the LO phases to the IQ-mixer of the auxiliary path. A TRX-switch that for reliability reasons uses TX ON switches connects the antenna with the transmitter and the receiver for TDD mode operation. The receiver consists of two LNA stages, the first one being single-ended and the second one differential. In addition, the transceiver contains LO IQ signal generation and distribution.

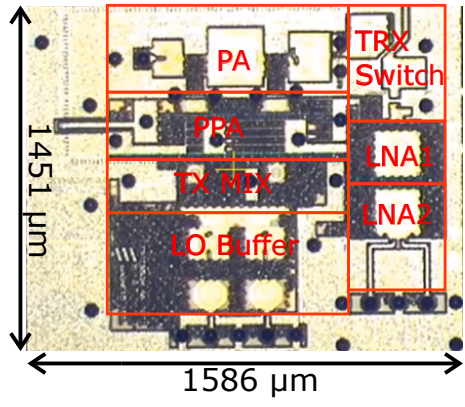

Figure 73: Die photo of the complete transceiver.

7.3.2 Scientific Contribution

- First detailed analysis of performance gains brought by adaptive bias for high bandwidth modulated signals.
- Record high modulated transmitter output power and efficiency for high integration mmW transceivers.
- Record low noise figure for high integration mmW transceivers.
- State-of-the-art efficiency for OFDM signals above 400 MHz bandwidth.

- Novel TRX-switch with built in LNA impedance matching is presented and analysed in detail.
- Fundamental BW problem associated with Doherty amplifiers was eliminated at the input, using quadrature mixers to generate 90° phase shift.
- LO leakage calibration achieving excellent measured leakage levels presented.
- Design strategy for output combining network of a current combining Doherty amplifier.
- Design methodology for power supply network and decoupling.

7.3.3 Comparison With State-of-the-art

A performance summary and comparison with state-of-the-art of the most relevant works at the time of the publication of paper III can be seen in table 9. The transmitter part of the transceiver front-end achieves the highest measured saturated output power and the highest modulated output power for 400/800/1600 MHz IBW for all OFDM modulated signals. Ignoring power consumption related to frequency conversion (LO-Buffers and Mixers), [69] reports high modulated efficiency, however not for OFDM but for single-carrier signals. For OFDM transmissions, the transmitter achieves the highest modulated efficiency for all modulation BW exceeding 400 MHz. Furthermore, the transmitter is measured with signals with the highest bandwidth and raw data rate. The receiver NF is the lowest reported for all front-ends.

7.3.4 My Contribution

I'm the creator of the complete chosen circuit architecture and responsible for the design, circuit analysis, simulations, and layout of the Doherty PA, including adaptive bias and output combiner network. In addition I had a supervising/guiding role of the other blocks in the transceiver. I did the lab measurement and manuscript writing. From the design phase and onwards, the second author provided valuable feedback, support and impact on the architecture and block designs and with design responsibility of the TRX-switch. The third author was responsible for the design of LNA1. The fourth author was responsible for the design of the mixers. The fifth author was responsible for the design of the PPA. The sixth author was responsible for the design of the LNA2. The seventh author was responsible for the design of IQ-generation and LO distribution. The eighth author provided significant assistance with modulated signal measurements. The ninth author did analysis of modulated transmitter performance during the design phase. The last author provided valuable feedback during the whole process and with emphasis on finalizing the manuscript.

Table 9: Paper III Performance Summary and Comparison [3]

Parameter	**This Work**		Yi [70] [JSSC'22]		Quadrelli [71] [JSSC'22]	Pashaeifar [72] [JSSC'21]	Zhu [69] [TMTT'21]		Pang [73] [JSSC'20]		Park [74] [ISSCC'19]	
Technology	**22 nm CMOS FDSOI**		65 nm CMOS Bulk		28 nm CMOS Bulk	40 nm CMOS Bulk	65 nm CMOS Bulk		65 nm CMOS Bulk		28 nm CMOS Bulk	
TRX	**Single TRX**		4xTRX		Single TRX	Single TX	4xTRX		8xTRX		Single TRX	
TRX type	**TX zero IF RX RF only**		TRX RF Only		TRX RF-IF	TX zero IF No RX	RF only		Bidirectional RF only 2-pol		RF only	
	24.25-29.5 GHz		24-29.5 GHz		22-31 GHz	24-30 GHz	24-28 GHz		26.5-29.5 GHz		25.8-29 GHz	
Area (mm^2)	**2.3**		4.5[3]		1.4	1.38	0.94		0.58		0.67	
Freq. Upconversion	**Yes**		No		Yes	Yes	No		No		No	
Switch	**Yes**		Yes		Yes	No	Yes		No		Yes	
Switch Loss (dB)	**1.1 in TX** **1.2 in RX**		1 in TX 0.5 in RX		N.R. N.R.	n/a n/a	1		n/a		n/a	
Compared Freq.	**26.5 GHz**		26 GHz		28 GHz	27 GHz	28 GHz		28 GHz		28 GHz	
TX/Element												
P1dB (dBm)	**15.9**		17.6		11.5	17.9[2]	16.1		10.3[2]		13.9	
Psat (dBm)	**18.3**		18		14	18.9[2]	18.1		14.1[2]		17	
PAE P1dB (%)	**13.4**		20.4		17	27[2]	16.6		6.4[2]		16	
Gain (dB)	**24**		23		20	20.7[2]	33		19[2]		17	
RF BW (GHz)	**2.4**		6		9	6[3]	4		3[3]		4	
RX/Element					7,17							
NF (dB)	**4.2**		4.3		8.5	n/a	4.5		5.2[2]		4.5	
P1dB (dBm)	**>-19.6**		-22.2		-18.8	n/a	-16.1		-26.8[2]		N.R.	
Gain (dB)	**23**		14.2		29	n/a	27.2		15[2]		N.R.	
RF BW (GHz)	**4**		6		9	n/a	4		3[3]		4	
PDC (mW)	**65.6**		82		110	n/a	45		112		44	
Modulated Meas.	**Cond.**		Cond.		Cond.	Cond.	Cond.		OTA		Cond.	
IRR (dB)	**44.3**		n/a		40	>50	n/a		n/a		n/a	
LO-leakage (dB)	**-41.4**		n/a		-30	-45	n/a		n/a		n/a	
Signal type	**5G OFDM**		5G OFDM		5G OFDM	5G OFDM	S.C. wideband		5G OFDM		OFDM	
Signal BW (MHz)	**16x100**	**16x100**	1200	400	400	8x100	540	405	400	400	100	100
QAM-Constellation	**16**	**64**	64	256	64	64	64	256	64	256	64	256
Symb. Rate Gs/s	**1.52**	**1.52**	1.14	0.38	0.38	0.76	0.4	0.3	0.38	0.38	0.1	0.1
Raw Data Rate Gb/s	**6.1**	**9.1**	6.84	3.04	2.28	4.56	2.4	2.4	2.3	3.04	0.6	0.8
EVM (%)	**10.2**	**6.7**	5.6	3.5	5.01	4.41	4.8	2.9	1.4	1.4	4.9	3.1
Pout (dBm)	**12.2**	**9**	3.9	6	5.38	7.3[2]	13.5	10.2	5.4[2]	4.5[2]	8.45	7.1
TX PAE (%)	**7.7**	**5.1**	2[1]	3.7	4.5[1]	5.8	8.4	3.5	<2	<2	5[1]	3.5[1]
TX PAE incl freq conv (%)	**5.3**	**3.2**	n/a	n/a	<2.5[1]	3.5	n/a	n/a	n/a	n/a	n/a	n/a

N.R. Not reported, n/a Not applicable, S.C. Single Carrier, [1] Graphically estimated. [2] Assuming 1.1 dB TRX-switch loss for comparison. [3] Ignoring BW limitations by TRX-switch. Yi JSSC'22: For 400 MHz BW 10.3 dBm P_{out} with 7.6 % PAE is reported. Pashaeifar JSSC'21: TX PAE and TX PAE including frequency conversion is estimated after TRX-switch loss and using power break down graph in Fig. 10. Pang JSSC'20: Modulated P_{out} is calculated from EIRP for 8 elements and combined PCB losses and antenna gain of -4 dB. TX PAE is estimated from figure 19. Park ISSCC'19: TX PAE is graphically estimated from figure 9.8.3.

7.4 Paper IV: Analysis and Design of a GHz Bandwidth Adaptive Bias Circuit for an mmW Doherty Amplifier

7.4.1 Overview

Paper IV derives fundamental equations for output power, impedance, and efficiency of the main and auxiliary amplifier as well as the complete Doherty amplifier, using a simplified ideal frequency agnostic transistor model suitable for hand calculations. Even with this ideal transistor model it is found that the Doherty amplifier is fundamentally nonlinear due to saturation of the main amplifier and the class C nonlinearity of the auxiliary. Furthermore, the paper analyses the use of a signal dependent bias level, commonly referred to as adaptive bias, which offers the possibility to control the output current characteristic and conduction angle of the auxiliary amplifier. The impact of

having a band-limited adaptive bias signal is investigated in combination with an OFDM modulated output signal. To verify the theoretical predictions, the design and measurements of the integrated adaptive bias circuit tailored for high PAR high bandwidth OFDM signals, for the mmW Doherty amplifier in 22nm CMOS FD-SOI in paper III, are investigated in more detail. Controllability of the adaptive bias circuit, to enhance the gains brought by it, are measured using continuous wave tone stimuli. The complete mmW transmitter is measured and compared with and without adaptive bias.

7.4.2 Scientific Contribution

- A theoretical explanation of the fundamental nonlinearities in an ideal Doherty amplifier using ideal transistors.
- First theoretical analysis of performance gains brought by adaptive bias for high bandwidth modulated signals.
- A high bandwidth, highly controllable adaptive bias circuit design is presented analysed and measured.
- Theoretical predictions are verified using circuit measurements.

7.4.3 Comparison With State-of-the-art

Due to lack of similar works no comparison table was compiled.

7.4.4 My Contribution

I'm the initiator of the idea to analytically analyse the basic operation of a Doherty amplifier, and to analyse the impact of using adaptive bias to linearise the output current from the class C biased auxiliary amplifier. I did the analytical circuit derivations and produced the resulting plots, and the supporting measurements and comparisons with the theory. I compiled the manuscript. The last author provided valuable feedback during the whole process and with emphasis on finalizing the manuscript.

Chapter 8

Conclusions, Discussion, and Future Work

Three mmW power amplifier circuits together with a theoretical investigation is the foundation of this dissertation. The implementation complexity, of the three PAs, is gradually increased, resulting in a highly advanced and close to complete mmW transmitter in paper III. The fourth paper, which theoretically investigates adaptive bias in combination with the Doherty amplifier, wraps up the thesis by analyzing the idea of using adaptive bias as introduced in paper II and the details for how it was successfully used in paper III.

8.1 Conclusions and Discussion

Papers I and II can be seen as steps in a trajectory towards paper III. The main outcome from the first step (paper I) was the successful demonstration of power combining two class AB PAs, which led to state-of-the-art saturated output power and 1 dB compression point. In the next step (paper II), the power combination was replaced by using an on-chip balun for impedance transformation to reach high output power, and the adaptive bias concept was introduced and tested, but not in detail analysed and verified for modulated high bandwidth signals. In addition, the TRX-switch, supporting TDD mode operation, was added between the PA and the antenna. To reduce power consumption, CSCG unit cells, instead of current steering, were introduced in the PPA for changing its gain. In the last step (paper III), the TRX-switch was kept but modified to incorporate the impedance matching of the LNA. The impedance transformation to reduce the load impedance and boost the output power was also kept. The efficiency was addressed by introducing a Doherty amplifier, which also brought back the power combination, which is inherent in the Doherty amplifier. Adaptive bias was kept, with key learnings from paper II, but now in detail analyzed mainly in paper IV to linearise the fundamentally nonlinear Doherty PA. A small power saving modification was added to the PPA variable gain functionality as the design was reverted back to current steering, but to a dedicated supply, which enables possibility to save power by reducing the supply voltage that the unused signal current is steered into. Two complex

quadrature mixers were used, one for the main path and one for the auxiliary. The required 90° phase shift of the input signal was achieved by rearranging the LO phases to the auxiliary mixer, which eliminated the input signal part of the fundamental Doherty bandwidth problem. Moreover, to reduce supply net impedance at base band frequencies below the signal bandwidth, a decoupling strategy using small damping resistors was developed.

8.2 Future Work

Naturally, there is a never ending list of possible ideas of research opportunities that could be investigated, and here are some ideas that could be interesting to pursue.

- In paper III, as a proof of concept, digital predistortion of the input signal to the transmitter could be investigated. In principle this should be possible, however, there might be unknown effects since the adaptive bias itself linearises the PA, but with limited bandwidth, which could complicate the implementation of a DPD.
- In paper III, fundamental bandwidth problems were solved for the input part of the Doherty amplifier, and it would be interesting to investigate what could be done to further improve the bandwidth of the output combiner.
- In paper IV, next step is to deploy adaptive bias also on the main amplifier to perform linearisation in two steps 1) linearisation of transistor compression 2) linearisation of fundamental nonlinearity from the Doherty amplifier as concluded in paper IV. The analysis should be carried out both for CW and modulated signals.
- Further increase the gains brought by the adaptive bias by optimising the current consumption at back-off as described in [75]. The idea is to use the adaptive bias signal to turn off or reduce the bias to a larger portion of the auxiliary path, such as the PPA and the mixer. The technique has potential to significantly improve the transmitter back-off efficiency.
- Add downconversion and baseband part to the receiver in paper III. An interesting detail to investigate could be to enable a high IRR by a phase/amplitude calibration.
- Add an integrated LO solution on the same die.
- Investigate the performance of the PA with a mismatched load and integrate the Doherty architecture together with a novel, or any of the known techniques, to increase its VSWR resilience.

Appendix A

Frequency Spectrum of a Two Tone Stimuli of an Instantaneous Third-Order Nonlinear System

A third-order instantaneous nonlinear system is described by equation 53 in chapter 5 and here repeated for convenience.

$$v_{out}(t) = k_1 v_{in}(t) + k_2 v_{in}^2(t) + k_3 v_{in}^3(t) \tag{88}$$

The input input signal is chosen as two cosine tones with amplitude A_1 and A_2 with angular frequency ω_1 and ω_2.

$$v_{in}(t) = A_1 cos(\omega_1 t) + A_2 cos(\omega_2 t) \tag{89}$$

The frequency components of v_{out} of equation 88 with input v_{in} as in Eq. 89 is summarized in table 10 below [30].

Table 10: Third Order Nonlinear Two Tone Test Response

Frequency	**Amplitude**	**Name**	**Order**
ω_1	$k_1 A_1$	linear	1
ω_2	$k_1 A_2$		1
ω_1	$k_3 \frac{3}{4} A_1^3$	compression	3
ω_2	$k_3 \frac{3}{4} A_2^3$		3
ω_1	$k_3 \frac{3}{2} A_1 A_2^2$	cross modulation	3
ω_2	$k_3 \frac{3}{2} A_1^2 A_2$		3
0	$k_2 A_1^2$	dc shift	2
0	$k_2 A_2^2$		2
$2\omega_1$	$k_2 \frac{1}{2} A_1^2$	harmonic	2
$2\omega_2$	$k_2 \frac{1}{2} A_2^2$		2
$3\omega_1$	$k_3 \frac{1}{4} A_1^3$		3
$3\omega_2$	$k_3 \frac{1}{4} A_2^3$		3
$\omega_1 + \omega_2$	$k_2 A_1 A_2$	intermodulation	2
$\omega_1 - \omega_2$	$k_2 A_1 A_2$		2
$2\omega_1 + \omega_2$	$k_3 \frac{3}{4} A_1^2 A_2$		3
$\omega_1 + 2\omega_2$	$k_3 \frac{3}{4} A_1 A_2^2$		3
$2\omega_1 - \omega_2$	$k_3 \frac{3}{4} A_1^2 A_2$		3
$2\omega_2 - \omega_1$	$k_3 \frac{3}{4} A_1 A_2^2$		3

References

[1] C. Elgaard, A. Axholt, E. Westesson, and H. Sjöland, "A 26GHz 22.2dBm Variable Gain Power Amplifier in 28nm FD-SOI CMOS for 5G Antenna Arrays," in *2018 Asia-Pacific Microwave Conference (APMC)*, 2018, pp. 965–967.

[2] C. Elgaard, S. Andersson, P. Caputa, E. Westesson, and H. Sjöland, "A 27 GHz Adaptive Bias Variable Gain Power Amplifier and T/R Switch in 22nm FD-SOI CMOS for 5G Antenna Arrays," in *2019 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2019, pp. 303–306.

[3] C. Elgaard, M. Özen, E. Westesson, A. Mahmoud, F. Torres, S. B. Reyaz, T. Forsberg, R. Akbar, H. Hagberg, and H. Sjöland, "Efficient Wideband mmW Transceiver Front End for 5G Base Stations in 22-nm FD-SOI CMOS," *IEEE Journal of Solid-State Circuits*, vol. 59, no. 2, pp. 321–336, 2024.

[4] C. Elgaard and H. Sjöland, "Analysis and Design of a GHz Bandwidth Adaptive Bias Circuit for an mmW Doherty Amplifier," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Submitted Feb 2024.

[5] C. Elgaard and L. Sundström, "A 491.52 mhz 840 uw crystal oscillator in 28 nm fd-soi cmos for 5g applications," in *ESSCIRC 2017 - 43rd IEEE European Solid State Circuits Conference*, 2017, pp. 247–250.

[6] S. Ek, T. Påhlsson, C. Elgaard, A. Carlsson, A. Axholt, A.-K. Stenman, L. Sundström, and H. Sjöland, "A 28-nm FD-SOI 115-fs Jitter PLL-Based LO System for 24–30-GHz Sliding-IF 5G Transceivers," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 7, pp. 1988–2000, 2018.

[7] Ericsson, "Ericsson mobility report," https://www.ericsson.com/assets/local/reports-papers/mobility-report/documents/2023/ericsson-mobility-report-june-2023.pdf, June 2023, [Online; accessed 3-Aug-2023].

[8] Cisco, "Cisco Annual Internet Report," https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html, March 2020, [Online; accessed 11-Nov-2023].

[9] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 4, pp. 623–656, 1948.

[10] H. Friis, "A note on a simple transmission formula," *Proceedings of the IRE*, vol. 34, no. 5, pp. 254–256, 1946.

[11] E. Johnson, "Physical limitations on frequency and power parameters of transistors," in *1958 IRE International Convention Record*, vol. 13, 1965, pp. 27–34.

[12] P. Henricsson, "Första mm-vågsantennen i kisel," Internet:https://etn.se/index.php/nyheter/63189-foersta-millimetervagsantennen-i-kisel, Feb. 2017.

[13] 3GPP, "Technical Specification 38.104 (V15.2.0) Base Station (BS) Radio Transmission and Reception," Jul. 2018.

[14] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits.* 2nd ed. Cambridge, UK: Cambridge University Press, 2004.

[15] F. Churchill, G. Ogar, and B. Thompson, "The Correction of I and Q Errors in a Coherent Processor," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-17, no. 1, pp. 131–137, 1981.

[16] J. Groe, "Polar Transmitters for Wireless Communications," *IEEE Communications Magazine*, vol. 45, no. 9, pp. 58–63, 2007.

[17] L. Rong, "All Digital Polar Transmitter Design for Software Defined Radio," Ph.D. dissertation, Royal Institute of Technology, Stockholm, Sweden, 2012.

[18] J. Lindstrand, "Integrated Transmitters for Cellular User Equipment-Wideband CMOS Power Amplifier and Antenna Impedance Tuners," Ph.D. dissertation, Lund University, Lund, Sweden, 2019.

[19] H. Zhang and E. Sánchez-Sinencio, "Linearization Techniques for CMOS Low Noise Amplifiers: A Tutorial," *IEEE Trans. Circuits Syst. I*, vol. 58, no. 1, pp. 22–36, Jan. 2011.

[20] S. Cripps, *RF Power Amplifiers for Wireless Communications*, 1st ed. Norwood, MA, USA: Artech House, 1999.

[21] P. Reynaert and M. Steyaert, *RF Power Amplifiers for Mobile Communications*, 1st ed. Dordrecht, The Netherlands: Springer, 2006.

[22] D. Chowdhury, L. Ye, E. Alon, and A. M. Niknejad, "An Efficient Mixed-Signal 2.4-GHz Polar Power Amplifier in 65-nm CMOS Technology," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 8, pp. 1796–1809, Aug 2011.

[23] H. Kobayashi, J. M. Hinrichs, and P. M. Asbeck, "Current-mode class-D power amplifiers for high-efficiency RF applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 12, pp. 2480–2485, Dec 2001.

[24] N. O. Sokal and A. D. Sokal, "Class E-A new class of high-efficiency tuned single-ended switching power amplifiers," *IEEE J. Solid-State Circuits*, vol. 10, no. 3, pp. 168–176, 1975.

[25] A. Grebenikov and N. O. Sokal, *Switchmode RF Power Amplifiers*. 1st ed. Boston, MA, USA: Newnes, 2007.

[26] S. Hykin, *Communication Systems*, 2nd ed. New York, NY, USA: John Whiley & Sons, 1983.

[27] J. Vuolevi and T. Rahkonen, *Distortion in RF Power Amplifiers*. 1st ed. Norwood, MA, USA: Artech House Professional, 2003.

[28] Wikipedia The Free Encyclopedia, "Nonlinear systems," Internet: https://en.wikipedia.org/wiki/Nonlinearsystem, Aug. 2020.

[29] E. Bedrosian and S. O. Rice, "The output properties of Volterra systems (nonlinear systems with memory) driven by harmonic and Gaussian inputs," *Proceedings of the IEEE*, vol. 59, no. 12, pp. 1688–1707, 1971.

[30] M. Lantz, "Systematic Design of Linear Feedback Amplifiers," Ph.D. dissertation, Lund University, Lund, Sweden, 2002.

[31] Wikipedia The Free Encyclopedia, "Volterra Series," Internet: https://en.wikipedia.org/wiki/Volterraseries, Aug. 2020.

[32] S. Narayanan, "Transistor distortion analysis using volterra series representation," *Bell Systems Technical Journal*, vol. 46, no. 5, pp. 991–1024, 1967.

[33] K. Sanghoon, C. Byounggi, and K. Bumman, "Linearity analysis of CMOS for RF application," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 3, pp. 972–977, 2003.

[34] Y. Tsividis and C. McAndrew, *Operation and Modeling of the Mos Transistor*, 3rd ed. New York, NY, USA: Oxford University Press, 2011.

[35] D. Webster, J. Scott, and D. Haigh, "Control of circuit distortion by the derivative superposition method [MMIC amplifier]," *IEEE Microwave and Guided Wave Letters*, vol. 6, no. 3, pp. 123–125, 1996.

[36] J. Kim, B. Fehri, S. Boumaiza, and J. Wood, "Power Efficiency and Linearity Enhancement Using Optimized Asymmetrical Doherty Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 2, pp. 425–434, 2011.

[37] J. H. K. Vuolevi, T. Rahkonen, and J. P. A. Manninen, "Measurement technique for characterizing memory effects in RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 8, pp. 1383–1389, 2001.

[38] E. Schurack, W. Rupp, T. Latzel, and A. Gottwald, "Analysis and measurement of nonlinear effects in power amplifiers caused by thermal power feedback," in *[Proceedings] 1992 IEEE International Symposium on Circuits and Systems*, vol. 2, 1992, pp. 758–761.

[39] S. Lee, S. Lee, S. Kang, and S. Hong, "A 1.7-GHz GaN MMIC Doherty power amplifier using an adaptive bias circuit with a quadrature coupler," in *2017 IEEE International Symposium on Radio-Frequency Integration Technology (RFIT)*, 2017, pp. 90–92.

[40] R.-Z. Zhan, Y. C. Li, and H. Zhang, "A Highly Efficient 60 GHz CMOS Doherty Power Amplifier With Adaptive Gate Biases," in *2020 IEEE MTT-S International Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications (IMWS-AMP)*, 2020, pp. 1–3.

[41] H. Zhang, R.-Z. Zhan, Y. C. Li, and J. Mou, "High Efficiency Doherty Power Amplifier Using Dual-Adaptive Biases," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 8, pp. 2625–2634, 2020.

[42] C. R. Chappidi, T. Sharma, Z. Liu, and K. Sengupta, "Load Modulated Balanced mm-Wave CMOS PA with Integrated Linearity Enhancement for 5G applications," in *2020 IEEE/MTT-S International Microwave Symposium (IMS)*, 2020, pp. 1101–1104.

[43] Y. Jin and S. Hong, "A 24-GHz CMOS Power Amplifier With Dynamic Feedback and Adaptive Bias Controls," *IEEE Microwave and Wireless Components Letters*, vol. 31, no. 2, pp. 153–156, 2021.

[44] S. Chen, G. Wang, Z. Cheng, P. Qin, and Q. Xue, "Adaptively Biased 60-GHz Doherty Power Amplifier in 65-nm CMOS," *IEEE Microwave and Wireless Components Letters*, vol. 27, no. 3, pp. 296–298, 2017.

[45] H. T. Nguyen, T. Chi, S. Li, and H. Wang, "A 62-to-68GHz linear 6Gb/s 64QAM CMOS doherty radiator with 27.5leveraging high-efficiency multi-feed antenna-based active load modulation," in *2018 IEEE International Solid-State Circuits Conference - (ISSCC)*, 2018, pp. 402–404.

[46] H. T. Nguyen and H. Wang, "A Coupler-Based Differential Doherty Power Amplifier with Built-In Baluns for High Mm-Wave Linear-Yet-Efficient Gbit/s Amplifications," in *2019 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2019, pp. 195–198.

[47] H. Zhang, R.-Z. Zhan, Y. C. Li, and J. Mou, "High Efficiency Doherty Power Amplifier Using Dual-Adaptive Biases," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 8, pp. 2625–2634, 2020.

[48] Z. Zong, X. Tang, K. Khalaf, D. Yan, G. Mangraviti, J. Nguyen, Y. Liu, and P. Wambacq, "A 28-GHz SOI-CMOS Doherty Power Amplifier With a Compact Transformer-Based Output Combiner," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 6, pp. 2795–2808, 2021.

[49] M. Pashaeifar, L. C. N. de Vreede, and M. S. Alavi, "A Millimeter-Wave CMOS Series-Doherty Power Amplifier With Post-Silicon Inter-Stage Passive Validation," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 10, pp. 2999–3013, 2022.

[50] D. Webster, D. Haigh, J. Scott, and A. Parker, "Derivative superposition-a linearisation technique for ultra broadband systems," in *IEE Colloquium Wideband Circuits, Modelling and Techniques*, 1996, pp. 3/1–3/14.

[51] Analog Devices, "Pick a Plug n Play Linearizer for Your 5G RF Power Amplifier," https://www.analog.com/en/technical-articles/pick-a-plug-n-play-linearizer-for-your-5g-rf-power-amplifier.html, Jan 2024, [Online; Accessed 12-Jan-2024].

[52] S. Jung, H. Park, M. Kim, G. Ahn, J. Van, H. Hwangbo, C. Park, S. Park, and Y. Yang, "A New Envelope Predistorter With Envelope Delay Taps for Memory Effect Compensation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 1, pp. 52–59, 2007.

[53] F. Roger, "A 200mW 100MHz-to-4GHz 11th-order complex analog memory polynomial predistorter for wireless infrastructure RF amplifiers," in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2013, pp. 94–95.

[54] N. Rostomyan, J. A. Jayamon, and P. M. Asbeck, "15 GHz Doherty Power Amplifier With RF Predistortion Linearizer in CMOS SOI," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 3, pp. 1339–1348, 2018.

[55] Wikipedia The Free Encyclopedia, "Nyqvist Frequency," Internet: https://en.wikipedia.org/wiki/Nyqvistfrequency, Jan. 2021.

[56] A. F. Molisch, *Wireless Communications*. 2nd ed. New York, NY, USA: John Whiley & Sons, 2011.

[57] W. Doherty, "A New High Efficiency Power Amplifier for Modulated Waves," *Proceedings of the Institute of Radio Engineers*, vol. 24, no. 9, pp. 1163–1182, 1936.

[58] H. Chireix, "High Power Outphasing Modulation," *Proceedings of the Institute of Radio Engineers*, vol. 23, no. 11, pp. 1370–1392, 1935.

[59] S. Cripps, "PA Design: Steve Cripps on Outphasing PAs," Internet: https://www.youtube.com/watchv-dw5gvchDFQ, 2018.

[60] B. Shi and L. Sundstrom, "A 200-MHz IF BiCMOS signal component separator for linear LINC transmitters," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 7, pp. 987–993, 2000.

[61] T. W. Barton and D. J. Perreault, "Theory and Implementation of RF-Input Outphasing Power Amplification," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 12, pp. 4273–4283, 2015.

[62] D. N. Martin and T. W. Barton, "Inphasing Signal Component Separation for an X-Band Outphasing Power Amplifier," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 3, pp. 1661–1674, 2021.

[63] F. Raab, "Efficiency of Outphasing RF Power-Amplifier Systems," *IEEE Transactions on Communications*, vol. 33, no. 10, pp. 1094–1099, 1985.

[64] Y. Zhang and P. Reynaert, "A high-efficiency linear power amplifier for 28ghz mobile communications in 40nm cmos," in *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2017, pp. 33–36.

[65] P. Indirayanti and P. Reynaert, "A 32 ghz 20 dbm-psat transformer-based doherty power amplifier for multi-gb/s 5g applications in 28 nm bulk cmos," in *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2017, pp. 45–48.

[66] S. Shakib, H.-C. Park, J. Dunworth, V. Aparin, and K. Entesari, "20.6 a 28ghz efficient linear power amplifier for 5g phased arrays in 28nm bulk cmos," in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, 2016, pp. 352–353.

[67] B. Park, D. Jeong, J. Kim, Y. Cho, K. Moon, and B. Kim, "Highly linear cmos power amplifier for mm-wave applications," in *2016 IEEE MTT-S International Microwave Symposium (IMS)*, 2016, pp. 1–3.

[68] F. Torres, M. De Matos, A. Cathelin, and E. Kerhervé, "A 31 ghz 2-stage reconfigurable balanced power amplifier with 32.6db power gain, 25.5and 17.9dbm psatin 28nm fd-soi cmos," in *2018 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2018, pp. 236–239.

[69] W. Zhu, J. Wang, X. Zhang, W. Lv, B. Liao, Y. Zhu, and Y. Wang, "A 24–28-GHz Four-Element Phased-Array Transceiver Front End With

21.1%/16.6% Transmitter Peak/OP1dB PAE and Subdegree Phase Resolution Supporting 2.4 Gb/s in 256-QAM for 5-G Communications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 6, pp. 2854–2869, 2021.

[70] Y. Yi, D. Zhao, J. Zhang, P. Gu, Y. Chai, H. Liu, and X. You, "A 24–29.5-GHz Highly Linear Phased-Array Transceiver Front-End in 65-nm CMOS Supporting 800-MHz 64-QAM and 400-MHz 256-QAM for 5G New Radio," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 9, pp. 2702–2718, 2022.

[71] F. Quadrelli, D. Manente, D. Seebacher, F. Padovan, M. Bassi, A. Mazzanti, and A. Bevilacqua, "A broadband 22–31-ghz bidirectional image-reject up/down converter module in 28-nm cmos for 5g communications," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 7, pp. 1968–1981, 2022.

[72] M. Pashaeifar, L. C. N. de Vreede, and M. S. Alavi, "14.4 A 24-to-30GHz Double-Quadrature Direct-Upconversion Transmitter with Mutual-Coupling-Resilient Series-Doherty Balanced PA for 5G MIMO Arrays," vol. 64, pp. 223–225, 2021.

[73] J. Pang, Z. Li, R. Kubozoe, X. Luo, R. Wu, Y. Wang, D. You, A. A. Fadila, R. Saengchan, T. Nakamura, J. Alvin, D. Matsumoto, B. Liu, A. T. Narayanan, J. Qiu, H. Liu, Z. Sun, H. Huang, K. K. Tokgoz, K. Motoi, N. Oshima, S. Hori, K. Kunihiro, T. Kaneko, A. Shirane, and K. Okada, "A 28-GHz CMOS Phased-Array Beamformer Utilizing Neutralized Bi-Directional Technique Supporting Dual-Polarized MIMO for 5G NR," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 9, pp. 2371–2386, 2020.

[74] J. Park, S. Lee, D. Lee, and S. Hong, "A 28GHz 20.3%-Transmitter-Efficiency 1.5°-Phase-Error Beamforming Front-End IC with Embedded Switches and Dual-Vector Variable-Gain Phase Shifters," in *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2019, pp. 176–178.

[75] C. Elgaard, S. Mattisson, and S. Andersson, "Envelope Controlled Biasing of an Auxiliary Transmitter of a Doherty Amplifier," US Patent Application US2023179150 (A1), 2023.

Paper I

Paper I

A 26GHz 22.2dBm Variable Gain Power Amplifier in 28nm FD-SOI CMOS for 5G Antenna Arrays

C. Elgaard, A. Axholt, E. Westesson, and H. Sjöland, "A 26GHz 22.2dBm Variable Gain Power Amplifier in 28nm FD-SOI CMOS for 5G Antenna Arrays," © 2018 IEEE, reprinted from *Proc. of IEEE APMC*, Kyoto, Japan, November 2018.

A 26GHz 22.2dBm Variable Gain Power Amplifier in 28nm FD-SOI CMOS for 5G Antenna Arrays

Christian Elgaard*, Andreas Axholt*, Eric Westesson*, and Henrik Sjöland*

* *Ericsson AB, Mobilvägen 1, SE-221 83, Lund, Sweden*

christian.elgaard@ericsson.com, henrik.sjoland@ericsson.com

***Abstract*—A 26 GHz power amplifier (PA) targeting millimeter wave 5G mobile systems is presented. The two stage PA, integrated in a complete transmitter in a 28 nm FD-SOI CMOS process, only occupies a die area of 0.144 mm^2. It uses stacking of two transistors to handle high voltage swing, and a transformer based power combiner at the output to reduce output load impedance. The first stage, the pre-PA (PPA), has 31 unit cells, that can accurately set the gain from 14.7-33.1 dB. The design has been optimized for device stress to provide function for >10 years. Saturated output power and 1 dB compression point reach state-of-the-art performance of 20.6 / 22.2 dBm and 18.8 / 20.7 dBm for 1.5 / 1.8 V supply respectively. For 1.5 V supply, PAE is peaking at 22.6 % and AM-PM is below 5 degrees up to the 1 dB compression point.**

***Index Terms*—5G; CMOS; Antenna Arrays; Integrated; Microwave Power Amplifier; Power Amplifier; SOI;**

I. Introduction

Fifth generation mobile systems will support millimeter-wave frequencies and antenna arrays systems (AAS). AAS featuring tens or hundreds of transceivers, where the antenna elements are separated by roughly $\lambda/2 \approx 6\,\mathrm{mm}$ at 26 GHz and assuming two power amplifiers (PAs) per antenna, one for each polarization, will not have space for external PAs, and thus integration in the CMOS transceiver chip is imperative. The 5G AAS require massive digital signal processing, but the chip size and power consumption are limited by the dense antenna array at 26 GHz. A technology node with small devices must thus be chosen, which makes the design of PAs with high output power a challenge. This paper presents the design and measurements of a variable gain PA, consisting of a PPA and two parallel PA output stages targeting 3GPP band n258 i.e. 24.25 - 27.5 GHz [1]. It is placed in a fully integrated transceiver chain, i.e. the PA input signal is generated on-chip by the preceding blocks, see Fig. 1. High output power is reached by using load impedance reduction, power combining, and device stacking. Device stacking is used to distribute the voltage stress, allowing the use of minimum oxide thickness transistors, without compromising long term reliability. In Section II the circuit design of the PPA, PA output stage, and output combiner is presented, followed by measurement results in Section III. Finally conclusions are drawn in Section IV.

II. Circuit Design

A. PPA

The inputs of the PPA are connected to transistor gates, see Fig. 2a, resulting in a capacitive input impedance. At 26 GHz it is equivalent to C = 120 fF in parallel with R = 416 Ω. Using a 312 pH inductor, with a Q value of 20, at the output of the preceding mixer will cancel the capacitive part, resulting in a real impedance of 296 Ω at 26 GHz. Most published PAs use an input impedance of 50 Ω for good matching with external instruments, which is a too heavy load for the preceding mixer stage, and thus will require an additional intermediate stage consuming both chip area and power. To mitigate gain variations in preceding stages of the transceiver chain, the PPA was designed using 32 unit cells of CSCG (common source common gate) stages, which can be turned on/off using a 5-bit control word. Unit cells that are turned off continue to draw bias current to avoid being damaged by excessive voltage stress, and are connected to dump signal current into Vdd instead of the PPA resonator tank, see Fig. 2a. The center frequency of the PPA output resonator tank can be tuned with a 3-bit control word and its Q-value was carefully chosen as a trade-off between gain, bandwidth, and compression point. Since the capacitance is dominated by the large voltage

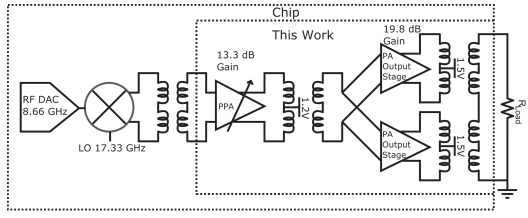

Fig. 1. The power amplifier architecture, part of a fully integrated transmitter

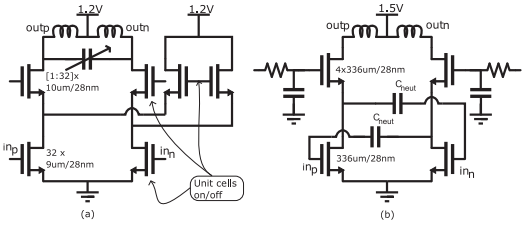

Fig. 2. (a) Schematic of the PPA (b) Schematic of one of the two PA output stages

dependent capacitance C_{gs} of the common source input stages of the PA, a higher Q-value will also result in a relatively larger gain shift as the center frequency moves downwards with increasing input signal [5]. This means that the center frequency of the PPA output resonance tank should ideally be placed slightly above the transmit frequency to optimize the compression point. Hence the Q-value is selected to 9 and the resonance frequency to 26.8 GHz for an output frequency of 26 GHz, to compensate for output stage gain compression.

B. PA Output Stage

The PA output stage, depicted in Fig. 2b, uses a fixed size CSCG (also known as cascode stage) stage to distribute both the DC and AC voltages over two transistors. A CG-stage normally has a signal grounded gate and thus low voltage signal swing at its source ($V_s \approx I_s/CG_{gm}$). However, to distribute the voltage swing evenly over the CS and CG transistors, which is effectively controlled by setting a ratio between the CG-stage C_{dg}, and a capacitor from gate to ground, through capacitive voltage division. In addition, this also reduces the effect of the CG C_{dg} capacitor on the output capacitance. Reducing the output capacitance is important to reduce losses in the output resonator tank, which scale proportionally with the parallel conductance $G = \omega C/Q$. The CS input stage uses cross coupled neutralization capacitors, C_{neut}, to reduce differential input capacitance and boost reverse isolation by cancelling input stage C_{gd}. After impedance reduction by the power combiner each side of the differential PA output stage is loaded by $12.5\,\Omega$. To drive this low impedance a large W/L ratio is required for the CSCG-stage, resulting in a large active area and increased layout challenges. The CSCG-stages of the PA output stage is about the same size as the output transformer, which makes it critical to reduce unwanted series inductance in the signal path. The CS gate bias voltage is set using a 5-bit current-DAC connected to a current mirroring transistor placed next the the active CS-stage, to quickly track thermal variations due to PA power dissipation. The CG gate bias voltage is controlled using a 3-bit resistor ladder DAC.

C. Output Combiner Design

The PA output combiner consists of two 1:1 transformers with the secondary sides connected in series. The inductance of the transformers is chosen to resonate with the drain capacitance at the center frequency. Combining power from two differential PA output stages comes at the cost of losses in the combining network. The simulated losses of the power combiner including routing to pad is 1.25-1.45 dB from 23-30 GHz. The primary and secondary coils of the transformers are octagon shaped with 12 μm wide metal traces, implemented in the thick upper copper / aluminum layers. To reduce the parasitic capacitance between the primary and secondary coils, the metal traces are offset by 4 μm from each other, which results in a diameter of 55 μm for the primary and 63 μm for the secondary. The slightly longer secondary loops boost the impedance transformation further.

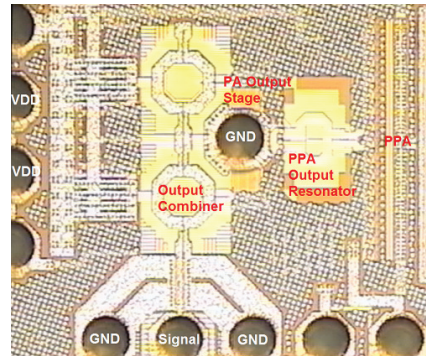

Fig. 3. Die photo showing the the PA. From right to left showing the PPA, PPA output resonator, output stage (partly placed under GND bump), and output combiner

III. Measurement Results

Two transmitter (TX) chains, containing the PA, were fabricated on a chip in 28 nm FD-SOI technology. It was flip-chip mounted in an FCBGA package and mounted on a PCB for measurements. The complete PA, including input resonance tank, PPA, PA output stages, output combiner, bias circuits and decoupling capacitors, occupied 0.144 mm^2, see Fig. 3.

The PA was measured using the on-chip RF DAC, LO generation, and mixer to generate the input signal to the PPA. The input signal amplitude was then controlled by a 9-bit digital value to the RF DAC. SSG (small signal gain) was measured with three different tuning settings of the resonance tank between the PPA and PA output stage, see Fig. 4a, on three different samples, see Fig. 4b. Since the input signal to the PA was not generated from an external instrument its absolute level was calculated from simulations. Including PPA the 3 dB bandwidth was measured to 3.3 GHz. Fig. 4c shows the SSG for five of the 31 available gain steps. The SSG peak shifts downwards by some 0.5 GHz as the gain is increased over its range. This effect is attributed to the increased capacitive loading of the PPA output balun, introduced by the increasing number of unit cells being active. The AM-AM performance was measured including the nonlinearities of the DAC. The individually generated LO signals within the two parallel TX-chains are phase locked to the same high frequency reference clock, as described in [2]. The AM-PM performance could therefore be obtained by measuring the phase difference between the two output signals, keeping the signal level constant in one chain and varying the input signal to the other. Measured and simulated AM-AM, AM-PM, and PAE are shown in Fig. 5. For 1.5 V supply the PA achieves 20.6 dBm saturated output power and 18.8 dBm of 1 dB compression point, 5 degrees maximum AM-PM offset up to 1 dB compression, and 16.6 % PAE_{1dB} excluding the PPA power consumption. For 1.5 V supply the PA is designed

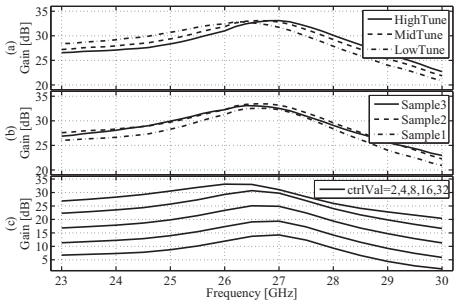

Fig. 4. (a) SSG for three different PPA tune values for sample 2 and 1.5 V supply, (b) SSG measured for three samples at mid PPA tune and 1.5 V supply, (c) PA SSG for five different PPA gain control settings and mid PPA tune and 1.5 V supply

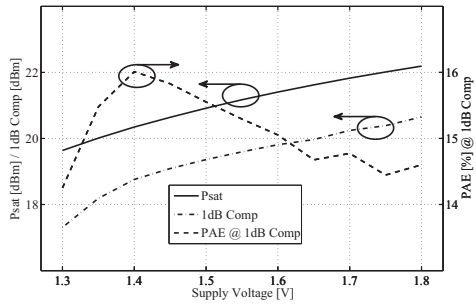

Fig. 6. Psat, 1dB compression point and PAE vs supply voltage.

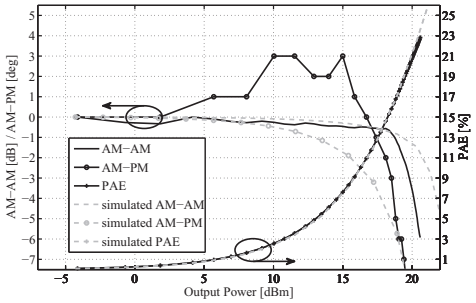

Fig. 5. Measured and simulated AM-AM, AM-PM and PAE plotted versus output power at 26 GHz and 1.5 V supply

to work for at least 10 years, using circuit simulator ELDO ageing tool, with an average on-time of 20 % using an OFDM signal with a PAR of 7 dB and an average output power of 12.7 dBm. Depending on requirement for saturated output power, compression point, efficiency, and numbers of operating hours, the supply voltage can be reduced or increased. Their respective dependencies are shown in Fig. 6. Table I summarises the performance and compares with state-of-the-art PAs in 28 nm CMOS for the same frequencies.

IV. Conclusions

An efficient, highly linear variable gain power combining 26 GHz power amplifier in 28 nm FD-SOI CMOS has been presented. Compared to previously published 28 nm CMOS designs in the same frequency range, it reaches state-of-the-art performance for both saturated output power and 1 dB compression point without sacrificing life span. The PA is integrated in a transmitter chain and includes a variable gain PPA.

TABLE I
Performance Summary and Comparison

Parameter	This Work		RFIC'17 [3]	RFIC'17 [4]	ISSCC'16 [5]		IMS'16 [6]	
Tech. [nm]	28 SOI CMOS		28 Bulk CMOS	28 Bulk CMOS	28 Bulk CMOS		28 Bulk CMOS	
Freq. [GHz]	26		27	32	30		28	
Pwr.Comb.	2		2	2	1		1	
Gain Control	5 bits, 19dB		None	None	None		None	
No. of Stages	2		2	2	2		1	
Vdd [V]	1.5	1.8	1	1	1	1.15	1.1	2.2
P_{sat} [dBm]	20.6	22.2	18.1	19.8	14	15.3	14.8	19.8
P_{1dB} [dBm]	18.8	20.7	16.8	16	13.2	14.3	14.0	18.6
PAE_{max} [%]	22.6(1)	21.3(1)	41.5	21	35.5	36.6	36.5	43.3
PAE_{1dB} [%]	16.6(1)	14.6(1)	37.6	12.8	34.3	35.8	35.2	41.4
PAE_{max} [%]	20(2)	19.6(2)	41.5	21	35.5	36.6	36.5	43.3
PAE_{1dB} [%]	14.5(2)	13.5(2)	37.6	12.8	34.3	35.8	35.2	41.4
Gain [dB]	33.1(3)	33.9(3)	20.5	22	15.7	16.3	10.0	13.6
Area [mm^2]	0.144		0.361	0.59	0.16		0.28	

(1) Not including PPA power consumption.
(2) Including PPA power consumption.
(3) Simulated value since input signal is generated internally. Also used for calculating PAE values.

References

[1] 3GPP TS 38.104 V15.0.0 (2017-12), Technical Specification, 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; Base Stations (BS) radio transmission and reception (Release 15). http://www.3gpp.org/ftp/Specs/archive/38_series/38.104/

[2] S. Ek, T. Påhlsson, A. Carlsson, A. Axholt, A. K. Stenman and H. Sjöland, "A 16–20 GHz LO system with 115 fs jitter for 24-30 GHz 5G in 28 nm FD-SOI CMOS," ESSCIRC 2017 - 43rd IEEE European Solid State Circuits Conference, Leuven, 2017, pp. 251-254.

[3] Y. Zhang and P. Reynaert, "A high-efficiency linear power amplifier for 28GHz mobile communications in 40nm CMOS," 2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), Honolulu, HI, 2017, pp. 33-36.

[4] P. Indirayanti and P. Reynaert, "A 32 GHz 20 dBm-PSAT transformer-based Doherty power amplifier for multi-Gb/s 5G applications in 28 nm bulk CMOS," 2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), Honolulu, HI, 2017, pp. 45-48.

[5] S. Shakib, H. C. Park, J. Dunworth, V. Aparin and K. Entesari, "A 28GHz efficient linear power amplifier for 5G phased arrays in 28nm bulk CMOS," 2016 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, 2016, pp. 352-353.

[6] B. Park, Daechul Jeong, J. Kim, Y. Cho, Kyunghoon Moon and B. Kim, "Highly linear CMOS power amplifier for mm-wave applications," 2016 IEEE MTT-S International Microwave Symposium (IMS), San Francisco, CA, 2016, pp. 1-3.

Paper II

Paper II

A 27 GHz Adaptive Bias Variable Gain Power Amplifier and T/R Switch in 22nm FD-SOI CMOS for 5G Antenna Arrays

C. Elgaard, S. Andersson, P. Caputa, E. Westesson, and H. Sjöland,, "A 27 GHz Adaptive Bias Variable Gain Power Amplifier and T/R Switch in 22nm FD-SOI CMOS for 5G Antenna Arrays," © 2019 IEEE, reprinted from *Proc. of IEEE RFIC*, Boston, USA, June 2019.

A 27 GHz Adaptive Bias Variable Gain Power Amplifier and T/R Switch in 22nm FD-SOI CMOS for 5G Antenna Arrays

Christian Elgaard[#*1], Stefan Andersson[#], Peter Caputa[#], Eric Westesson[#], Henrik Sjöland[#*2]
[#]Ericsson, Lund, Sweden
[*]Lund University, Lund, Sweden
[1]christian.elgaard@ericsson.com, [2]henrik.sjoland@ericsson.com

***Abstract*— A 27 GHz fully integrated, variable gain, two stage Power Amplifier (PA) and a Transmit/Receive (T/R) switch targeting 5G antenna array systems are presented. The PA uses adaptive bias, tracking the input signal amplitude, which improves saturated output power ($\mathrm{P_{sat}}$) with 1.4 dB and 1 dB output compression ($\mathrm{OP_{1dB}}$) by 3 dB. For a supply voltage of 1.2 V, the PA reaches a $\mathrm{P_{sat}}$ of 17.4 dBm and an $\mathrm{OP_{1dB}}$ of 16.5 dBm, with a power added efficiency of 19.5 % and 17.3 %, respectively. The power gain can be controlled with 5-bits from 5.2 to 34 dB. The T/R-switch has an insertion loss of 1.63/1.46 dB in TX/RX mode, and for reliability reasons all switch devices are on in TX-mode. The complete PA and T/R-switch only occupies 0.146 $\mathrm{mm^2}$ in a 22 nm FD-SOI CMOS technology.**

***Keywords*— power amplifier, adaptive bias, T/R-switch, millimeter wave integrated circuits.**

I. Introduction

Fifth generation mobile communication (5G) will support millimeter-wave frequencies and beamforming, using large antenna array systems (AAS) with tens or even hundreds of antennas, where each antenna supports two polarizations [1]. Size and separation of the antennas in an AAS is typically about half a wave length, which at 30 GHz amounts to 5 mm. To generate transmit signals for such densely spaced antennas, full integration of the PA and the transmit/receive (T/R) switch into the CMOS transceiver chip is necessary. Effectively, many transceivers will be placed on one chip. In addition to the analog part, each chip needs to handle massive digital signal processing associated with beam-forming. To limit the size of the digital part, a short channel length advanced CMOS technology, must be used. A challenge with such technologies is the sensitivity to high voltage levels, which makes the implementation of the PA and T/R-switch particularly difficult. The T/R-switch must handle the voltage swing from the PA with low insertion loss, while protecting the low noise amplifier (LNA) from the strong PA output signal. Many published works focus either on the PA or the T/R-switch, thereby not addressing the issues arising when combining them.

Digital predistortion (DPD) is often applied to the PA input signal, which improves adjacent channel leakage ratio and error vector magnitude at the PA output. A DPD system can effectively combat amplitude to phase (AM-PM) variation introduced by the PA, and within limits also reduce the amplitude to amplitude (AM-AM) error. One limitation of AM-AM reduction for a PA using a DPD, however, is the increase of peak to average ratio of the input signal. This requires an increased dynamic range in the TX digital to analog converter (DAC), and an increased maximum signal level in the entire analog part of the transmitter, leading to more stringent requirements on linearity and output power capability.

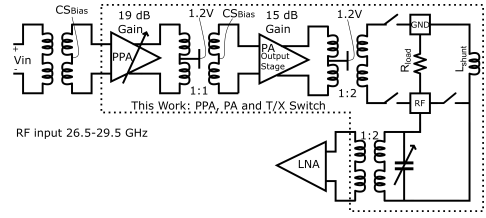

Fig. 1. The proposed power amplifier and T/R switch architecture, part of a fully integrated transceiver.

To the authors knowledge, this paper presents the design and measurement of the first published fully integrated PA and T/R switch in 22 nm FD-SOI CMOS, see Fig. 1, targeting the 3GPP band n257, i.e. 26.5 - 29.5 GHz [2]. The PA, consisting of a variable gain pre-PA (PPA) and a PA output stage, utilizes analog linearization by dynamically adapting the PA bias to track the input envelope. Turning on the adaptive bias increases saturated output power ($\mathrm{P_{sat}}$) with 1.4 dB and output referred 1 dB compression ($\mathrm{OP_{1dB}}$) by 3 dB. Prior to the T/R-switch, the PA reaches a $\mathrm{P_{sat}}$ of 17.4 dBm and an $\mathrm{OP_{1dB}}$ of 16.5 dBm when the adaptive bias is active. The switch has an insertion loss of 1.63/1.46 dB in TX/RX-mode The results show that adaptive bias can reduce the dynamic range of a DPD required to combat AM-AM distortion, effectively reducing the required dynamic range for the TX DAC, TX-BB, and mixer. It is also shown that a low loss T/R-switch can be fully integrated in 22 nm FD-SOI CMOS in a reliable way.

II. Circuit Design

A. PPA

The PPA depicted in Fig. 2 is a common source (CS) common gate (CG) stage (i.e. a cascode stage) where both the CS and the CG transistors are implemented as 32 unit cells to control the gain. The gates of the CS transistors in all unit cells on each differential side, are all connected to the same bias voltage and input signal. The unit cells are switched on/off by controlling the CG transistor gate voltage bias, which will control the effective width of both the CG and

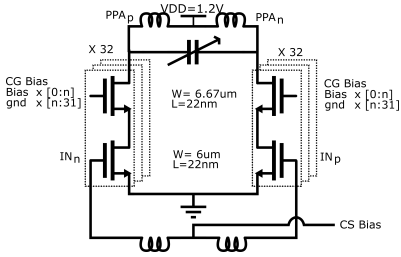

Fig. 2. The PPA.

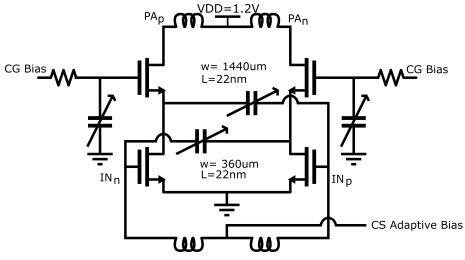

Fig. 3. The PA output stage.

CS transistors. This results in a reduced bias current when the PPA gain is reduced. The input impedance is dominated by the CS transistor $\mathrm{C_{gs}}$ capacitance, which is resonated by the inductive part of the preceding transformer. A 5-bit digitally controlled capacitor is used to compensate for capacitance variations when changing the PPA gain. At resonance the input impedance becomes real and is simulated to $260\,\Omega$ for a Q-value of 20. The high input impedance is essential to reduce loading of the preceding mixer and to avoid the need for an additional gain stage between the mixer and the PPA. At the output transformer, between the PPA and PA output stage, a 3-bit digitally controlled capacitor is used to control the resonance frequency.

B. PA Output Stage

The PA output stage, shown in Fig. 3, uses a similar CSCG stage as the PPA, but with 2-bit programmable neutralization capacitors and digital control of CG gate voltage swing. Assuming high resistance from the bias source to the CG gate, the voltage swing at the CG gates is set by the output voltage swing and the capacitive voltage division between $\mathrm{C_{dg}}$ and the gate capacitor to ground. The swing of the CG gate voltage in turn sets that of the CS drain. To divide the output voltage swing evenly between the CG and CS transistors, a 2-bit digitally controllable capacitor between the CG gate and signal ground is added. The real part of the input impedance, determined mainly by the Q-value of the input transformer of the PA output stage, is simulated to $170\,\Omega$ at resonance.

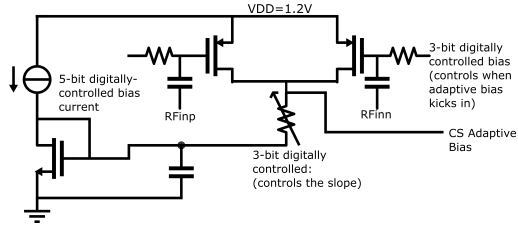

Fig. 4. Adaptive bias circuit for PA output stage CS transistor.

C. Adaptive Bias

To boost $\mathrm{OP_{1dB}}$ and $\mathrm{P_{sat}}$, an adaptive bias circuit, see Fig. 4, is used to increase the CS transistor gate bias voltage in the PA output stage as the input signal amplitude increases. The reference current to the adaptive bias circuit can be controlled with a 5-bit digital word. For the adaptive bias circuit to function correctly, it is important that it starts to increase the voltage bias at the input level where the PA output stage starts to compress, and that it also increases the bias voltage with the correct slope to compensate the gain compression at higher input signal levels. To a first approximation the input level starting point can be controlled by the 3-bit bias voltage level, effectively setting at which input amplitude the rectifying PMOS pair will start conduct current. The slope of the increase in gate bias voltage is set with the 3-bit tunable resistor, transforming the DC current from the PMOS pair into a DC voltage. In total 64 different adaptive bias settings can be selected (8 starting points and 8 slopes). Figure 5 shows measured results for output power and CS voltage bias as a function of the input signal level for all 64 combinations. For test reasons the CS bias voltage is also routed, so that it can be measured outside the chip, through an auxiliary path. The output signal of the adaptive bias circuit is designed to track the envelope of the input signal and the bandwidth is limited by the output impedance of the adaptive bias circuit, which is below $1\,\mathrm{k}\Omega$ and the input capacitance of the PA output stage CS transistors, which is approximately $2 \times \mathrm{C_{gs}} = 340\,\mathrm{fF}$. The envelope of an OFDM-modulated signal has the features with the highest frequency components when the amplitude is low, but as the adaptive bias circuit does not change the bias level at low amplitudes, this is not a problem, and it is enough if medium to high signal levels can be tracked.

D. T/R Switch

Figure 6 shows the layout of the T/R switch, implemented in the thick upper copper and aluminum metal layers, including RF and supply pads, the TX 1:2 balun, and the RX 2:1 balun. The upper three pads show GND-Signal-GND connecting to antenna, and the lower two pads are VDD supplying the PA through the center tap of the balun. The architecture of the T/R switch is chosen such that all the three active switches, as shown in Fig. 1, are conducting when in TX-mode. This ensures low voltage across the switch transistors, thereby avoiding the risk of damaging them by

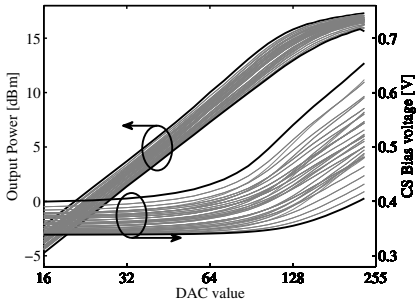

Fig. 5. Measured output power at 27 GHz and CS bias voltage from adaptive bias circuit vs. DAC signal level for all 8x8 adaptive bias settings. The two thick curves are the extremes for the adaptive bias when enabled.

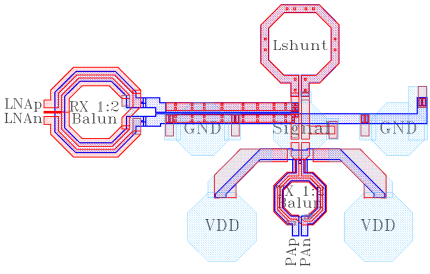

Fig. 6. The layout of the T/R switch and the TX/RX baluns, omitting active parts, vias, and lower metal layers.

high voltage stress in TX mode, without the need for transistor stacking, which has the drawback of increased switch on-resistance R_{on}. In addition, non-conducting transistor switches in TX mode would risk introducing additional non-linearities, as the switches could start to conduct current during voltage peaks. In TX mode the inductor L_{shunt} is used to resonate the parasitic capacitance of the RF pad whereas in RX mode it is used to conduct the received signal to the LNA together with the primary side of the RX 2:1 balun. Capacitive tuning is applied at the input of the LNA for final matching. In RX mode, parasitic capacitance of the switch between the RF pad and the TX balun will leak signal power into TX, which increases the loss in RX mode. To mitigate this, an extra switch is added between the TX balun and the GND pad, which places the parasitic capacitances of the two switches in series, effectively reducing the signal leakage. Figure 7 shows simulated losses in the T/R switch. At 27 GHz, TX and RX insertion loss (IL) are simulated to 1.63 and 1.46 dB, respectively. In addition, for the two 1:2 baluns TX and RX simulated losses are 1.22 and 1.41 dB respectively. Isolation from TX to RX is simulated to 22.1 dB, which at P_{sat} gives a 130 mV_{rms} input signal to the LNA, which is acceptable since the circuit is targeted to operate in time division duplex.

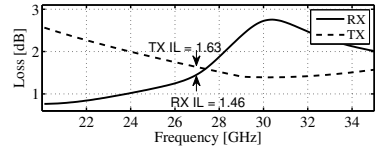

Fig. 7. Simulated T/R-switch losses in TX/RX mode.

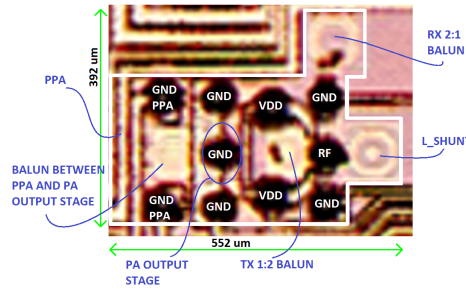

Fig. 8. Die photo showing the the PA and T/R-switch. The area inside the white polygon is 0.146 mm^2.

III. Measurement Results

The chip was fabricated in a 22 nm FD-SOI CMOS process and flip-chip mounted in a package that was mounted on a PCB. The PA, including decoupling capacitors, bias generation, and pads for both supply and RF occupies a die area of 0.129 mm^2. The complete PA and T/R-switch occupies a die area of 0.146 mm^2, see Fig. 8. The input signal to the PA is generated on-chip by an integrated DAC, fed through an on chip analog base band, and up-converted to carrier frequency by a mixer directly preceding the PPA. The PA was measured, using an R&S NRP2 Power Meter and an R&S FSM Signal & Spectrum Analyzer, with both adaptive and constant CS gate bias voltage, configured such that for low input amplitudes the CS-bias voltage was the same for both cases. Measurement results were obtained with identical settings for figures 9 - 11. Figure 9 shows the output power and the CS gate bias voltage as a function of TX-DAC value. The increased CS bias voltage at high input amplitudes when using adaptive bias increases P_{sat} by 1.4 dB. Measured drain efficiency for the PA output stage and normalized power gain for the complete PA is shown in Fig. 10. The gain is normalized to a simulated small signal power gain of 34 dB. When using adaptive bias, drain efficiency peaks at 20.1 %. Furthermore, OP_{1dB} increases by 3 dB from 13.5 to 16.5 dBm, when using adaptive bias. A two tone test with 50 MHz separation around 27.2 GHz, was carried out to assess the linearity of the PA. Figure 11 shows the total power in the two IM3 tones, both with constant and adaptive bias, vs. output power for the fundamental tones. As can be seen, the power in the IM3 products when using adaptive bias is lower for all output power levels reachable when using constant bias. A 2.5 dB reduction in IM3 power can be seen at the highest constant bias two tone output power,

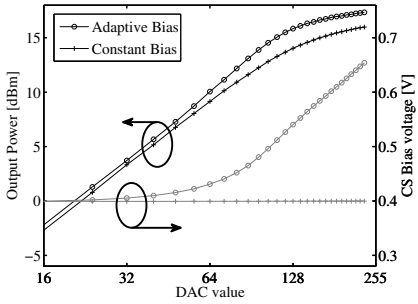

Fig. 9. Output power at 27 GHz vs. input signal to the TX-DAC.

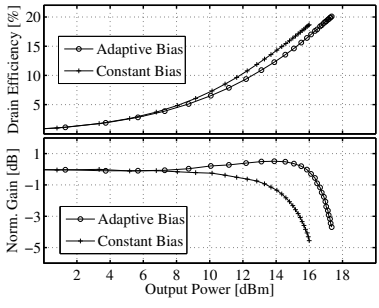

Fig. 10. Upper plot: Drain efficiency vs. output power for adaptive bias and constant bias at 27 GHz. Lower plot: Normalized gain vs. output power for adaptive bias and constant bias at 27 GHz.

i.e. 11.3 dBm. In addition, the adaptive bias case offers >2 dB extended fundamental output power. The 1 dB bandwidth was measured to more than 1.2 GHz, frequency ranging from 26.3 GHz to 27.5 GHz. Table 1 summarises the performance and compares with state-of-the-art PAs in 28 nm CMOS, since no published PAs in 22 nm CMOS process were found for similar frequencies. For fair comparison all PA performance are prior to the losses in the T/R-switch and the proposed PA

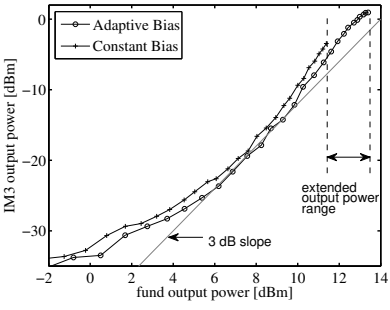

Fig. 11. IM3 tones output power vs. the fundamental tones output power for constant and adaptive bias. A 3 dB slope is added for reference.

Table 1. Performance Summary and Comparison

Parameter	**This Work**	APMC'18 [3]	RFIC'18 [4]	RFIC'17 [5]	RFIC'17 [6]
Technology [nm]	**22FDSOI CMOS**	28FDSOI CMOS	28FDSOI CMOS	28Bulk CMOS	28Bulk CMOS
Frequency [GHz]	**27**	26	31	27	32
Pwr.Combining	**1**	2	1	2	2
Gain Control [dB]	**28.8**	19	10	None	None
No. of Stages	**2**	2	2	2	2
Vdd [V]	**1.2**	1.5	0.7/1.98	1	1
P_{sat} [dBm]	**17.4**	20.6	17.9	18.1	19.8
P_{1dB} [dBm]	**16.5**	18.8	11.6	16.8	16
PAE_{max} [%]	**19.5**$^{(1)}$	22.6$^{(1)}$	25.5	41.5	21
PAE_{1dB} [%]	**17.3**$^{(1)}$	16.6$^{(1)}$	10	37.6	12.8
Power Gain [dB]	**34**$^{(2)}$	33.1$^{(2)}$	32.6	20.5	22
Area [mm^2]	**0.129**	0.144	0.508	0.361	0.59

$^{(1)}$Calculated from measured drain efficiency and simulated power gain. Not including PPA power consumption.
$^{(2)}$Simulated value since input signal is generated internally.

achieves the highest power gain, largest gain control range, and smallest die area.

IV. Conclusions

This paper introduces a variable gain PA at 27 GHz, utilizing adaptive bias that tracks the envelope of the modulated signal, and a T/R-switch. The adaptive bias increases the $\mathrm{OP_{1dB}}$ and $\mathrm{P_{sat}}$ from 13.5 to 16.5 dBm and from 16 to 17.4 dBm, respectively. In addition, for a two tone test with 50 MHz frequency difference, the adaptive bias reduces the power of the IM3-tones by 2.5 dB. The T/R-switch achieves a simulated TX/RX insertion loss of 1.63/1.46 dB. For reliability reasons, the transistors in the T/R-switch are all in highly conductive mode when transmitting. Both the PA and the T/R-switch are implemented in a 22 nm FD-SOI CMOS process and integrated on the same chip, in a transceiver targeted for 5G AAS.

References

[1] B. Sadhu et al., "A 28-GHz 32-Element TRX Phased-Array IC With Concurrent Dual-Polarized Operation and Orthogonal Phase and Gain Control for 5G Communications," in IEEE Journal of Solid-State Circuits, vol. 52, no. 12, pp. 3373-3391, Dec. 2017. doi: 10.1109/JSSC.2017.2766211

[2] 3GPP TS 38.104 V15.0.0 (2017-12), Technical Specification, 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; Base Stations (BS) radio transmission and reception (Release 15). http://www.3gpp.org/ftp/Specs/archive/38_series/38.104/

[3] C. Elgaard, A. Axholt, E. Westesson and H. Sjöland, "A 26GHz 22.2DBM Variable Gain Power Amplifier in 28NM FD-SOI CMOS for 5G Antenna Arrays," 2018 Asia-Pacific Microwave Conference (APMC), Kyoto, 2018, pp. 965-967.

[4] F. Torres, M. De Matos, A. Cathelin and E. Kerhervé, "A 31 GHz 2-Stage Reconfigurable Balanced Power Amplifier with 32.6dB Power Gain, 25.5% PAEmaxand 17.9dBm Psatin 28nm FD-SOI CMOS," 2018 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), Philadelphia, PA, 2018, pp. 236-239.

[5] Y. Zhang and P. Reynaert, "A high-efficiency linear power amplifier for 28GHz mobile communications in 40nm CMOS," 2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), Honolulu, HI, 2017, pp. 33-36.

[6] P. Indirayanti and P. Reynaert, "A 32 GHz 20 dBm-PSAT transformer-based Doherty power amplifier for multi-Gb/s 5G applications in 28 nm bulk CMOS," 2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), Honolulu, HI, 2017, pp. 45-48.

Paper III

Paper III

Efficient Wideband mmW Transceiver Front-End for 5G Base Stations in 22 nm FD-SOI CMOS

C. Elgaard, M. Özen, E. Westesson, A. Mahmoud, F. Torres, S. Bint Reyaz
T. Forsberg, R. Akbar, H. Hagberg, and H. Sjöland, "Efficient Wideband mmW Transceiver Front-End for 5G Base Stations in 22 nm FD-SOI CMOS," *IEEE Journal of Solid-State Circuits (JSSC)*,pp. 1-16, June 2023.

Efficient Wideband mmW Transceiver Front-End for 5G Base Stations in 22 nm FD-SOI CMOS

Christian Elgaard, Mustafa Özen, Eric Westesson, Ahmed Mahmoud, Florent Torres, Shakila Bint Reyaz
Therese Forsberg, Rehman Akbar, *Student Member, IEEE*, Hans Hagberg, Henrik Sjöland, *Senior Member, IEEE*

***Abstract*—This paper presents a fully integrated millimeter-wave transceiver front-end covering 24.25 to 29.5 GHz. It features a wide band Doherty power amplifier utilizing adaptive bias, and a transmit/receive switch (TRX-switch) that has embedded low noise amplifier to antenna matching. The phase shift of 90 degrees to the Doherty auxiliary amplifier is achieved using a separate IQ-mixer with rearranged phases in the auxiliary path, ensuring a wideband 90-degree phase shift, and avoiding 3 dB loss from RF input power splitting. Special emphasis is on the analysis of adaptive bias, the Doherty output combiner network, the decoupling capacitors, and the TRX-switch. Including TRX-switch losses of 1.1 dB in transmit mode, the transmitter reaches a saturated output power of 18.3 dBm with a 1 dB output compression point of 15.9 dBm. Stimulated with a 400 MHz 16-QAM OFDM IQ-signal at base-band, without digital IQ-compensation and predistortion, the transmitter delivers a 26.5 GHz modulated signal with an output power (P_{out}) of 12.8 dBm, and an error vector magnitude (EVM) of -20.2 dB. The complete transmitter, including quadrature local oscillator drivers, then achieves a power added efficiency (PAE) of 5.8%. For a 1600 MHz wide 64-QAM OFDM signal, P_{out} is 9.0 dBm, with EVM=-23.3 dB and a complete transmitter PAE of 3.2%. In receive mode including TRX-switch, at 27.25 GHz, the noise figure is below 4 dB with a gain of 23 dB and a third order input referred intercept point of -9 dBm. The active part of the die, manufactured in 22 nm FD-SOI CMOS, occupies 2.3 mm^2.**

***Index Terms*—adaptive bias, decoupling, Doherty power amplifier, image rejection, low noise amplifier (LNA), LO-leakage, millimeter-wave (mmW), mixer, transceiver (TRX), TRX-switch.**

I. Introduction

5G mmW-band base stations (BS) typically use large antenna array systems (AAS). This increases the number of deployed transceivers, which drives the BS front-end requirements, for both cost and radio performance, to become more similar to those of user equipment (UE), than to previous generations of cellular BS. Transmit (TX) output power requirements can be significantly reduced since AAS beamforming gain increases the effective isotropic radiated power (EIRP), whereas efficiency requirements are mainly driven by thermal aspects, however, recently an increased focus on cost savings by reduced power consumption is observed [1]. On the receive (RX) side, the uplink (UL) effective isotropic sensitivity (EIS) is boosted by AAS beamforming gain, similar to the EIRP in downlink (DL). But, superior BS to UE total radiated power (TRP) results in an asymmetric DL/UL performance, which is shown in the link budget presented in Section II and in [2]. To some extent this is also desired due to the DL heavy data traffic in today's cellular networks. Nevertheless, to enhance cell edge coverage and avoid a too asymmetric DL/UL capacity in the network, the RX noise figure (NF) remains a key target to improve in BS AAS, effectively favouring RX NF over TX output power. This paper presents a fully integrated 22 nm FD-SOI CMOS transceiver front-end for BS AAS covering 3GPP 5G NR bands n257, n258, and n261 (24.25-29.5 GHz) [3].

In recent years, many 5G mmW transceivers have been published [2], [4]–[18], all using class AB biased PAs to meet the stringent linearity requirements, of the high peak to average ratio (PAR) 5G NR OFDM signal. Class AB PAs demonstrate good linearity at the expense of poor efficiency when transmitting high PAR signals [19]. Most of the published works also leave out frequency conversion [4]–[14], which avoids fundamental problems arising in the base-band (BB) and mixer, such as LO-leakage and finite image rejection (IR). None of the published complete transceiver front-ends use the increasingly popular technique of adaptive bias of the PA, however, standalone PAs have been published that investigate the concept [20]–[26]. Unfortunately, improvements are, at best, demonstrated using continuous wave (CW) tones, and in [25] which demonstrate excellent wideband modulated performance, no assessment is carried out regarding the gains brought by the adaptive bias, neither using simulations nor measurements. For successful use of adaptive bias for high bandwidth (BW) signals a rigorous analysis and design is required. Furthermore, possibility to control and adjust the adaptive bias parameters is needed to accurately linearise the amplifier, which unfortunately is missing in all the published adaptive bias enhanced amplifiers, with the exception of [20]. Multiple successful mmW Doherty amplifier designs also exist [23]–[26], but none that is integrated in a transceiver architecture. The traditional Doherty amplifier has fundamental BW limitations, both from input phase shifting and output power combination network. The TRX-switch often favours TX over RX since loss in TX mode is assumed to cost more DC power to retrieve [15], however, from a full system perspective, in a cellular BS AAS, loss in RX mode is typically more important.

To summarise, this motivates the research of a 5G mmW transceiver front-end with a Doherty PA, linearised by an adaptive bias capable of tracking high BW modulated signals and with high level of controllability. The presented circuit also includes frequency translation using two IQ-mixers, with built in LO-leakage calibration, which can produce a close to ideal 90-degrees phase shift for all covered frequencies at the auxiliary Doherty input, thereby eliminating one of the BW bottlenecks. The circuit also features a TRX-switch with low RX insertion loss with integrated LNA input impedance

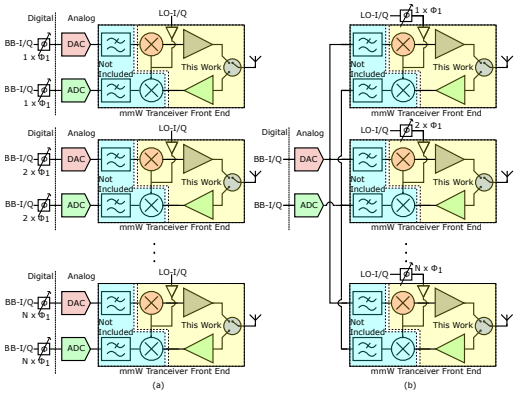

Fig. 1. AAS for the mmW transceiver front-end using (a) Digital beamforming and (b) Analog LO beamforming.

matching.

This paper is organized as follows. Section II places the front end circuit in a system context, Section III motivates the chosen transceiver front-end architecture, and Section IV provides key design details about the different sub-blocks. Measurement results and comparison with state-of-the-art for the complete transceiver front-end in both RX and TX mode are presented in Section V. Finally, the paper is concluded in Section VI.

II. System Context

There are different architectures for beamforming AAS, and the main categorization is digital and analog beamforming [27]. Digital beamforming is more flexible, for instance being able to form multiple beams and to approximate true time delays (TTD). Digital architectures are typically associated with higher cost as each antenna element requires a complete signal chain from antenna to digital. This means that it requires separate analog to digital (ADC) and digital to analog converter (DAC) for each antenna element in the array. Another disadvantage with digital beamforming is the digital signal processing part that will be large and power consuming for wideband signals and large array sizes. Analog beamforming is less flexible than digital since it can only form a singe beam at the time, also it typically lacks the possibility to approximate TTD when applying the phase shift, but on the other hand it requires less hardware and signal processing. It can also provide better immunity to interfering signals outside the beam direction. There are different analog beamforming architectures, which can be separated by where the phase shift is performed. The phase shift can be performed on the RF signal, so called RF beamforming, on the local oscillator (LO) signal, so called LO beamforming, and on the BB signal, so called analog BB beamforming. All have in common that the signal is phase shifted rather than delayed. The system context of the proposed mmW transceiver front-end, for a 1-dimensional beamforming AAS example, is shown using digital beamforming in Fig. 1 (a) and using analog LO beamforming in Fig. 1 (b). In large arrays, combinations of analog and digital beamforming, so called hybrid beamforming, can be used to find a good balance between flexibility and complexity. In general, the larger the array and the wider the signal bandwidth, the more attractive analog beamforming becomes. This means that at high operating frequencies, where the signal bandwidths are higher and larger array sizes are required, analog beamforming is preferred. There is a trend in 5G, and presumably also for 6G, towards high operating frequencies and bandwidths, calling for analog beamforming. Both the digital beamforming and the analog LO beamforming of Fig. 1 (a) and (b) requires careful considerations for effectively distributing the LO signal used for up/down-conversion. The options typically becomes to use a centralized LO generation followed by a large and relatively power hungry distribution buffer tree, or to use a local LO generation at each transceiver, as proposed in [28], but perhaps the most attractive option, especially for large AAS, is to use a hybrid version of centralized and local LO generation.

AAS base stations for 5G mmW can be configured in different power classes, depending on the intended range. Table I presents an indicative link budget for UL and DL for three potential BS classes, local area (LA), medium range (MR), and wide area (WA).

TABLE I
Link Budget Per Polarization

BS Class	**WA**		**MR**		**LA**	
Eq. Free Space Distance	1000 m		500 m		125 m	
Antennas BS, UE	32 x 16, 4		16 x 16, 4		8 x 8, 4	
Link	DL	UL	DL	UL	DL	UL
BW [MHz]	800	100	800	200	800	800
QAM-Mod	64	16	64	16	64	16
Raw Data Rate [Gbit/s]	4.8	0.4	4.8	0.8	4.8	3.2
TX Element Pout [dBm]	10	13	10	13	10	13
TRP [dBm]	37	19	34	19	28	19
TX Antenna Gain [dB]	3	3	3	3	3	3
EIRP [dBm]	**67**	**28**	**61**	**28**	**49**	**28**
Path Loss [dB]	121	121	115	115	103	103
RX Antenna Gain [dB]	3	3	3	3	3	3
RX Beamforming Gain [dB]	6	27	6	24	6	18
Received Power [dBm]	**-45**	**-63**	**-45**	**-60**	**-45**	**-54**
Required SNR [dB]	27	21	27	21	27	21
Margin [dB]	5	5	5	5	5	5
RX noise floor [dBm]	**-77**	**-89**	**-77**	**-86**	**-77**	**-80**
RX Noise Figure [dB]	8	5	8	5	8	5
BW Noise Increase [dB]	89	80	89	83	89	89
kT [dBm/Hz]	-174	-174	-174	-174	-174	-174

The phased array beamforming gain increases linearly with array size, boosting DL and UL performance equally. However, the larger the AAS becomes the more unsymmetrical the DL/UL performance becomes. The reason is that with each added antenna element, a PA is also added, and hence the BS increases its TRP linearly when adding more antenna elements. As a result, an LA BS can have relatively symmetrical UL/DL performance, whereas the quality of service for a WA BS becomes highly limited by the UL.

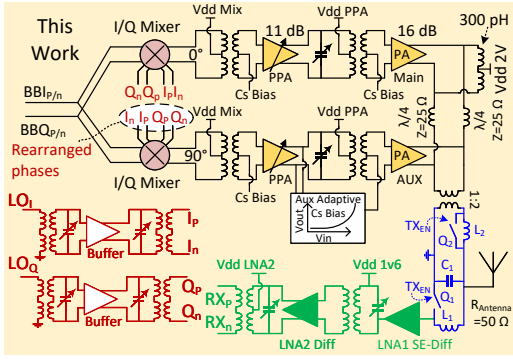

Fig. 2. The transceiver architecture. Black: Transmitter, Green: Receiver, Blue: TRX-switch, Red: LO distribution.

III. Transceiver Architecture

A. Transmitter

Figure 2 shows the architecture of the entire transceiver front-end. For high efficiency with modulated signals a Doherty power amplifier is used, where the required 90 degrees phase shift between the input of the main and auxiliary amplifiers is achieved by using two IQ-mixers, with rearranged local oscillator (LO) phases to the auxiliary path, as previously demonstrated in [29]. This avoids the fundamental 3 dB loss of the input signal, that is associated with a single input Dohery amplifier using RF power splitting and a $\lambda/4$-transmission line to generate the 90 degree phase difference. Disadvantages of the 3 dB loss include more stringent requirements on the preceding PA driver, such as 3 dB more gain, 3 dB higher output signal, and 3 dB higher output compression point, which will lead to an increased power consumption. On the other hand, one may argue that having two IQ-mixers, two Pre Power Amplifiers (PPAs), and twice the LO-Drivers will increase the power consumption. Theoretically though, the current consumption will not increase as each of the two PPAs or mixers only have to drive half the load. This is possible if the Q value of inductances (Q_L) and capacitances (Q_C) do not change between the two cases. The impedance at parallel resonance is $R_p = Q/(\omega_0 C)$, where $Q = Q_L//Q_C$. If Q is unchanged, the impedance becomes inversely proportional to the capacitance, so that when the circuit is divided into two halves, each with half the capacitance, the impedance of each halve is doubled. However, even more importantly than avoiding the 3 dB loss, the proposed dual input Doherty amplifier also avoids the fundamental problem to produce the desired 90 degrees input signal phase shift over a wide BW. Theoretically, the BW for Doherty operation is now only limited by the output combination network. Any LO IQ imbalance is expected to have two consequences, first it will translate to an incorrect phase difference of the output signals of the main and auxiliary amplifiers. This problem is of minor concern as the required 90 degree phase shift anyhow only occurs at a single frequency, and a small phase shift does not significantly affect the Doherty operation. Second, due to the individual LO routing to the main and auxiliary amplifier paths and the rearranged phases to the two mixers, there can be different LO IQ imbalance in the main and auxiliary paths. Since their contributions to the combined output signal varies with the amplitude, there can be an amplitude dependent image signal. Two other popular techniques to create the 90 degree phase shift of the input signal are multi-stage poly phase filter (PPF) and quadrature hybrid coupler, as in [29]. Both multi-stage PPF and quadrature hybrid couplers has the fundamental 3 dB problem as described above and in addition both have losses, which in the case of multi-stage PPF can be quite severe [30]. Hybrid couplers can consume significant chip area, especially for the lower range of mmW bands. In addition the multi-stage PPF has problems covering a wide relative bandwidth of 20 % such as required to cover the 3GPP band n257, n258, and n261 [30]. Since the LO signal is used not only for a single up-conversion mixer in the transmitter, but also creating the 90 degree phase shift, it is worth mentioning that there is no additional requirement on the LO signal phase noise for the proposed architecture. To support a 120 kHz subcarrier spacing, with a PLL bandwidth of 2 MHz, it would thus be sufficient with an inband phase noise of -105 dBc/Hz for an LO phase noise EVM contribution below -35 dB, as reported in [28].

Adaptive bias is used in the auxiliary amplifier of the Doherty power amplifier, with the overall objective to linearise the output amplitude of the power amplifier, similar to what is demonstrated in [20]–[22]. For the Doherty amplifier, this becomes even more important, as the turn-on of the auxiliary amplifier is associated with amplitude distortion. Deployment of adaptive bias in a mmW Doherty amplifier has been demonstrated in [23]–[26]. The adaptive bias has two tasks, controlling the turn-on of the class C biased auxiliary amplifier, and compensating for compression at large output signal levels. However, not so frequently discussed, the adaptive bias must track the envelope of the modulated signal, which for high modulation bandwidths may contain frequency components at multiple GHz. Unsuccessful tracking of the envelope will result in reduced efficiency and increased distortion. Inadequate adaptive bias bandwidth can manifest itself as distortion memory effects as well as reduction of the suppression of the amplitude distortion.

To cover a wider frequency range, while still enabling the use of high impedance levels in the resonance tanks to reduce power consumption, digitally controllable capacitors are used in the mixers, in the PPAs driving the PA inputs, and in the LO buffer inputs and outputs.

The BB filter required to suppress repetitive spectrum from the digital to analog converter (DAC) providing the IQ-input signal, was left out of the design to enable possibility to freely evaluate performance of transmission of various bandwidths.

As the transceiver uses a single-ended (SE) antenna and the PA is differential, an on-chip balun is used for differential to SE conversion between the PA and TRX-switch. Naturally, losses in the balun reduces both the output power and the efficiency.

B. Receiver and Switch

A novel TRX-switch, where the LNA matching i.e. the series gate inductance, is incorporated into the switch is proposed. In RX-mode the switch provides lower loss and significantly less area than a quarter wavelength transmission-line based TX-ON switch. The first low noise amplifier (LNA1) is single-ended (SE), to avoid deteriorating RX noise figure with the loss of a balun. Furthermore, since the switch is connected to the secondary side of the TX-balun, it offers protection from electro static discharge (ESD) for the receiver, via the grounding of the TX-balun secondary side. An inductively source degenerated design was chosen, optimized for high gain, exceeding 15 dB according to simulations. Due to more than 20 dB of combined gain of LNA1, balun, and LNA2, the NF of the succeeding blocks in the receiver, the mixer, analog low-pass filter, and ADC, will have minor impact on the overall RX NF. Similar to the TX, digitally controllable capacitors are used to tune the resonance frequency at the output of LNA1 and LNA2.

IV. Circuit Design

A. LO Distribution and Mixer

Two single ended 90 degree phase shifted (quadrature) LO signals are generated off-chip. Integrated baluns are then used to produce on-chip differential versions of the LO signals, which in turn drive two differential buffers, one I and one Q buffer, see Fig. 2. Besides from producing the differential versions of the LO signals, the integrated baluns also avoid the need for any additional ESD protections. Each buffer is implemented as a differential pair with a tuned transformer as output load. The output of the I- and Q-buffers provide the four LO phases needed for the complex up-conversion mixers. To reduce possible mismatch in the buffer core caused by process gradients on the wafer, a common centroid layout is used, where the transistors of the differential pairs are divided into several subparts, placed in an interleaved fashion as proposed in [31]. One challenge is to distribute the four phases to the two mixers, while preserving the quadrature accuracy, when routing the long and closely spaced wires of the LO distribution network. Phase and amplitude imbalance causes reduced image rejection in the complex mixers, which will ultimately compromise error vector magnitude (EVM) when transmitting OFDM signals. Figure 3 shows the passive LO-distribution network, simulated using the ADS Momentum electromagnetic simulator to verify its performance, shown without metal ground shield beneath the wires for picture clarity.

Active double balanced mixers, also known as Gilbert mixers, were chosen for the frequency upconversion. Advantages include current mode output for easy combination of I and Q mixer output signals, limited LO-amplitude requirements, and signal gain. The output currents are fed into a tuned transformer at each IQ-mixer, which converts them into a voltage signal. To obtain good matching the two IQ-mixers in the main and auxiliary paths are identical, except for the rearranged phases of the local oscillator (LO) signal to the auxiliary path. LO-leakage to the mixer outputs is due to

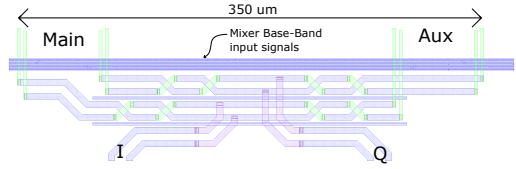

Fig. 3. Layout of the LO-distribution network. The multiple horizontal dark blue wires is the distribution of the base-band input signals to the mixers.

direct leakage and due to DC-offsets in the Gilbert mixer cores. Suppression of the LO-leakage from each Gilbert cell is achieved by adding two 6-bit digitally controllable DC-current sources at the output of the baseband input transistors in each double-balanced mixer, see Fig. 4, which for the two IQ-mixers gives in total eight individual controllable current sources. The ranges for the current sources are designed to cover any expected DC-offset due to process variations. The LO-leakage calibration is similar to the procedure demonstrated in [32], and the simulation results can be seen in Fig. 4. When performing the LO-leakage calibration in an AAS, the LO-leakage would have to be characterized during a dedicated calibration state. The output signal from one transmitter can then be measured using a transmit observation receiver or by activating another receiver in the AAS and measure the TX signal through the antenna coupling. The DC leakage currents can then be appropriately adjusted. This procedure would have to be repeated when conditions such as temperature or operating frequency change.

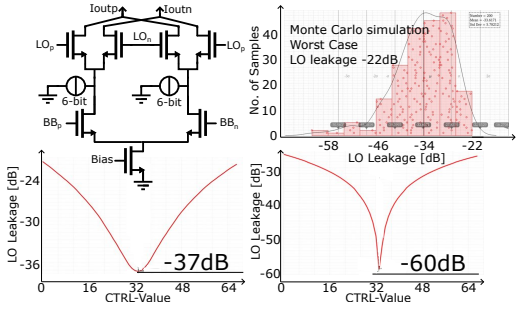

Fig. 4. Schematic of the mixer with LO-leakage calibration and raw LO-leakage performance due to mismatch and simulated performance in first and second sweep of current control settings.

B. Pre Power Amplifier and Power Amplifier

The PPA, see Fig. 5, employs a common source (CS) common gate (CG) architecture with cross-coupled neutralization capacitors and with current steering to achieve 16 controllable coarse gain steps, that are measured and shown in Fig. 21. The use of tunable body-bias voltage Vbb on the common source stage provides fine gain steps. Vbb can be controlled in eight levels ranging from 0 V to 1 V, and gives a simulated PPA

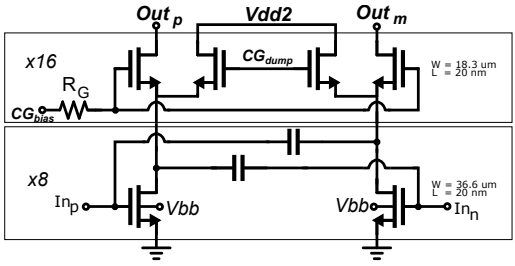

Fig. 5. Schematic of the PPAs

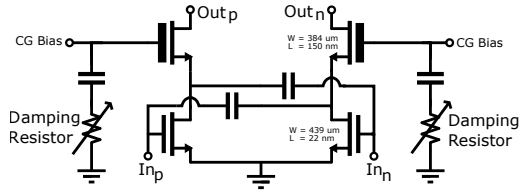

Fig. 6. Schematic of the Main and Auxiliary PAs

gain variation of about 2 dB, providing a fine step resolution < 0.5 dB, between the 16 course gain steps. The design is based on [33], with the difference of the signal dumping into a second dedicated supply voltage $Vdd2$. For power reduction purposes, this second supply voltage $Vdd2$ can be set to a lower value than the PPA supply voltage Vdd. A four bit tunable capacitor is implemented in the output resonance tank and allows the PPA to cover a wider frequency range by steering the output resonance frequency.

The main and auxiliary amplifiers of the Doherty PA are identical, see Fig. 6. Both consist of a thin oxide CS and a thick oxide CG stage, and have cross-coupled neutralization capacitors which for stability are conservatively dimensioned to about half of C_{gd}, as introduced for mmW PAs by [34] and theoretically explained by [35] to improve reverse isolation and to reduce input capacitance. The input capacitance of the CS stage resonates with the transformer between the PA and the PPA and thereby presents a real and high impedance to increase PPA voltage gain. For appropriate distribution with most of the large output voltage swing over the thick oxide CG transistor, the voltage swing at the gate of the CG should be minimized, for an increased effective g_m of the CG stage. The bias voltage is set using a resistive ladder and a relatively large capacitor to ground is used to provide low impedance for high frequencies. Unfortunately any parasitic inductance between the CG transistor and the CG decoupling capacitance, risk forming part of a resonance tank of a parasitic oscillator. To mitigate this problem a controllable damping resistance is added. The damping resistance is dimensioned to have minimum impact on the desired performance while still effectively suppressing the loop gain of the parasitic oscillator. Since the damping resistor is very small, tens of ohms, combined with very low gain from the CG-gate to the output, its noise impact will be negligible.

C. Adaptive Bias

A good implementation of an adaptive bias circuit must fulfill four requirements. First, possibility to control the small-signal operating point of the amplifier. Second, control at what signal level the adaptive bias should start to increase. Third, control of the slope of the bias increase after the turn-on point. Fourth, being able to drive the PA common mode input impedance, with a bandwidth exceeding that of the modulation. The adaptive bias generation circuit shown in Fig. 7 (a) uses a current mirror combined with a digitally controlled current source to set the small-signal operating point of the input common source (CS) transistor of the PA. Adjusting settings for turn-on and slope control has an unwanted impact on the DC-level output from the rectifying pair in Fig. 7 (a), at low input signal amplitudes, which should ideally be independent on large signal settings. To orthogonalize the settings of small-signal bias-level, turn-on point, and slope control, a negative feedback loop including an operational amplifier (OP-amp) and a downscaled replica of the rectifying pair in Fig. 7 (a), is therefore used to make the small-signal adaptive bias level independent of turn-on and slope control settings. The negative feedback loop includes, besides the OP-amp, a transistor with output current I_1, which produces a voltage together with slope control resistor R_1 in the rectifying pair replica. In the rectifying pair, the turn-on point of current I_2, that also sets the turn-on of the adaptive bias, is controlled using a resistive ladder controlling the gate bias of the PMOS CS pair. Slope control, after turn-on, is controlled using a variable resistor R_1, effectively transforming current I_2 into a voltage. Another feature of the adaptive bias generation circuit is the possibility to swap bias mode between constant GM mode for class AB operation and voltage reference mode for class C operation. In class AB bias mode I_0 is transformed to a voltage over a classic current mirror diode and in class C bias mode I_0 is transformed to a voltage over a 1k Ω resistor. Using a CW-tone input signal to the PA, the adaptive bias circuit produces expected amplitude dependent bias levels, which can be measured using a high impedance DC-path, see Fig. 17 and 18 in Section V-B. The adaptive bias settings, i.e. the level of current I_0 and settings for turn on and slope, can be adjusted for optimal performance for each sample, but also for large changes of conditions such as temperature or operating frequency. Neither of temperature and operation frequency are expected to have a major impact on the operation of the adaptive bias circuit itself, however, and it should produce a robust and controllable amplitude dependent bias voltage for the PA.

The proposed power amplifiers use neutralizing capacitors, which reduces input capacitance at the carrier frequency for differential signals, since the neutralizing capacitors ideally cancel the effect of C_{gd} of the CS transistors. For signals in common mode, like the adaptive bias, however, the input capacitance is increased by the neutralizing capacitors rather than decreased. Ignoring all capacitors, except for C_{gs} and

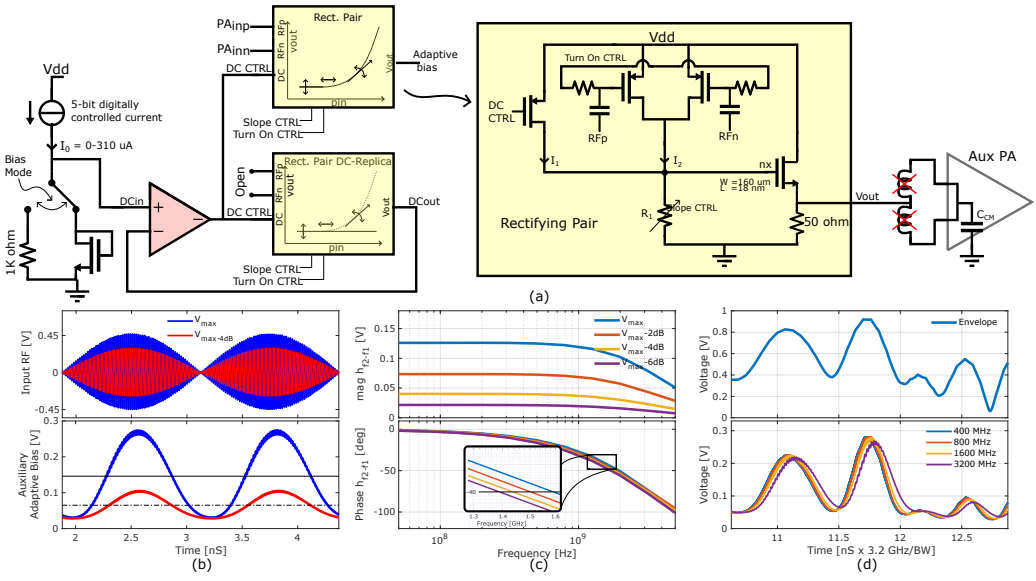

Fig. 7. (a): Simplified schematic of the adaptive bias generation circuit and the rectifying pair sub block. (b): Upper: Time domain two tone input signal to adaptive bias circuit. Lower: Time domain output signal from adaptive bias circuit. (c): Upper: Amplitude response for $f_{h_2-h_1}$ of adaptive bias circuit vs. tone separation for various input signal amplitudes. Lower: Phase response for $f_{h_2-h_1}$ of adaptive bias circuit vs. tone separation for various input signal amplitudes. (d): Upper: Envelope of an 800 MHz OFDM signal. Lower: Simulated corresponding adaptive bias signal for the auxiliary PA.

C_{gd}, the common mode input capacitance, that the adaptive bias need to drive, can be expressed as

$$C_{input_{CM}} = 2C_{gs} + 4C_{gd}(1 - G_{BB}) \quad (1)$$

where G_{BB} is the voltage gain at base-band frequencies, here for simplicity estimated as $G_{BB} \approx -g_{m_{CS}}/g_{m_{CG}} \approx -1$. This gives an input capacitance of $C_{input_{CM}} = 2C_{gs} + 8C_{gd}$, and becomes $C_{input_{CM}} \approx 840\,\text{fF}$ with simulated values for $C_{gs} \approx 170\,\text{fF}$ and $C_{gd} \approx 62.5\,\text{fF}$, and on top of that routing parasitics and other neglected capacitances will be added, increasing the total capacitance, which is simulted to about $1.4 - 1.8\,\text{pF}$, depending on the bias level. To be able to effectively drive this capacitance with wide signal bandwidth, the rectifying pair is equipped with a resistively loaded common drain (CD) output buffer, which is scaled to achieve low enough output impedance. The DC-level shifting is automatically compensated for by the feedback loop. Unfortunately the chip provides no possibility to directly measure a high frequency adaptive bias signal, but Fig. 7 (b) - (d) shows simulated results, of a layout extracted design. Figure 7 (b) shows input and output signals from the adaptive bias circuit for a two tone simulation with 800 MHz tone separation and two different input amplitude levels. The output signal from the adaptive bias circuit follows the peaks of the envelope and goes towards a bias floor at the envelope minimas, and can thus be approximated by DC plus the beat frequency tone, i.e. at the separation frequency of the two tones ($f_{h_2-h_1}$). The tone separation represents the fastest bias change possible for a signal BW equal to the frequency difference of the two tones. However, the two tone test case is very pessimistic, since in the frequency domain, it has all the energy located at the edges of the signal BW, but for an OFDM signal the energy is evenly distributed. In Fig. 7 (c) the magnitude and phase of $f_{h_2-h_1}$ is plotted vs. increasing tone separation. At 1.6 GHz tone separation, the magnitude has lost $<1.3\,\text{dB}$ compared to low frequencies and the phase is lagging by 41 degrees. The dominant pole causing the phase lag of the adaptive bias signal is located at the output at the CD buffer and the phase lag ϕ can be expressed as $\phi \approx tan^{-1}(\omega R_{out} C_{CM})$, where $R_{out} = R/(g_m R + 1)$, g_m is the transconductance of the CD buffer, C_{CM} is the capacitive load of the auxiliary PA, and $R = 50\,\Omega$. Due to the voltage dependency of C_{gs}, C_{CM} increases with increased bias level, which would slow down the adaptive bias signal in the peaks. However, this highly undesired effect is counteracted by reduction of the CD buffer output impedance for large signals, due to increased g_m. The combined effect is that the adaptive bias signal is slightly faster for high output signals, which can be seen as an inset in Fig. 7 (c). Nevertheless, the adaptive bias will still, linearise the signal at this very high modulation frequency. For a real 1.6 GHz modulated OFDM signal the improvement that the adaptive bias brings will not largely differ from that of a modulated OFDM signal with lower BW, since the majority of the modulation energy of the 1.6 GHz BW signal has much lower frequency. This result is confirmed by measurements comparing performance improvements brought

by the adaptive bias vs.signal BW, which are shown in Fig. 24 in Section V-B. Figure 7 (d) shows the input signal envelope and corresponding simulated output signal from the adaptive bias circuit for 400-3200 MHz BW OFDM input signals. To ease a comparison of the output bias signals, for the various OFDM bandwidths, the x-axis was normalized with respect to the modulation bandwidth.

D. Output Combiner Network

The output combiner should provide load modulation of the main amplifier when the class-C biased auxiliary amplifier delivers output power. This is achieved by the impedance inverting functionality of the $\lambda/4$-transmission line, connected between the main and auxiliary amplifiers. The $\lambda/4$-transmission line also adds 90 degrees of phase shift to the signal, from the main amplifier to the output. Therefore, for constructive combination at the output of the Doherty amplifier, the input signal to the auxiliary amplifier is phase shifted by the same amount. Figure 8 shows the design flow of the output combiner. Step (1) shows the ideal schematic

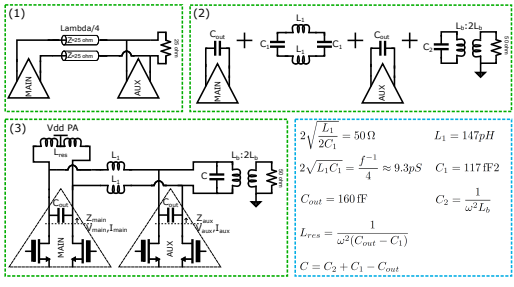

Fig. 8. Design steps (1)-(3) for the output combiner.

for a current combining Doherty amplifier, similar to as it was depicted in the original publication by W.H. Doherty in 1936 [36], but loaded by a 25 Ω differential load. In step (2), we implement the transmission line in step (1) using lumped components, which can merge the output capacitance C_{out} of the main and auxiliary amplifiers with the parallel capacitance C_1. In addition, we also add a balun at the output, transforming from a differential 25 Ω to a single-ended 50 Ω impedance. In step (3) the circuit parts in step (2) are merged, and inductor L_{res} and capacitor C are added to resonate with any remaining susceptance in the respective nodes. Since C_{out} is a parasitic of the PA, the simulated output current and impedance for the main and auxiliary amplifier is prior to this parasitic, i.e. inside the output transistor. Figure 9 shows the top-level layout of the output combiner with the TRX-switch to the right. To reduce impedance in ground and supply nets, multiple ground (Gnd) and supply (Vdd) pads are used in parallel.

E. LNA1 and LNA2

Figure 10 (a), shows LNA1, which is a single ended inductively source-degenerated cascode structure with series resistance of the capacitor at the cascode-gate to ensure stability.

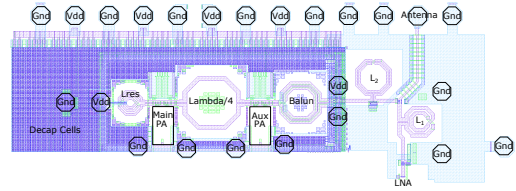

Fig. 9. The upper metal layers of the output combiner and TRX-switch network.

To dimension the input transistor the first step is to decide the V_{gs} bias level giving optimal $g_m/C_{gs} = \omega_t$ of the transistor, to maximise the gain. For good matching the real part of the input impedance should be equal to the antenna impedance, which becomes $Re(Z_{in}) = g_m L_s/C_{gs} = \omega_t L_s = R_s = 50\,\Omega$. The optimal width of the input transistor is then found by sweeping the width using a circuit simulator, while temporarily adding a series inductance L_g, matching the LNA input reactance $Im(Z_{in}) = \omega L_s - 1/(\omega C_{gs})$, for each point. When the optimal width is found, also the optimal LNA NF impedance, $Z_{opt} = 50 + j40\,\Omega$ is determined. The series gate inductor L_g, used to find the optimal input transistor width and Z_{opt} of the LNA, is then removed as the TRX-switch will provide the reactive part of the LNA-to-antenna matching. The output load is a tuned balun, which in turn, drives the LNA2, see Fig. 10 (b), with a differential signal.

LNA2 drives an output transformer with 100 Ω differential impedance, since the differential output signal from the secondary side of the transformer drives two 50 Ω single-ended wires, through package and PCB. An off-chip balun transforms the differential output signal to a 50 Ω single ended signal for measurements.

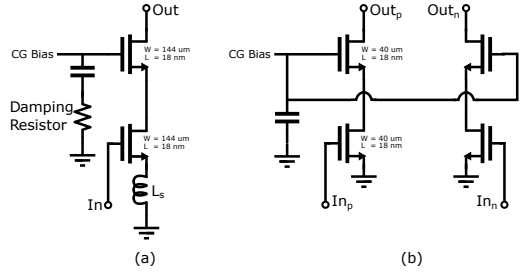

Fig. 10. (a) Schematic of LNA1. (b) Schematic of LNA2.

F. TRX-switch

A schematic and a 3D layout of the TRX-switch can be seen in Fig. 11. In TX mode both switch transistors (Q_1 and Q_2) are conducting. Inductor L_1 then resonates with C_1 and parasitics of Q_2, which results in a high impedance load at the antenna node from the RX port. L_2 is shunted by Q_2 and therefore the switch presents a low impedance and low loss from the PA to

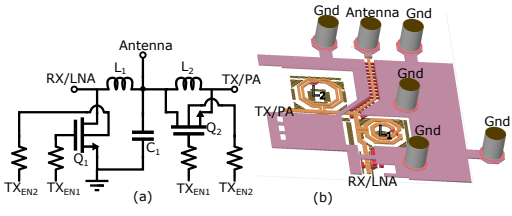

Fig. 11. (a) Schematic of the TRX-switch. (b) 3D model of the TRX-switch layout including copper pillars connecting to the package.

the antenna, see Fig. 12 (Upper). Since both Q_1 and Q_2 are conducting in TX mode, the switches must handle a large AC current, but there will be no high voltage across the transistors, that would cause reliability issues. While the Q_2 drain and source carry the full PA output voltage, the SOI transistor gate will track the other terminals, by using a large resistor in series with the gate, so signal voltages between terminals will be low. To further reduce the ON-resistance of Q_1 and Q_2 in TX mode, also the back back-gate is pulled high, but to a slightly higher voltage than the gate. The isolation between the PA and LNA input gate was simulated to >35 dB, see Fig. 12 (Lower), which gives a peak value of about 50 mV at the LNA input gate when the PA is delivering full output power.

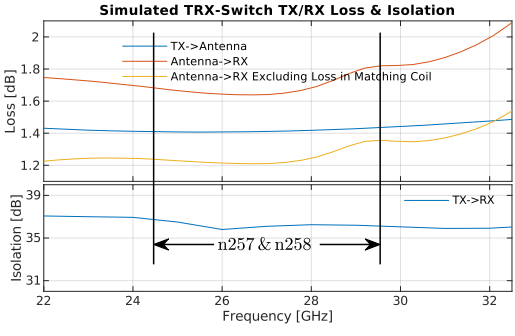

Fig. 12. Simulated loss and isolation in TX mode. The loss includes simulated losses of about 0.3 dB in package.

In RX mode both Q_1 and Q_2 are off. The drain to source capacitance of Q_2 is then in resonance with L_2, which minimizes loading from the PA branch on the antenna node in RX mode. The design of the TRX-switch has three steps. First, dimension the reactive components based on a known optimal source impedance for the LNA. Second, size transistor Q_1 for correct output capacitance C_{out1}. Third, size transistor for Q_2 to achieve a good balance between losses in TX and RX-mode. Figure 12 (Upper) shows the complete simulated loss from antenna to RX. In addition, to compare with switches without impedance matching, the estimated minimum loss required to achieve the LNA matching, without the switch functionality, was simulated and deducted from the complete loss from antenna to RX. At center frequency the difference is about 0.4 dB.

In this section, the element values are derived analytically. The admittance at the RX port is designed to equal the optimal source impedance $Z_{opt} = 1/Y_{opt}$ for LNA minimum noise figure (NF), which here is known from the LNA circuit design. Susceptance B_{L1} is equal to $-B_{C1}$ since L_1 is in resonance with C_1. The input admittance seen from the LNA port can then be written as:

$$Y_{opt} = G_{opt} + jB_{opt} = B_{L1}^2/G_L + j(B_{Cout1} + B_{L1}) \quad (2)$$

where B_{L1}, B_{Cout1} are the susceptances of L_1 and the output capacitance of Q_1 irrespectively, and G_L is the load conductance. From the equation above, B_{L1} and B_{Cout1} follow as:

$$B_{L1} = -\sqrt{G_L G_{opt}} \quad (3)$$

$$B_{Cout1} = B_{opt} + \sqrt{G_L G_{opt}} \quad (4)$$

A trade-off between losses in TX-mode and how much the transmitter loads the receiver when in RX-mode exists. In TX-mode Q_2 shunts L_2 and should be made large for low series on resistance, but in RX-mode the receiver will be loaded by the parallel resonance of L_2 and the output capacitance of Q_2, i.e. C_{cout2}. Table II summarizes the analytically derived ideal component values, which are further optimized when implemented in the layout of the TRX-switch.

TABLE II
TRX-SWITCH COMPONENT VALUES

Component	Realization/Value
Z_{opt}	50+j40 Ω
R_L	50 Ω
L_1	364 pH
L_2	609 pH
C_1	89 fF
Q_1	$W_g = 150\,\mu$m, $L_g = 18$ nm, $C_{out1} = 33$ fF
Q_2	$W_g = 240\,\mu$m, $L_g = 18$ nm, $C_{out2} = 53$ fF

The switch transistors Q_1 and Q_2 cause losses in both RX and TX mode, in RX mode due to the finite Q-value of their off-state output capacitance, and in TX mode due to their on-resistance. To investigate the impact of these losses the TRX-switch was simulated with the extracted views of Q_1 and Q_2 replaced by ideal, and appropriately sized, capacitors in RX mode and with ideal shorts in TX mode. Having Q_1 ideal gives 0.5/0.2 dB reduced loss in RX/TX-mode at 27 GHz, and having both Q_1 and Q_2 ideal gives 0.7/0.5 dB reduced loss in RX/TX-mode.

G. Decoupling Capacitors

Low voltage PAs with high output power require large supply currents. For a PA to be efficient, such as the Doherty PA, its supply current must follow the envelope of the input signal. The supply current of the Doherty amplifier thus have significant AC signal components. The bandwidth of

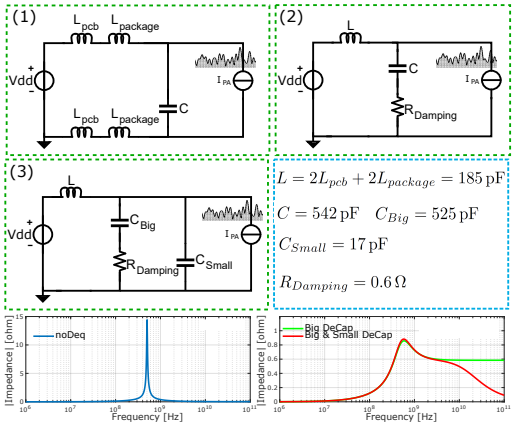

Fig. 13. Step (1)-(3) in supply network design. Significant reduction of peak impedance is achieved by adding a series resistor for the major part of the decoupling capacitors.

the envelope is wider than the modulation bandwidth due to the nonlinear transfer from Cartesian coordinates to envelope, $ENV = \sqrt{I^2 + Q^2}$, and consequently the AC signal will have a bandwidth that exceeds that of the modulation. Decoupling capacitors will form a resonance tank together with parasitic inductances from on chip routing, chip-to-package copper pillar interconnects, package wire routing, package-to-PCB bumps, and finally PCB wires to the large off chip capacitors. Figure 13 shows the design flow of the decoupling network estimating the parasitic resonance and how to reduce the peak impedance. The available area surrounding the PA is filled with as much decoupling capacitance as possible, in total this becomes $542\,\text{pF}$. The supply and ground inductors are estimated to sum up to $185\,\text{pH}$, which gives an oscillation frequency of $0.5\,\text{GHz}$. Both decoupling capacitors and the parasitic inductors have high Q-value, with the overall quality factor of the parasitic resonance tank estimated to be 25, which gives a peak impedance of $15\,\Omega$, see Fig. 13, blue curve. This will result in strong ringing as the PA injects a current signal, containing wideband power, into the supply resonance tank. Furthermore, frequency dependent node impedances, at base-band, fundamental, and harmonics, are the root cause for electrical memory effects [37], which are highly unwanted in any PA design. To mitigate the problem, the parasitic resonance tank is damped down to a Q-value of 1, which will limit the magnitude of the impedance peak, see Fig. 13, green curve. In addition a small part of the decoupling capacitor is left without any damping to further reduce supply network impedance at high frequencies, at a small expense of the performance at the parasitic oscillation frequency, see Fig. 13, red curve. The exact sizing of the small undamped decoupling capacitor is not very crucial. It should be small enough not to significantly affect the impedance at the parasitic resonance frequency, but still large enough to significantly reduce the impedance at high frequencies. Here, mostly due to layout considerations, it was chosen to $17\,\text{pF}$.

V. Measurement Results

Figure 14 (Left) shows a photo of the chip fabricated in a 22 nm FD-SOI CMOS process. It was flip-chip mounted in a package that was soldered to a PCB, see Fig. 14 (Middle). The PCB has DC-coupled, differential IQ transmitter input signals, single-ended IQ-LO signals, a single-ended RF antenna port, and a differential receiver RF output, all conducted, matched to $50\,\Omega$, and measured in room temperature. In total the chip uses four analog supplies in transmit mode, Vdd PA (2 V), Vdd PPA (1 V), Vdd Mixer (1 V), and Vdd Analog (0.8 V). In receive mode the chip uses two analog supplies, Vdd LNA (1.6 V) and Vdd Analog (0.8 V).

Figure 14 (right) shows simplified versions of the transmitter and receiver test benches used when measuring the performance of the transceiver. Insertion loss in PCB routing and cables were also measured and de-embedded. For all transmitter measurements the IQ LO signal to the two complex mixers with rearranged phases, used for up-conversion and phase shifting of input signal to main and auxiliary amplifiers, were fed from a Keysight VSG M9348B vector signal generator. For transmitter CW-tone measurements, the RF output signal was fed to the Keysight UXA N9040B signal analyser and the measured output power was verified using an R&S NRP Z31 power sensor. When performing modulated measurements the RF output signal was connected to a Keysight UXR 0404A real-time oscilloscope. For all receiver measurements an R&S SMW 200A vector signal generator, with built in noise source, was used for input signal and the differential $100\,\Omega$ output signal from LNA2 was converted to a single-ended $50\,\Omega$ signal and fed to an R&S FSW 67 spectrum analyzer.

A. Receiver Results

Figure 15 shows measured small signal gain (SSG) and NF for LNA1 and LNA2. The gain exceeds 20 dB for low, mid, and high frequency tuning and the NF, including losses in the TRX-switch, is below 4 dB at 27.25 GHz. The 3 dB BW exceeds 3 GHz. A combined power consumption was measured to 65.6 mW (41 mA from a 1.6 V supply) for the two LNAs. Since the receiver does not have down conversion and base-band filter, which would affect the out of band linearity, focus is on measuring inband linearity performance. Therefore, a two tone test at 24-28 GHz, with a narrowband tone separation of 10 MHz, was used to characterize the large signal gain and inband linearity of the receiver, which is summarized in Fig. 16. For the measured frequency range the inband IIP3 is above -10 dBm.

B. Transmitter Results

1) CW-Tone Measurements: Fundamental transmitter performance was measured using CW-tone stimuli. Figure 17 shows output power vs. input power and the aggregated AM-AM of the complete transmitter, including the conversion gain of the mixers, the gain of PPA and the PA. Adaptive bias, for the PA CS-transistors, becomes a DC-voltage for CW-tone stimuli, and can be measured through a low frequency

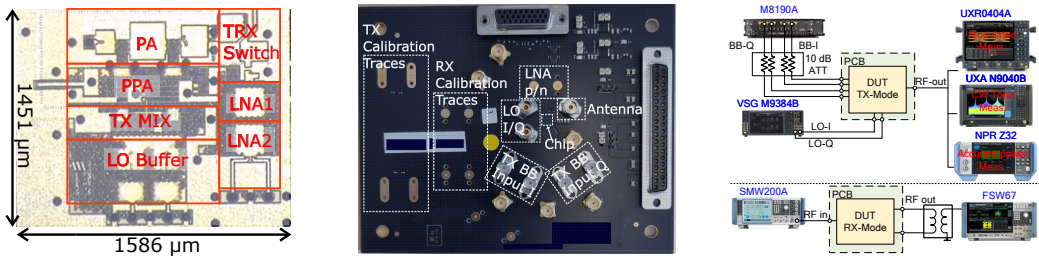

Fig. 14. Left: Chip photo of the complete transceiver front-end. Middle: PCB with the mounted chip in package near the middle. Right Upper: Simplified transmitter measurement setup. Right Lower: Simplified receiver measurement setup.

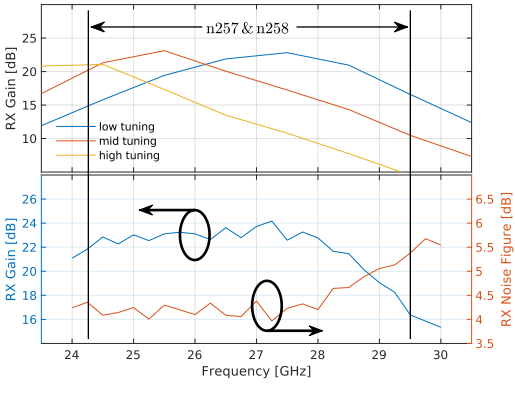

Fig. 15. Upper: RX gain vs. frequency for low/mid/high tuning. Lower: RX NF and gain vs. frequency with optimized settings for each frequency. For both plots NF and Gain are including TRX-switch, LNA1, and LNA2.

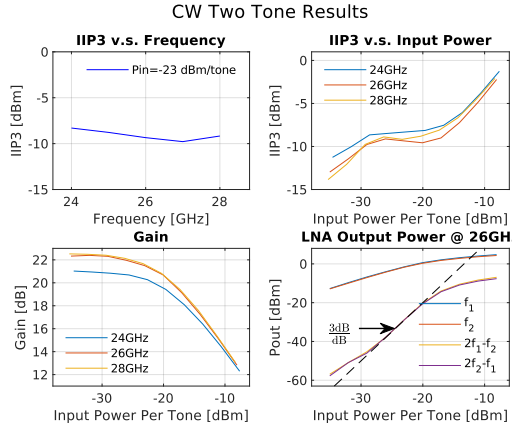

Fig. 16. Summarized results for receiver two tone test.

output. The lower left sub-figure of Fig. 17 shows how the bias level for the auxiliary amplifier gradually starts to increase for low input signal levels and reaches roughly the same bias level as the main amplifier at maximum output power. For an accurate efficiency estimate of the power amplifier the output power was measured using an R&S NRP Z31 power sensor. At the PA output the 3 dB compression point is at 18.7 dBm, which can be considered the maximum useful output power. The efficiency at 6 dB back-off from the 3 dB compression point is 13.7 % for the Doherty amplifier. The corresponding efficiency for a class B and class A amplifier with the same peak efficiency would be 7.6 % and 3.4 %, which demonstrates a back-off efficiency boost of 1.8 x and 4 x, respectively.

To evaluate the impact when adjusting the two adaptive bias parameters, turn-on control and slope control for the auxiliary PA bias, as explained in Section IV-C, CW-tone measurements were carried out for nine different settings, low, mid, and high for both turn-on and slope control. The circuit offers 8 settings for each parameter, which results in $8 \times 8 = 64$ possible settings. Figure 18 compares the AM-AM characteristics, and how the PA CS bias voltage changes with the input signal power, for the nine chosen adaptive bias settings that were measured. The slope of the increase of the adaptive bias, after turn-on, is controlled by adjusting resistor R_1 and the turn-on of adaptive bias is controlled by adjusting the bias level of the rectifying pair, both shown in Fig. 7 (b). Having low gate bias voltage to the PMOS rectifying pair shifts the turn-on to occur at lower input signal amplitudes. The selected setting, as used in modulated measurements, is $R_1 = Mid$ and $DC_{ctrl} = Low$. which improves the 1 dB compression point by more than 3 dB, compared to having $R_1 = Low$ and $DC_{ctrl} = High$, which to all practical means is identical to turning off the adaptive bias, i.e. applying a static bias.

Amplitude modulation to phase modulation (AMPM), shown in Fig. 19, was measured by gradually increasing a DC signal level for the I and Q baseband input signals and comparing the PA output signal phase with one of the LO signals. For the adaptive bias case the phase deviation at 1 dB compression point is about 13 degrees compared to the phase at low output signal level, and as can be seen the adaptive bias reduced AMPM at all output power levels.

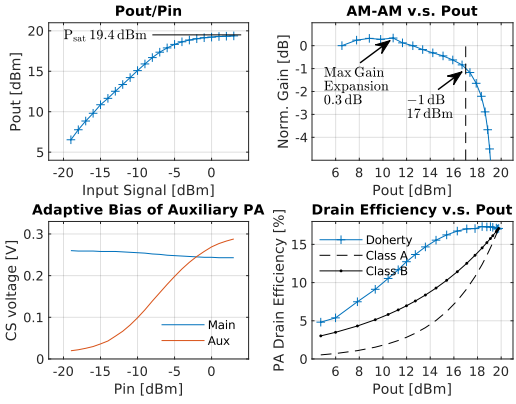

Fig. 17. PA 26.5 GHz CW-tone measurements. Gain is for complete transceiver. PA output P_{sat} is 19.4 dBm, 1 dB compression point is 17 dBm, and peak efficiency of the PA is 17.3 %.

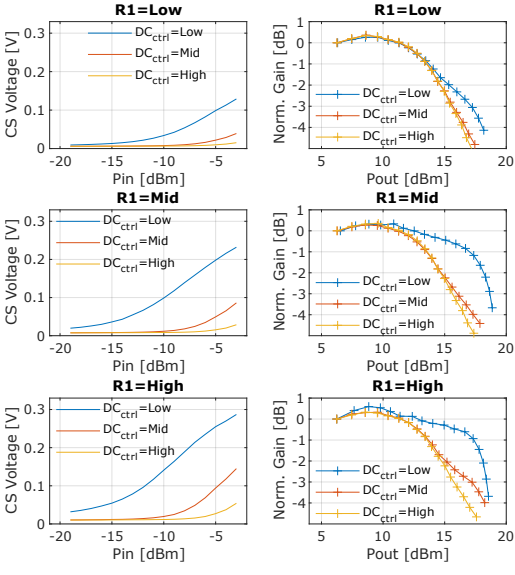

Fig. 18. 26.5 GHz CW-tone measurements. Comparison of AM-AM for nine different adaptive bias settings.

To test the BB gain flatness and RF output frequency response, the transmitter was measured using three different LO frequency settings 24.25 GHz, 26.5 GHz, and 29.5 GHz. For each of the three LO frequencies the transmitter was retuned for optimal performance, after which the frequency of the complex IQ-input signal was swept from -2 GHz to +2 GHz without doing any modification to the operation of the circuit. First the test was carried out with a large input signal, giving saturated output power for low BB offset frequencies, see Fig.

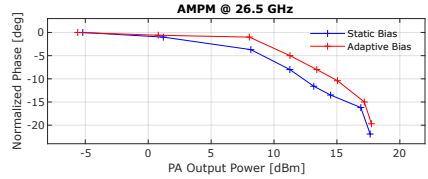

Fig. 19. AMPM from 26.5 GHz CW-tone measurement.

20. As the offset frequency was increased the measured output power was reduced both due to reduced signal reaching the PA, due to BW limitations in the mixer and PPA, but also due to frequency dependent transfer function in the output combiner and the TRX-switch. It should be noted that the RF center frequency can be fine tuned with much higher resolution than the three selected frequencies in Fig. 20. Second, SSG and PPA

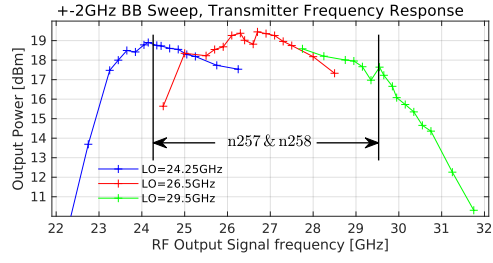

Fig. 20. Transmitter frequency response when sweeping base-band input signal frequency ± 2 GHz, using 24.25 GHz, 26.5 GHz, and 29.5 GHz local oscillator.

gain steps were measured using the same measurement set-up for all the 16 PPA gain steps, but with a low input signal, and only for an LO frequency of 26.5 GHz, see Fig. 21. The results in the figures show that the output power reduces by less than 1.5 dB for a BW that exceeds 3 GHz, and the SSG has a 1 dB BW of 1.2 GHz and a 3 dB BW of 2.4 GHz, fairly independent of PPA gain setting.

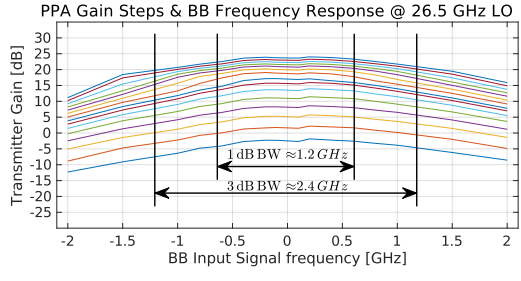

Fig. 21. Transmitter frequency response when sweeping base-band input signal frequency ± 2 GHz, using 26.5 GHz local oscillator for the 16 PPA gain steps.

2) Modulated Measurements: One major challenge when amplifying an OFDM signal is the high peak-to-average ratio (PAR). The high PAR of the OFDM signal is due to combining a large number of sub-carriers that are statistically independent. The theoretical maximum output signal amplitude is A x N, where A is the amplitude of each individual sub-carrier and N is the number of sub-carriers. However, this is extremely unlikely to occur when N is large. The I/Q signals then approaches asymptotically a Gaussian distribution, with zero average, and the resulting amplitude then follows a Rayleigh distribution, where the PAR value can be expressed as a probability function, see Eq. 5. [38]

$$PAR = \sqrt{ln(\frac{1}{P_0})} \tag{5}$$

which for a probability of $P_0 = 10^{-4}$ gives a PAR equal to $9.6\,\mathrm{dB}$.

To assess the performance of the transmitter for high PAR high bandwidth modulated signals, it was tested with 5G NR OFDM signals with 16-QAM and 64-QAM modulation and with varying bandwidth. The input signal was configured as 4x, 8x, or 16x 100 MHz wide signals, for a total signal bandwidth ranging from 400 MHz to 1.6 GHz. Each of the 100 MHz wide 5G NR OFDM signals have 792 subcarriers with a subcarrier spacing of 120 kHz. In total that gives a bandwidth of $792 \times 120\,\mathrm{kHz} = 95.04\,\mathrm{MHz}$, giving a utilization degree of 95 %. This means that a 16×100 MHz 5G NR 64-QAM signal will transmit $16 \times 792\,\mathrm{sym} \times 120\,\mathrm{kHz} = 1.52064\mathrm{Gsym/s}$ and each symbol using 64-QAM corresponds to 6 bits, resulting in a raw data rate of $9.123840\,\mathrm{Gbit/s}$.

Figure 22 exhibits the measured spectrum and constellation diagrams, as instrument screen dumps, including pilot symbols using lower order modulations. No predistortion or IQ-correction was used in any of the modulated measurements. The LO-leakage calibration algorithm was successfully used, suppressing LO-leakage to more than 40 dB below the modulated output power, for all measured bandwidths. When increasing the signal bandwidth only a minor penalty in output power, efficiency, and linearity was noted. Both high base-band frequency for CW-tones, as shown in Fig. 20, and fast adaptive bias that accurately tracks the envelope of the input signal as simulated and shown in Fig. 7 (b)-(d), are essential to success.

To estimate the improvement that the adaptive bias brings, the circuit was measured both using adaptive bias and with a static class-C bias level for the auxiliary amplifier. Figure 23 clearly shows the improvement for EVM and ACLR with modulated output power. Due to the high PAR of the input signal the adaptive bias starts improving the performance already at relatively low output power levels, and for intended maximum modulated output power level about 3 dB improvement is observed for EVM and ACLR, with neglectable reduction in efficiency. In addition, measurement to estimate BW dependency of the improvement offered by the adaptive bias are shown in Fig. 24. The relatively BW independent improvement, for the tested BWs, is in agreeance with the predictions from the adaptive bias circuit simulations in Section IV-C. Image rejection, LO-leakage ratio, inband power, $ACLR_1$, and $ACLR_2$ were measured using a 100 MHz wide 5G NR signal located at 200 MHz base band frequency and up-converted to 26.7 GHz using an LO frequency of 26.5 GHz. The LO leakage was suppressed using the proposed method from Section IV-A. LO IQ signals originate from a lab instrument and therefore the phase difference between the I and Q signals is assumed to be unknown when reaching the on-chip baluns. To circumvent this problem, which would not occur when the LO signal is generated locally in the chip, the phase difference between LO I and LO Q is adjusted on the signal generator until the image is minimum. It should be pointed out that since the LO IQ signals are used on the IQ-mixers in both the main and auxiliary paths it is not possible with the measurement setup to compensate for on chip LO IQ imbalance after the signal is split to the two paths. The LO phases are also rearranged between the two mixers and the mismatch will be unique for each mixer. The wideband frequency response is shown in Fig 25, without compensating for any base-band IQ measurement set-up mismatches the transmitter demonstrates excellent image rejection of -44.3 dB. The LO-leakage is measured to -26.1 dBm, giving a leakage ratio of 41.4 dB. The power amplifier delivers an inband power of 15.3 dBm with a DE of 15.5% for an EVM of -19.1 dB.

A power consumption break down for the transmitter, including all on chip circuitry needed for transmission with the measured performance in Fig. 22, is shown in table III.

TABLE III
TRANSMITTER POWER CONSUMPTION BREAK DOWN @ 26.5 GHZ

QAM Mod.	BW [MHz]	PA 2 V [mW]	PPA 1 V [mW]	Mixer 1 V [mW]	Analog(1) 0.8 V [mW]	TX Complete [mW]	Avg. Pout PA [dBm]
16	400	169.4	62.4	61.9	33.9	327.6	13.9
16	800	156.4	62.4	61.9	33.8	314.5	13.4
16	1600	154.6	60.7	61.9	33.6	310.8	13.3
64	400	116.4	57.9	61.9	33.1	269.3	11.4
64	800	114.2	57.8	61.9	33	266.9	11.2
64	1600	99.4	56.9	61.9	33	251.2	10.1

(1) Of which the LO-Buffers consume about 27 mW, approximately 4 mW is consumed by the adaptive bias generation, and general bias generation consumes about 2.5 mW.

A summary of the transmitter performance for modulation, bandwidth, output power, EVM, drain efficiency (DE) of PA, data rate, and power added efficiency (PAE) is provided in table IV.

TABLE IV
TRANSMITTER 5G NR OFDM PERFORMANCE @ 26.5 GHZ

QAM Mod.	BW [MHz]	Avg. Pout PA [dBm]	Avg. Pout TRX Switch [dBm]	EVM [dB]	Raw Data Rate [G Bit/s]	Avg. D.E. PA [%]	Avg. PAE TX [%]	Avg. PAE TX Incl. Freq.Conv [%]
16	400	13.9	12.8	-20.2	1.52	14.3	8.2	5.8
16	800	13.4	12.3	-20.3	3.04	14.1	7.8	5.4
16	1600	13.3	12.2	-19.8	6.08	13.8	7.7	5.3
64	400	11.4	10.3	-23.0	2.28	11.9	6.1	4.0
64	800	11.2	10.1	-22.7	4.56	11.6	5.9	3.8
64	1600	10.1	9.0	-23.3	9.12	10.4	5.1	3.2

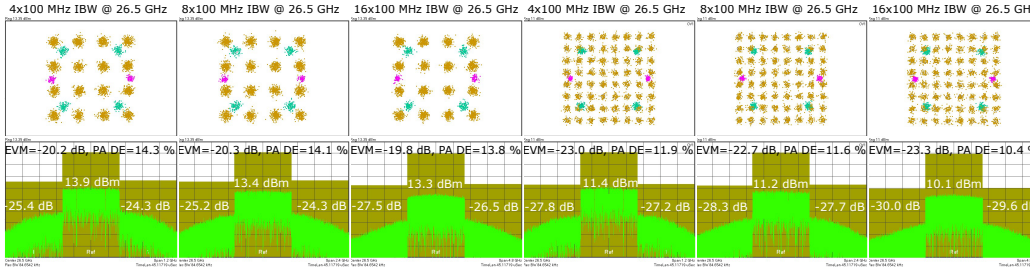

Fig. 22. Measured performance using modulated 5G NR 16QAM/64QAM-OFDM signals for (400 MHz, 800 MHz, 1600 MHz) BW, and with PAR value that follows Eq. 5. Pilot symbols can be seen in cyan and purple color. Output power and DE stated in the figure is average value for the modulated signal.

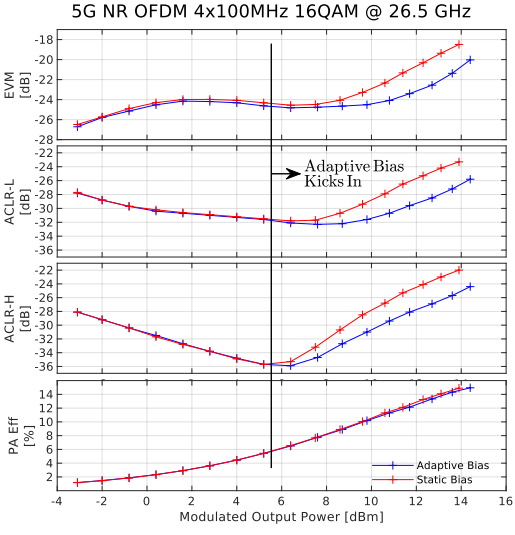

Fig. 23. Measured performance of EVM/ACLR/DE for adaptive bias and static class-C bias on auxiliary amplifier, using a 5G NR 400 MHz OFDM 16QAM-OFDM signal.

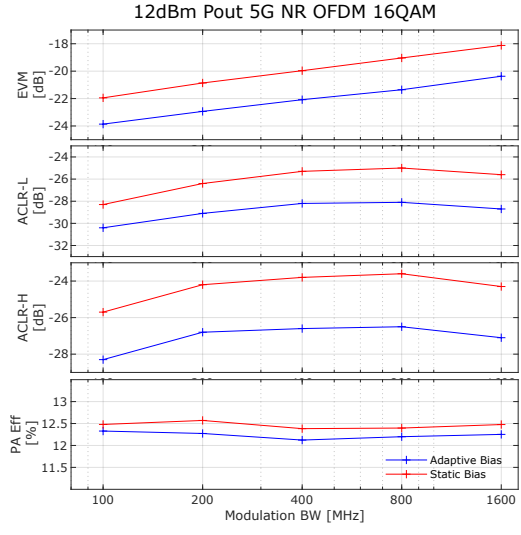

Fig. 24. Measured performance of EVM/ACLR/DE for adaptive bias and static class-C bias on auxiliary amplifier using a 5G NR OFDM 16QAM-OFDM signal with increasing BW.

C. Comparison Table

The performance of the proposed transceiver is summarized in table V and compared with published state-of-the-art transceivers in the same frequency range. All compared numbers are referring to chip output, i.e. including losses in the TRX-switch. This work is the only transceiver front-end that includes frequency up-conversion from base-band IQ signals and including TR-switch. It is also one of few that takes BW limitations of the BB part of the transmitter into consideration. Frequency up-conversion adds significant complexity and contributes to EVM due to non-ideal image rejection, LO-leakage, non-flat BB frequency response, and mixer distortion. Furthermore, the design is carried out in a 22 nm process node, which enables higher level of integration with digital circuitry, but makes it harder to reach high output power. Nevertheless, it still has the highest measured saturated output power and the highest modulated output power for 400/800/1600 MHz IBW for all OFDM modulated signals, which have significantly higher PAR than single carrier signals. Without power consumption related to frequency conversion (LO-Buffers and Mixers), [12] reports high transmitter modulated efficiency, but for single-carrier signals. When comparing to OFDM transmissions, the proposed transceiver front-end achieves the highest modulated efficiency for all modulation BW exceeding 400 MHz. Finally, the receiver NF is the lowest reported for all front-ends.

VI. Conclusion

The design of an efficient transceiver front-end suitable for 5G base stations covering 24.25-29.5 GHz frequency has been

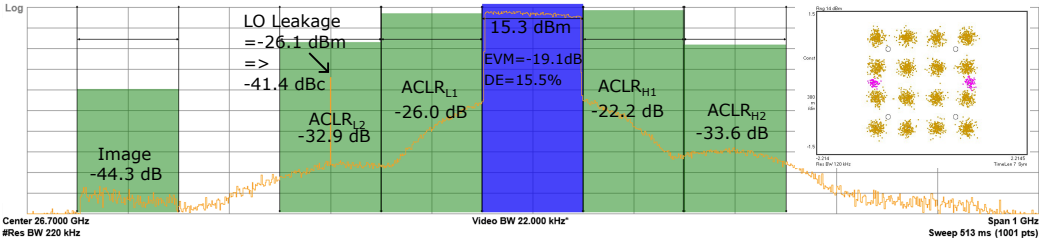

Fig. 25. 100 MHz modulated 5G NR 16QAM signal centered at 200 MHz BB frequency, with LO at 26.5 GHz, measured at PA output, showing inband power, EVM, DE, $ACLR_1$, $ACLR_2$, LO leakage, and image rejection. Constellation diagram as an inset.

TABLE V
COMPARISON WITH STATE-OF-THE ART MMW TRANSCEIVER FRONT-ENDS

Parameter	**This Work**		Yi [13] [JSSC'22]		Quadrelli [39] [JSSC'22]	Pashaeifar [29] [JSSC'21]	Zhu [12] [TMTT'21]		Pang [8] [JSSC'20]		Park [6] [ISSCC'19]	
Technology	**22 nm CMOS FDSOI**		65 nm CMOS Bulk		28 nm CMOS Bulk	40 nm CMOS Bulk	65 nm CMOS Bulk		65 nm CMOS Bulk		28 nm CMOS Bulk	
TRX	**Single TRX**		4xTRX		Single TRX	Single TX	4xTRX		8xTRX		Single TRX	
TRX type	**TX zero IF RX RF only**		TRX RF Only		TRX RF-IF	TX zero IF No RX	RF only		Bidirectional RF only 2-pol		RF only	
	24.25-29.5 GHz		24-29.5 GHz		22-31 GHz	24-30 GHz	24-28 GHz		26.5-29.5 GHz		25.8-29 GHz	
Area (mm^2)	**2.3**		4.5$^{(3)}$		1.4	1.38	0.94		0.58		0.67	
Freq. Upconversion	**Yes**		No		Yes	Yes	No		No		No	
Switch	**Yes**		Yes		Yes	No	Yes		No		Yes	
Switch Loss (dB)	**1.1 in TX**		1 in TX		N.R.	n/a	1		n/a		n/a	
	1.2 in RX		0.5 in RX		N.R.	n/a						
Compared Freq.	**26.5 GHz**		26 GHz		28 GHz	27 GHz	28 GHz		28 GHz		28 GHz	
TX/Element												
P1dB (dBm)	**15.9**		17.6		11.5	17.9$^{(2)}$	16.1		10.3$^{(2)}$		13.9	
Psat (dBm)	**18.3**		18		14	18.9$^{(2)}$	18.1		14.1$^{(2)}$		17	
PAE P1dB (%)	**13.4**		20.4		17	27$^{(2)}$	16.6		6.4$^{(2)}$		16	
Gain (dB)	**24**		23		20	20.7$^{(2)}$	33		19$^{(2)}$		17	
RF BW (GHz)	**2.4**		6		9	6$^{(3)}$	4		3$^{(3)}$		4	
RX/Element					7,17							
NF (dB)	**4.2**		4.3		8.5	n/a	4.5		5.2$^{(2)}$		4.5	
P1dB (dBm)	**$>$-19.6**		-22.2		-18.8	n/a	-16.1		-26.8$^{(2)}$		N.R.	
Gain (dB)	**23**		14.2		29	n/a	27.2		15$^{(2)}$		N.R.	
RF BW (GHz)	**4**		6		9	n/a	4		3$^{(3)}$		4	
PDC (mW)	**65.6**		82		110	n/a	45		112		44	
Modulated Meas.	**Cond.**		Cond.		Cond.	Cond.	Cond.		OTA		Cond.	
IRR (dB)	**44.3**		n/a		40	$>$50	n/a		n/a		n/a	
LO-leakage (dB)	**-41.4**		n/a		-30	-45	n/a		n/a		n/a	
Signal type	**5G OFDM**		5G OFDM		5G OFDM	5G OFDM	S.C. wideband		5G OFDM		OFDM	
Signal BW (MHz)	**16x100**	**16x100**	1200	400	400	8x100	540	405	400	400	100	100
QAM-Constellation	**16**	**64**	64	256	64	64	64	256	64	256	64	256
Symb. Rate Gs/s	**1.52**	**1.52**	1.14	0.38	0.38	0.76	0.4	0.3	0.38	0.38	0.1	0.1
Raw Data Rate Gb/s	**6.1**	**9.1**	6.84	3.04	2.28	4.56	2.4	2.4	2.3	3.04	0.6	0.8
EVM (%)	**10.2**	**6.7**	5.6	3.5	5.01	4.41	4.8	2.9	1.4	1.4	4.9	3.1
Pout (dBm)	**12.2**	**9**	3.9	6	5.38	7.3$^{(2)}$	13.5	10.2	5.4$^{(2)}$	4.5$^{(2)}$	8.45	7.1
TX PAE (%)	**7.7**	**5.1**	2$^{(1)}$	3.7	4.5$^{(1)}$	5.8	8.4	3.5	$<$2	$<$2	5$^{(1)}$	3.5$^{(1)}$
TX PAE incl freq conv (%)	**5.3**	**3.2**	n/a	n/a	$<$ 2.5$^{(1)}$	3.5	n/a	n/a	n/a	n/a	n/a	n/a

N.R. Not reported, n/a Not applicable, S.C. Single Carrier, $^{(1)}$ Graphically estimated. $^{(2)}$ Assuming 1.1 dB TRX-switch loss for comparison. $^{(3)}$ Ignoring BW limitations by TRX-switch. Yi JSSC'22: For 400 MHz BW 10.3 dBm P_{out} with 7.6 % PAE is reported. Pashaeifar JSSC'21: TX PAE and TX PAE including frequency conversion is estimated after TRX-switch loss and using power break down graph in Fig. 10. Pang JSSC'20: Modulated P_{out} is calculated from EIRP for 8 elements and combined PCB losses and antenna gain of -4 dB. TX PAE is estimated from figure 19. Park ISSCC'19: TX PAE is graphically estimated from figure 9.8.3.

presented. Besides an analysis of the architecture choices, the article provides detailed examination of adaptive bias in combination with a Doherty amplifier, design strategy for the output combination network of an integrated mmW Doherty amplifier, a novel TRX-switch optimized for RX mode to favour UL due to UE limited output power, and a design methodology for supply network and decoupling for low voltage/high power efficient PAs. Measurements, both CW-

tone and wide-band modulated, confirm functionality and performance gains of the adaptive bias, LO-leakage calibration, wideband BB/RF output power, and Doherty back-off efficiency. An excellent RX NF confirms TRX-switch integration with LNA impedance matching. When compared to relevant CMOS transceiver front-ends in the same frequency range the transmitter delivers state-of-the-art modulated output power for 5G NR 16/64 QAM OFDM signals for all measured bandwidths (400 MHz, 800 MHz, 1600 MHz) and state-of-the-art efficiency above 400 MHz BW. The receiver reaches state-of-the-art NF among CMOS transceiver front-ends. Furthermore, the transceiver front-end occupies an area small enough for convenient integration in a large AAS.

VII. Acknowledgment

The authors would like thank Rohan Raghunath and Annie Ydström for their strong support in automated test development, de-embed measurements and actual performance testing. Furthermore the authors would like to acknowledge, Magnus Sandgren for top level layout work, Martin Hansson for package design, and Housam Elfeituri for design of PCB RF traces.

References

[1] Ericsson, "On the road to breaking the energy curve," Internet:https://www.ericsson.com/4aa14d/assets/local/about-ericsson/sustainability-and-corporate-responsibility/documents/2022/breaking-the-energy-curve-report.pdf, Oct. 2022.

[2] J. Pang *et al.*, "A 28-GHz CMOS Phased-Array Transceiver Based on LO Phase-Shifting Architecture With Gain Invariant Phase Tuning for 5G New Radio," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 5, pp. 1228–1242, 2019.

[3] 3GPP, "Technical Specification 38.104 (V17.2.0) Base Station (BS) Radio Transmission and Reception," Jul. 2021.

[4] S. Shakib, M. Elkholy, J. Dunworth, V. Aparin, and K. Entesari, "A Wideband 28-GHz Transmit–Receive Front-End for 5G Handset Phased Arrays in 40-nm CMOS," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 7, pp. 2946–2963, 2019.

[5] S. Mondal, R. Singh, and J. Paramesh, "21.3 A Reconfigurable Bidirectional 28/37/39GHz Front-End Supporting MIMO-TDD, Carrier Aggregation TDD and FDD/Full-Duplex with Self-Interference Cancellation in Digital and Fully Connected Hybrid Beamformers," in *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2019, pp. 348–350.

[6] J. Park, S. Lee, D. Lee, and S. Hong, "A 28GHz 20.3%-Transmitter-Efficiency 1.5°-Phase-Error Beamforming Front-End IC with Embedded Switches and Dual-Vector Variable-Gain Phase Shifters," in *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2019, pp. 176–178.

[7] L. Gao and G.M. Rebeiz, "A 22–44-GHz Phased-Array Receive Beamformer in 45-nm CMOS SOI for 5G Applications With 3–3.6-dB NF," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 11, pp. 4765–4774, 2020.

[8] J. Pang *et al.*, "A 28-GHz CMOS Phased-Array Beamformer Utilizing Neutralized Bi-Directional Technique Supporting Dual-Polarized MIMO for 5G NR," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 9, pp. 2371–2386, 2020.

[9] X. Tang, Y. Liu, G. Mangraviti, Z. Zong, K. Khalaf, Y. Zhang, W.M. Wu, S.H. Chen, B. Debaillie, and P. Wambacq, "Design and Analysis of a 28 GHz T/R Front-End Module in 22-nm FD-SOI CMOS Technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 6, pp. 2841–2853, 2021.

[10] J. Pang *et al.*, "A CMOS Dual-Polarized Phased-Array Beamformer Utilizing Cross-Polarization Leakage Cancellation for 5G MIMO Systems," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 4, pp. 1310–1326, 2021.

[11] Y. Yin, B. Ustundag, K. Kibaroglu, M. Sayginer, and G.M. Rebeiz, "Wideband 23.5–29.5-GHz Phased Arrays for Multistandard 5G Applications and Carrier Aggregation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 1, pp. 235–247, 2021.

[12] W. Zhu, J. Wang, X. Zhang, W. Lv, B. Liao, Y. Zhu, and Y. Wang, "A 24–28-GHz Four-Element Phased-Array Transceiver Front End With 21.1%/16.6% Transmitter Peak/OP1dB PAE and Subdegree Phase Resolution Supporting 2.4 Gb/s in 256-QAM for 5-G Communications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 6, pp. 2854–2869, 2021.

[13] Y. Yi, D. Zhao, J. Zhang, P. Gu, Y. Chai, H. Liu, and X. You, "A 24–29.5-GHz Highly Linear Phased-Array Transceiver Front-End in 65-nm CMOS Supporting 800-MHz 64-QAM and 400-MHz 256-QAM for 5G New Radio," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 9, pp. 2702–2718, 2022.

[14] B. Sadhu, A. Paidimarri, W. Lee, M. Yeck, C. Ozdag, Y. Tojo, J.O. Plouchart, X. Gu, Y. Uemichi, S. Chakraborty, Y. Yamaguchi, N. Guan, and A. Valdes-Garcia, "A 24-to-30GHz 256-Element Dual-Polarized 5G Phased Array with Fast Beam-Switching Support for >30,000 Beams," in *2022 IEEE International Solid- State Circuits Conference (ISSCC)*, vol. 65, 2022, pp. 436–438.

[15] B. Sadhu, Y. Tousi, J. Hallin, S. Sahl, S. Reynolds, O. Renström, K. Sjögren, O. Haapalahti, N. Mazor, B. Bokinge, G. Weibull, H. Bengtsson, A. Carlinger, E. Westesson, J.E. Thillberg, L. Rexberg, M. Yeck, X. Gu, D. Friedman, and A. Valdes-Garcia, "7.2 A 28GHz 32-element phased-array transceiver IC with concurrent dual polarized beams and 1.4 degree beam-steering resolution for 5G communication," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, 2017, pp. 128–129.

[16] J.D. Dunworth, A. Homayoun, B.H. Ku, Y.C. Ou, K. Chakraborty, G. Liu, T. Segoria, J. Lerdworatawee, J.W. Park, H.C. Park, H. Hedayati, D. Lu, P. Monat, K. Douglas, and V. Aparin, "A 28GHz Bulk-CMOS dual-polarization phased-array transceiver with 24 channels for 5G user and basestation equipment," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, 2018, pp. 70–72.

[17] A. Verma *et al.*, "A 16-Channel, 28/39GHz Dual-Polarized 5G FR2 Phased-Array Transceiver IC with a Quad-Stream IF Transceiver Supporting Non-Contiguous Carrier Aggregation up to 1.6GHz BW," in *2022 IEEE International Solid- State Circuits Conference (ISSCC)*, vol. 65, 2022, pp. 1–3.

[18] J. Pang, Y. Zhang, Z. Li, M. Tang, Y. Liao, A.A. Fadila, A. Shirane, and K. Okada, "A Power-Efficient 24-to-71 GHz CMOS Phased-Array Receiver Utilizing Harmonic-Selection Technique Supporting 36dB Inter-Band Blocker Rejection for 5G NR," in *2022 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 65, 2022, pp. 434–436.

[19] S.C. Cripps, *RF Power Amplifiers for Wireless Communications*, 1st ed. Norwood, MA, USA: Artech House, 1999.

[20] C. Elgaard, S. Andersson, P. Caputa, E. Westesson, and H. Sjöland, "A 27 GHz Adaptive Bias Variable Gain Power Amplifier and T/R Switch in 22nm FD-SOI CMOS for 5G Antenna Arrays," in *2019 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2019, pp. 303–306.

[21] C.R. Chappidi, T. Sharma, Z. Liu, and K. Sengupta, "Load Modulated Balanced mm-Wave CMOS PA with Integrated Linearity Enhancement for 5G applications," in *2020 IEEE/MTT-S International Microwave Symposium (IMS)*, 2020, pp. 1101–1104.

[22] Y. Jin and S. Hong, "A 24-GHz CMOS Power Amplifier With Dynamic Feedback and Adaptive Bias Controls," *IEEE Microwave and Wireless Components Letters*, vol. 31, no. 2, pp. 153–156, 2021.

[23] S. Chen, G. Wang, Z. Cheng, P. Qin, and Q. Xue, "Adaptively Biased 60-GHz Doherty Power Amplifier in 65-nm CMOS," *IEEE Microwave and Wireless Components Letters*, vol. 27, no. 3, pp. 296–298, 2017.

[24] H.T. Nguyen and H. Wang, "A Coupler-Based Differential Doherty Power Amplifier with Built-In Baluns for High Mm-Wave Linear-Yet-Efficient Gbit/s Amplifications," in *2019 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2019, pp. 195–198.

[25] M. Pashaeifar, L.C.N. de Vreede, and M.S. Alavi, "A Millimeter-Wave CMOS Series-Doherty Power Amplifier With Post-Silicon Inter-Stage Passive Validation," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 10, pp. 2999–3013, 2022.

[26] H. Zhang, R.Z. Zhan, Y.C. Li, and J. Mou, "High Efficiency Doherty Power Amplifier Using Dual-Adaptive Biases," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 8, pp. 2625–2634, 2020.

[27] B. Sadhu, X. Gu, and A. Valdes-Garcia, "The more (antennas), the merrier: A survey of silicon-based mm-wave phased arrays using multi-ic scaling," *IEEE Microwave Magazine*, vol. 20, no. 12, pp. 32–50, 2019.

[28] S. Ek, T. Pählsson, C. Elgaard, A. Carlsson, A. Axholt, A.K. Stenman, L. Sundström, and H. Sjöland, "A 28-nm fd-soi 115-fs jitter pll-based lo system for 24–30-ghz sliding-if 5g transceivers," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 7, pp. 1988–2000, 2018.
[29] M. Pashaeifar, L.C.N. de Vreede, and M.S. Alavi, "14.4 A 24-to-30GHz Double-Quadrature Direct-Upconversion Transmitter with Mutual-Coupling-Resilient Series-Doherty Balanced PA for 5G MIMO Arrays," vol. 64, pp. 223–225, 2021.
[30] S. Kulkarni, D. Zhao, and P. Reynaert, "Design of an optimal layout polyphase filter for millimeter-wave quadrature lo generation," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 60, no. 4, pp. 202–206, 2013.
[31] S. Voinigescu, *High-Frequency Integrated Circuits*, 1st ed. Cambridge, UK: Cambridge University Press, 2013.
[32] B. Kang, J. Yim, T. Kim, S. Ko, W. Ko, H. Shin, I. Ryu, S.G. Yang, J.D. Bae, and H. Park, "Design and Analysis of an Ultra-Wideband Automatic Self-Calibrating Upconverter in 65-nm CMOS," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 7, pp. 2178–2191, 2012.
[33] C. Elgaard, A. Axholt, E. Westesson, and H. Sjöland, "A 26GHz 22.2DBM Variable Gain Power Amplifier in 28NM FD-SOI CMOS for 5G Antenna Arrays," in *2018 Asia-Pacific Microwave Conference (APMC)*, 2018, pp. 965–967.
[34] W.L. Chan, J.R. Long, M. Spirito, and J.J. Pekarik, "A 60GHz-band 1V 11.5dBm power amplifier with 11% PAE in 65nm CMOS," in *2009 IEEE International Solid-State Circuits Conference - Digest of Technical Papers*, 2009, pp. 380–381,381a.
[35] H. Asada, K. Matsushita, K. Bunsen, K. Okada, and A. Matsuzawa, "A 60GHz CMOS power amplifier using capacitive cross-coupling neutralization with 16 % PAE," in *2011 6th European Microwave Integrated Circuit Conference*, 2011, pp. 554–557.
[36] W. Doherty, "A New High Efficiency Power Amplifier for Modulated Waves," *Proceedings of the Institute of Radio Engineers*, vol. 24, no. 9, pp. 1163–1182, 1936.
[37] J. Vuolevi and T. Rahkonen, *Distortion in RF Power Amplifiers*. 1st ed. Norwood, MA, USA: Artech House Professional, 2003.
[38] A.F. Molisch, *Wireless Communications*, 2nd ed. New York, NY, USA: John Whiley & Sons, 2011.
[39] F. Quadrelli, D. Manente, D. Seebacher, F. Padovan, M. Bassi, A. Mazzanti, and A. Bevilacqua, "A broadband 22–31-ghz bidirectional image-reject up/down converter module in 28-nm cmos for 5g communications," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 7, pp. 1968–1981, 2022.

Christian Elgaard completed his M.Sc. degree in Engineering Physics from Lund University in 2002. Between 2003 and 2014 Christian held different positions at Ericsson, Nokia, and CSR and since 2015 he is working in the Integrated Radio Systems group, RF Frontend and PA, at Ericsson Research in Lund, Sweden, where he holds a Master Researcher position. Over the years, Christian has worked on many different analog blocks aimed for 2G - 5G radio systems, including LNAs, Mixers, VCOs, XO BB-filters, and PAs stretching from low frequencies to the mm-wave domain. Christian Elgaard is the inventor/co-inventor of more than 30 granted or pending patent applications and he has authored 5 international peer reviewed journal or conference papers. Christian is currently pursuing an industrial Ph.D. at the department of Electrical and Information Technology, Lund University with focus on mmW power amplifiers.

Mustafa Özen (S'10-M'14) received the B.Sc. degree in electrical engineering from Ankara University, Ankara, Turkey, in 2006. He received the M.Sc. degree in microwave engineering and Ph.D. degree from the Chalmers University of Technology, Göteborg, Sweden, in 2010 and 2014, respectively. He was a postdoctoral researcher at University of California, San Diego during Sep. 2016-Dec. 2017, working with CMOS power amplifiers. He is currently employed by Ericsson, Sweden as an RFIC design engineer. His research interests are high efficiency power amplifier architectures, linearization, and design of millimeter wave SiGe/Si front-ends.

Dr. Özen was the recipient of the 2011 Best Paper Award of the IEEE Wireless and Microwave Technology Conference. He was awarded international post-doctoral fellowship from Swedish Research Council in 2016.

Eric Westesson was born in Nässjö, Sweden, in 1973. He received the M.Sc. degree in electrical engineering from Lund University, Lund, Sweden, in 1998. In 1998, he joined the Department of Applied Electronics, Lund University as a Ph.D. Student, where he was involved in the linearization of RF power amplifiers. In 2001, he joined the Ericsson Mobile Platforms (later ST-Ericsson and Ericsson Modems), where he was involved design for the cellular handset industry. Since 2014, he has been with the Ericsson Radio, Lund, where he is involved in RFIC design for millimeter-wave infrastructure.

Ahmed Mahmoud received his M.Sc. degree in electronics and communication engineering from Fayoum University, Egypt, in 2011. Subsequently, he pursued his Ph.D. degree at Lund University, Lund, Sweden, completing it in 2017. Starting from 2012, he served as a Ph.D. student in the Mixed-Signal Group at the Department of Electrical and Information Technology, Lund University, where his research primarily focused on Digital Phase Locked Loop (DPLL) in CMOS technology. Currently, Ahmed is employed at Ericsson AB in Lund, Sweden, actively engaged in the field of RF ASIC design for cellular base stations.

Florent Torres received a Masters degree in Electrical Engineering from the University of Bordeaux, France in 2014. He continued his academic journey and obtained a Ph.D. degree in Electronics from the University of Bordeaux, France in 2018. During his doctoral studies, he conducted research in collaboration with the IMS Laboratory, Bordeaux, France, and STMicroelectronics, Crolles, France and focused on the design of power amplifiers for 5G applications in 28nm FD-SOI technology. Currently, he is a Senior Researcher within the RF Frontend and PA team of the Integrated Radio System research group at Ericsson Research, Lund, Sweden, where he joined in 2018. His current research interests include the design of power amplifier and frontends at RF and millimeter-wave frequencies using advanced technology nodes, device hardware research, as well as RF fingerprinting for physical layer security and trustworthiness applications.

Shakila Bint Reyaz Shakila Bint Reyaz received the M.Eng. degree in Telecommunications from NED University of Engineering and Technology, Karachi, Pakistan in 2007, and the Ph.D. degree from Uppsala University, Sweden, in 2015. Her Ph.D. thesis was focused on Reconfigurable and Wideband Receiver Components for System on Chip millimetre -wave Radiometer Front-Ends. She has worked with several EU-Projects with different types of RF circuits (switches and amplifiers etc.) using silicon and III-V based technologies. She joined Ericsson AB in 2018, where she has been working with the design of millimetre-wave transceivers in CMOS technology for cellular base stations.

Therese Forsberg Received the M.Sc. degree in electronics design engineering from Linköping University, Linköping, Sweden, in 2007, and the Ph.D. degree from Lund University, Lund, Sweden, in 2019. Her master's thesis was focused on RF coexistence issues when integrating GPS receivers in cell phones. Between 2007 and 2013, she was with Ericsson AB, Lund, where she was involved in the design and validation of analog transceiver RF application-specific integrated circuits for cell phones. From 2013, she was a Ph.D. student with the Analog RF Group, Department of Electrical and Information Technology, Lund University, Lund, where she was involved in the research of efficient millimeter-wave transmitter design in CMOS technology. Since she joined Ericsson AB in 2019, she has been working with the design of millimeter-wave transceivers in CMOS technology for cellular base stations.

Rehman Akbar (Graduate Student Member, IEEE) received the M.Sc. degree in electrical engineering from Tampere University of Technology (TUT), Tampere, Finland, in 2014. From 2012 to July 2015, he was a member of the RFIC Design Lab at TUT, where he worked on VHF range DC-DC converters for mobile devices. He is currently pursuing a Ph.D. degree in electrical engineering at the University of Oulu, Oulu, Finland. Since 2022, he has been an Experienced Researcher with the Radio Architecture and Integration Systems Group, Ericsson Research, Ericsson, Lund. His research interests include the design of wideband RF/mmWave/BB integrated circuits, primarily in CMOS technology, radio architecture, and integration for next-generation wireless systems.

Hans Hagberg received the M.Sc. degree in electronics design engineering from Lund University, Lunds Tekniska Högskola, Lund, Sweden, in 1995. Between 1995 and 2013 he was with Ericsson AB cell phone and modem business in Lund, where he was involved in design of analog transceiver RF application-specific integrated circuits, as well as the belonging digital signal processing, mainly focusing on polar transmitters and Envelope tracking schemes. From 2013 he has been working at Ericsson AB, base station business in Lund, where his main responsibility and focus has been on digital base band signal processing and algorithm design, including in particular algorithm research within digital pre-distortion and crest factor reduction techniques in digital transmitter chains.

Henrik Sjöland (Senior Member, IEEE) received the M.Sc. degree in electrical engineering from Lund University, Sweden, in 1994, and the PhD degree from the same university in 1997. In 1999 he was a postdoc at UCLA on a Fulbright scholarship. He has been an associate professor at Lund University since year 2000, and a full professor since 2008. Since 2002 he is also part time employed at Ericsson Research, where he is currently a Senior Specialist. He has authored/co-authored about 200 international peer reviewed journal and conference papers. He is also inventor/co-inventor of about 50 different inventions with granted patents, and of more than 100 inventions with patent applications pending. Prof. Henrik Sjöland has been an associate editor of IEEE Transactions on Circuits and Systems – II (2014-2015), and of IEEE Transactions on Circuits and Systems – I (2016-2023). He has been a technical program committee member of ESSCIRC (2011-2016), and currently of ISSCC (from 2024). He has successfully been the main supervisor of 14 PhD students to receive their degrees. His research interests include design of radio frequency, microwave, and mm wave integrated circuits and systems, primarily in CMOS technology.

Paper IV

Paper IV

Analysis and Design of a GHz Bandwidth Adaptive Bias Circuit for an mmW Doherty Amplifier

C. Elgaard and H. Sjöland, "Analysis and Design of a GHz Bandwidth Adaptive Bias Circuit for an mmW Doherty Amplifier," IEEE Transactions on Microwave Theory and Techniques, Regular Papers, Submitted Feb. 2024.

Analysis and Design of a GHz Bandwidth Adaptive Bias Circuit for an mmW Doherty Amplifier

Christian Elgaard[1,2], Henrik Sjöland[1,2]
[1]Department of Electrical and Information Technology, Lund University, Sweden
[2]Ericsson Research, Ericsson AB, Lund, Sweden; {christian.elgaard, henrik.sjoland}@ericsson.com

***Abstract*—This paper derives theoretical results for adaptive bias in Doherty amplifiers and presents the design and measurements of an integrated adaptive bias circuit tailored for high peak-to-average high bandwidth signals. Fundamental equations for output power, impedance, and efficiency of the complete Doherty amplifier are derived. Even with ideal transistor models, the Doherty amplifier is fundamentally nonlinear due to saturation of the main amplifier and class-C nonlinearity of the auxiliary. Increasing the transconductance of the auxiliary amplifier mitigates the distortion. Adaptive bias offers the possibility to control the output current characteristic of the auxiliary amplifier. This means that adaptive bias linearises and mitigates the need for an oversized auxiliary amplifier. Both methods, transconductance scaling and adaptive bias, are analysed and compared as well as having a band limited adaptive bias signal. The design of a multiple GHz bandwidth adaptive bias circuit is presented. To verify the circuit design and the theoretical predictions, an mmW Doherty amplifier in 22 nm CMOS-FD-SOI, utilizing the presented adaptive bias circuit, is measured and compared with and without adaptive bias. Comparison is conducted both using continuous-wave and modulated high bandwidth signals. Measured results confirm the predicted improvements by the adaptive bias as derived by the theoretical analysis.**

***Index Terms*— Adaptive bias, auxiliary amplifier bias, Doherty amplifier, efficiency linearity trade-off, millimeter-wave (mmW).**

I. Introduction

THE Doherty amplifier, shown in Fig. 1, is the most well known efficient power amplifier (PA) and an absolutely brilliant construction. Even though it is close to 100 years since its invention, first publication, and patent in 1936 by Willliam H. Doherty [1], [2], it is still highly present in today's wireless systems, and in the research for tomorrow's. However, with the exception of two recently published papers [3] and [4], the published 5G mmW transceivers during recent years all focus on class AB PAs to support the tough linearity requirements that follow transmissions of wideband, high peak to average (PAR), orthogonal frequency division multiplexing (OFDM) signals with high order modulations [5]–[20]. A growing interest for deploying adaptive bias, i.e. adjusting the bias of the power amplifier so that it follows the envelope of the modulated signal, is seen among mmW PA publications. Unfortunately, with the execption of [3], none of the PAs integrated in mmW frontends make use of the technique, but a few published standalone PAs do, and of these, some are class AB PAs [21]–[23] and some are Doherty amplifiers [24]–[29]. Gains brought by the adaptive bias are typically only evaluated for sinusoidal stimuli's, and [25], [28], [29] which reports good results for modulated signals in combination with adaptive bias, do not provide any evaluation or comparison to estimate the performance with and without the adaptive bias.

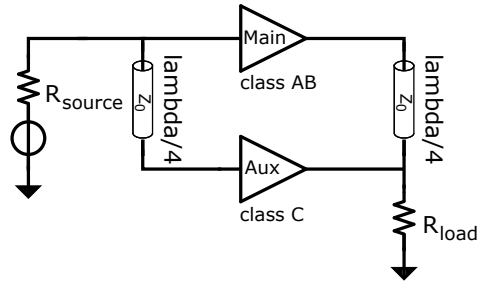

Fig. 1. A current combined Doherty PA

Since the adaptive bias actively changes the operating point of the amplifying transistor(s) in the PA, it will affect both the linearity and the power consumption. Therefore, deploying adaptive bias on a PA operating on wideband modulated signals requires careful analysis. For instance, the adaptive bias signal must be fast enough to follow most of the envelope of the signal, and it should also provide an appropriate amount of bias increase as the input signal amplitude increases. Published amplifiers utilizing adaptive bias are typically lacking all information about bandwidth considerations/requirements. One work [28] however, does mention that the adaptive bias circuit has GHz bandwidth, but no more details or analysis are provided about how this is accomplished or what the trade-offs are.

The efficiency of the Doherty amplifier is boosted by using a class C amplifier together with a class AB, from now on referred to as the auxiliary and main amplifier. Traditionally, the conduction angle of the class C amplifier is determined by its static bias level and the input signal amplitude. Another possibility to control the conduction angle is to use adaptive bias, where the bias level depends on the input signal amplitude. This leads to some fundamental research questions. It is well known that adaptive bias can be used to linearise a single transistor stage, but what can be achieved when it comes to adaptive bias for Doherty amplifiers? Furthermore, can the Doherty amplifier benefit from adaptive bias even when using ideal linear transistor models?

To summarize, a growing interest for adaptive bias, and in particular in combination with Doherty amplifiers is observed. However, there is a lack of both an explaining theory for why and how the adaptive bias improves the Doherty amplifier, and of measurements comparing performance with and without adaptive bias. Theoretical analysis regarding the fundamental nonlinearities in a Doherty amplifiers, due to lack of output current from the class C biased auxiliary amplifier have previously been presented, for instance in [30], [31], which however do not link this built in imperfection of the Doherty amplifier to the mitigation opportunity offered by adaptive bias.

This paper has five main objectives. First, to explain through analytical derivation why a Doherty amplifier with ideal linear transistor models still produces a nonlinear output. Second, to investigate how adaptive bias can be used to linearise the system. Third, to quantify the effect of different adaptive bias non-idealities on the linearity of the Doherty amplifier. Fourth, present the design and verification of a integrated highly configurable adaptive bias circuit for an mmW Doherty PA amplifying high bandwidth high PAR OFDM signals. Fifth, to quantify benefits brought by deploying the adaptive bias on the auxiliary amplifier, by comparing the performance with that of having a static auxiliary amplifier bias, on a measured 5G mmW PA that is integrated in a transceiver frond-end. In addition to the five main objectives, the paper also analyses the efficiency, load impedance, auxiliary amplifier conduction angle, and amplifier output currents, with and without adaptive bias.

The paper is organized as follows: A general circuit analysis of the Doherty amplifier is conducted in Section II. How to select transconductance ($\mathrm{g_m}$) ratio between main and auxiliary amplifier is derived in Section III, and the deviation from ideal output current of the auxiliary amplifier is analysed in Section IV. A method to provide ideal conduction angle for the auxiliary amplifier is proposed in Section V. In Section VI and VII the efficiency and distortion in a Doherty amplifier are analysed with ideal and non-ideal adaptive bias. A highly integrated 5G mmW transceiver front-end, with a Doherty PA, and the design of the adaptive bias circuit capable of biasing the auxiliary amplifier either in a static class C or in adaptive bias mode is presented in Section VIII. The measurement results of this are presented, analysed, and compared with the theory in Section IX. Finally, conclusions are drawn in Section X.

II. Circuit Analysis of Doherty Amplifier

In the schematic of Fig. 2, the main amplifier is represented by a first current source and the auxiliary amplifier by a second current source, proportional to the first but phase shifted by 90-degrees, due to the delayed input signal of the auxiliary amplifier.

The main amplifier current is labelled $\mathrm{I_1}$, the auxiliary current $\mathrm{I_2}$, and $\mathrm{I_3}$ is the current that comes out from the $\lambda/4$-transmission line. The relation between the three currents are defined as

$$I_2 = cI_3 = -jdI_1 \tag{1}$$

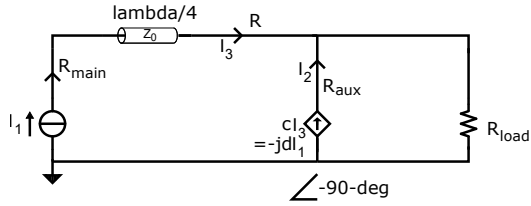

Fig. 2. The output network of a Doherty amplifier. Main and auxiliary amplifiers are represented by ideal current sources.

Now the goal is to express the impedances at the output of the main and auxiliary amplifiers as a function of d and the system load impedance $\mathrm{R_{load}}$. The impedance at the output of the transmission line R is load modulated by $\mathrm{I_2}$, which gives.

$$R = R_{load}(1+c) \tag{2}$$

Similarly, the impedance $\mathrm{R_{aux}}$ seen by $\mathrm{I_2}$ is load modulated by $\mathrm{I_3}$, which gives

$$R_{aux} = R_{load}(1+c)/c \tag{3}$$

The relation between impedance $\mathrm{R_{main}}$ and R is given by the transmission line impedance transformation

$$R_{main} = \frac{{Z_0}^2}{R} = \frac{{Z_0}^2}{R_{load}(1+c)} \tag{4}$$

And since we assume an ideal lossless transmission line, the power from $\mathrm{I_1}$ must be the same on both sides of the transmission line, which gives

$$I_3 = I_1\sqrt{\frac{R_{main}}{R}} = I_1\frac{Z_0}{R_{load}(1+c)} \tag{5}$$

Combining (5) with (1) gives

$$c = \frac{d}{\frac{Z_0}{R_{load}} - d} \tag{6}$$

Using (6) with (2), (3), and (4) gives the sought load impedances expressed in d and the system load impedance $\mathrm{R_{load}}$

$$R = R_{load}(1 + \frac{d}{\frac{Z_0}{R_{load}} - d}) \tag{7}$$

$$R_{aux} = R_{load}(1 + \frac{d}{\frac{Z_0}{R_{load}} - d})/\frac{d}{\frac{Z_0}{R_{load}} - d} \tag{8}$$

$$R_{main} = \frac{{Z_0}^2}{R_{load}(1 + \frac{d}{\frac{Z_0}{R_{load}} - d})} \tag{9}$$

The main amplifier provides current at all signal levels, but the auxiliary amplifier only when the signal level is high. This can be captured by making the variable d, as used in the results above, a function of the input signal level $\mathrm{d = f(v_{in})}$, and set equal to zero below a threshold point, called back-off. For a symmetrical Doherty amplifier, the back-off level is located

6 dB below the maximum output power, below that point d is equal to zero, and above it d increases, and at maximum output power d is equal to 1. It is clear that the main amplifier should have a linear gain, but it not as obvious what the gain should be for the auxiliary amplifier. Naturally, the target is to achieve a linear gain of the combined system up to saturated output power. For the symmetrical Doherty the main amplifier is designed to reach its maximum voltage amplitude at 6 dB back-off. Assuming a linear gain for the output current from the main amplifier, its power contribution will still continue to increase beyond this point as its load impedance is gradually reduced for output power levels above 6 dB back-off. However, the power delivered will no longer be linearly dependet on the input signal, instead the increase rate of the output power from the main amplifier will be lower and limited by the rate at which its load impedance is reduced. To achieve linear gain the power delivered by the auxiliary amplifier should exactly fill this gap, so that the combined output power becomes a linear amplification of the input signal. The target of the following calculations is to analytically derive how the ratio, labelled d in the analysis above, between the transconductance of the main and auxiliary amplifiers should then depend on the input signal level. For this case the transconductance is assumed to be the transconductance of the fundamental harmonic, h_1. The output power from the main amplifier is then directly given by (10), assuming the voltage amplitude not being saturated.

$$P_{main} = (v_{in} \cdot gm_{main_{h1}})^2 \frac{{Z_0}^2}{R_{load}(1 + \frac{d}{\frac{Z_0}{R_{load}} - d})} \tag{10}$$

Having $\mathrm{d} = 0$ gives a linear output power

$$P_{out} = (v_{in} \cdot gm_{main_{h1}})^2 \frac{{Z_0}^2}{R_{load}} \tag{11}$$

The output power from the auxiliary amplifier is

$$P_{aux} = (v_{in} \cdot gm_{aux_{h1}})^2 R_{load}(1 + \frac{d}{\frac{Z_0}{R_{load}} - d})/\frac{d}{\frac{Z_0}{R_{load}} - d} \tag{12}$$

Combining (10) and (12) and using $\mathrm{c} = \mathrm{d}/(\mathrm{Z_0}/\mathrm{R_{load}} - \mathrm{d})$ from (6)

$$P_{main} + P_{aux} = v_{in}^2 (gm_{main_{h1}}^2 \frac{{Z_0}^2}{R_{load}(1+c)} + gm_{aux_{h1}}^2 R_{load}(\frac{1}{c} + 1)) \tag{13}$$

For the case of a symmetrical Doherty, $\mathrm{Z_0} = 2\mathrm{R_{load}}$ and $\mathrm{gm_{aux_{h1}}} = \mathrm{d} \cdot \mathrm{gm_{main_{h1}}}$, which gives

$$P_{main} + P_{aux} = (v_{in} \cdot gm_{main_{h1}})^2 R_{load}(\frac{4}{(1+c)} + d^2(\frac{1}{c} + 1)) \tag{14}$$

Now using (6) but expressing d in c, gives

$$P_{main} + P_{aux} = (v_{in} \cdot gm_{main_{h1}})^2 R_{load}(\frac{4}{(1+c)} + (\frac{Z_0}{R_{load}} \frac{c}{1+c})^2(\frac{1}{c} + 1)) \tag{15}$$

Which after some simplification becomes

$$P_{main} + P_{aux} = (v_{in} \cdot gm_{main_{h1}})^2 \frac{Z_0^2}{R_{load}}(\frac{1}{(1+c)} + (\frac{c}{1+c})^2(\frac{1}{c} + 1)) \tag{16}$$

After further simplifications it becomes

$$P_{main} + P_{aux} = (v_{in} \cdot gm_{main_{h1}})^2 \frac{Z_0^2}{R_{load}} \tag{17}$$

Which is equal to the linear output power $\mathrm{P_{out}}$ from (11). And since the output power is independent of d, the results proves that any ratio d between main and auxiliary will provide a linear Doherty. However, this is only true as long as the output of the main amplifier does not compress in voltage, and not any ratio guarantees that. The question then becomes, how should d gradually be increased to achieve a linear characteristic, combined with a hard maximum output voltage level of the main amplifier? For high efficiency, the voltage amplitude of the main amplifier should also be constant and close to the maximum for signal levels above back-off. This means that the load impedance of the main amplifier should be inversely proportional to the input voltage above back-off. Such a requirement could be expressed as

$$v_{in_{6dB_{BO}}} \cdot gm_{main_{h1}} \frac{Z_0^2}{R_{load}} = v_{in} \cdot gm_{main_{h1}} R_{main} \tag{18}$$

$$v_{in_{6dB_{BO}}} \cdot gm_{main_{h1}} \frac{Z_0^2}{R_{load}} = v_{in} \cdot gm_{main_{h1}} \frac{{Z_0}^2}{R_{load}(1+c)} \tag{19}$$

$$c = \frac{v_{in} - v_{in_{6dB_{BO}}}}{v_{in_{6dB_{BO}}}} \tag{20}$$

and again using (6) gives

$$d = \frac{Z_0}{R_{load}} \frac{v_{in} - v_{in_{6dB_{BO}}}}{v_{in}} \tag{21}$$

For the symmetrical Doherty, where $\mathrm{Z_0} = 2\mathrm{R_{load}}$, $\mathrm{I_2}$ becomes

$$I_2 = 2gm_{main_{h1}}(v_{in} - v_{in_{6dB_{BO}}})\Big|_{v_{in} > v_{in_{6dB_{BO}}}} \tag{22}$$

As indicated by (22), after the auxiliary amplifier kicks in its output current $\mathrm{I_2}$ should increase linearly with the input signal level, but with twice the slope of the main amplifier in a symmetric Doherty amplifier. This guarantees that the main amplifier output voltage stays constant at increased input signal levels.

III. Sizing of Main and Auxiliary Amplifiers

In this section the transconductance g_m ratio of the main and auxiliary amplifiers is determined for ideal Doherty operation. To facilitate the discussion, Fig. 3 illustrates some key measures of the auxiliary amplifier and its signals, that is the gate-to-source voltage, drain current, threshold voltage V_{th}, max drain current I_{max}, conduction angle 2Φ, overdrive voltage v_{od}, and max input voltage v_{in}. In this paper, an

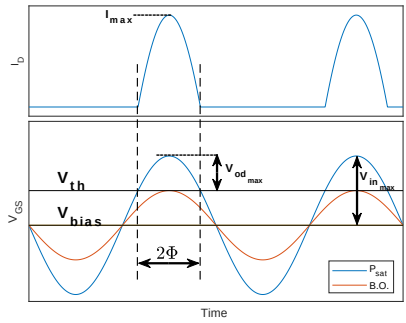

Fig. 3. Signals in the auxiliary amplifier, showing threshold voltage, maximum drain current, conduction angle, maximum overdrive voltage, and maximum input voltage.

ideal transistor model is used, and the reason for that is to separate effects originating from transistor imperfections from fundamental behaviour of the Doherty amplifier, shown also with ideal transistors. The ideal transistor model used is a voltage controlled current source with output current equal to $v_{ov}g_m$, where the overdrive voltage $v_{ov} = 0$ for $v_{in} \leq V_{th}$ and $v_{ov} = v_{in} - V_{th}$ for $v_{in} > V_{th}$, as long as the drain voltage is below maximum voltage. When output current is large enough to produce voltage larger than the maximum voltage, the output current is reduced so that the drain voltage reaches the maximum voltage level. For an amplifier modelled like this, I_{max} is equal to the maximum overdrive voltage of the input signal multiplied by the transconductance g_m. The bias level of the auxiliary amplifier is below the threshold voltage and controls when it starts to conduct current, which is set to occur at 6 dB back-off input signal amplitude, giving

$$I_{max} = g_m(v_{in_{max}} - v_{in_{6dB_{BO}}}) = g_m v_{od_{max}} \tag{23}$$

At max input signal, the conduction angle in figure 3 is equal to $2\pi/3$, and the auxiliary amplifier should then together with the main amplifier drive the load with full voltage swing, which will provide a value for the required transconductance g_m and in turn give the maximum drain current I_{max}. Furthermore, at max output power, main and auxiliary amplifiers should contribute with equal amount of fundamental signal current to the load. The fundamental current as a function of the conduction angle and I_{max} is found in [32] and repeated in (24).

$$I_{fund} = \frac{I_{max}}{2\pi}\frac{2\Phi - sin\,2\Phi}{1 - cos\,\Phi} \tag{24}$$

At maximum input signal both main and auxiliary amplifier should drive a signal current equal to $V_{dd}/2R_{load}$, obtained from (4) and (3) with $c = 1$. This gives a closed expression for the required g_m by combining (23) with (24) and solving for g_m.

$$g_m = \frac{1 - cos\,\Phi}{2\Phi - sin\,2\Phi}\frac{2\pi}{v_{od_{max}}}\frac{V_{dd}}{2R_{load}} \tag{25}$$

For the class B biased main amplifier $v_{od_{max}} = V_{in_{max}}$ and thus, with the conduction angle $2\Phi = \pi$ it becomes

$$g_{m_{main}} = \frac{2}{v_{in_{max}}}\frac{V_{dd}}{2R_{load}} \tag{26}$$

For the class C biased auxiliary amplifier $v_{od_{max}} = V_{in_{max}}/2$ and thus, with the conduction angle $2\Phi = 2\pi/3$ it becomes

$$g_{m_{aux}} = \frac{6\pi}{4\pi - 3\sqrt{3}}\frac{2}{v_{in_{max}}}\frac{V_{dd}}{2R_{load}} \tag{27}$$

The ratio γ of transconductance for the main and the auxiliary amplifiers for conduction angle $2\pi/3$ is calculated in (28) and plotted for various conduction angles in Fig. 4.

$$\gamma = \frac{g_{m_{aux}}}{g_{m_{main}}} = \frac{6\pi}{4\pi - 3\sqrt{3}} = 2.56 \tag{28}$$

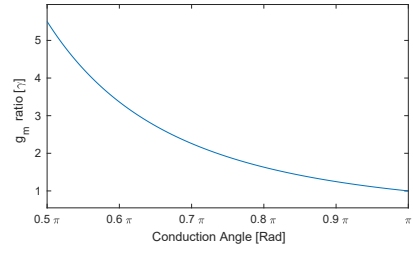

Fig. 4. Required g_m ratio in symmetric Doherty of main class B PA and class C biased auxiliary amplifier, versus maximum conduction angle.

It should be noted that in a traditional Doherty amplifier with fixed bias, the maximum conduction angle cannot be set freely, but be determined by the back-off level. Setting the class C amplifier to turn on at a back-off level of 6 dB will result in a maximum conduction angle equal to $2\pi/3$.

IV. Auxiliary Amplifier Output Current Deviation from Ideal Performance

The conduction angle of the class C amplifier is zero below back-off and above back-off it is $2\Phi = 2 \cdot \arccos(v_{in_{BO}}/v_{in})$. A linearly increasing output current with signal amplitude is required from the auxiliary amplifier for an overall linear system and optimal Doherty efficiency. The fundamental output current from the auxiliary amplifier is given by (29).

$$I_{aux} = I_2 = \frac{g_{m_{aux}} v_{od}}{2\pi} \frac{2\Phi - sin\, 2\Phi}{1 - cos\, \Phi} \tag{29}$$

The first part of the expansion has the desired linear behaviour, with I_{aux} being proportional to overdrive voltage. However, the second part introduces non-linearity, due to the conduction angle being dependent on the overdrive voltage, see Fig. 5.

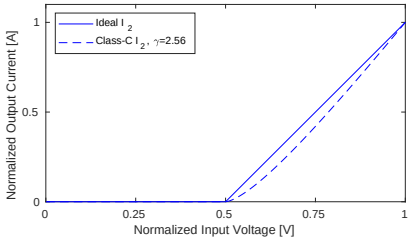

Fig. 5. Solid line -ideal linear auxiliary output current I_2. Dashed line -output current as obtained by a class C amplifier biased at 6 dB back-off with size ratio $\gamma = 2.56$

Studying Fig. 5 it becomes evident that a challenging part of the design of a Doherty amplifier is to control the output current from the auxiliary amplifier. Simply applying a bias level and an input signal that turns on the auxiliary amplifier at the back-off level will produce an output current that differs significantly from the wanted ideal case. Since the dashed curve is below the solid curve the main amplifier will go into voltage compression and the output power delivered by both the main and auxiliary amplifiers will be reduced. Naturally this results in nonlinear distortion since the output signal will deviate from linear amplification. If we for some reason would have a region where the dashed curve would be above the solid curve, for instance if we choose to adjust the bias so that the input signal turns on the auxiliary amplifier at a lower input signal level, i.e. before the main amplifier goes into compression, the load modulation would instead lower the impedance at the output of the main amplifier too much, resulting in reduced efficiency. That case would, however, still produce a linear output, since the reduction of the output power delivered from the main amplifier would be filled up exactly by the increase in power from the auxiliary amplifier, as derived in (17).

V. Adaptive Bias

A method to produce the ideal conduction angle and overdrive voltage for I_2 to satisfy (22), is to apply a signal dependent bias level, typically called adaptive bias. In Fig. 5 the difference between ideal I_2 and the I_2 obtained directly from the class C biased auxiliary amplifier does not seem that large. In a sense, this is true, and the deviation, as will be shown later, is of minor minor concern for this case. However, this is for an auxiliary/main ratio of $\gamma = 2.56$, which can be considered high. For implementation purposes, at least for high frequency applications, such a large scaling ratio is limiting and troublesome. Parasitic capacitances are increased, chip area increases, bandwidth is reduced, and efficiency drops due to increased losses, and so on. Some of these drawbacks could perhaps be acceptable if the increased size would result in an increased output power. Unfortunately, however, the auxiliary amplifier only delivers the same amount of output power as the 2.56 times smaller main amplifier. Using a reduced scaling ratio γ with fixed auxiliary bias, the bias level must be increased and thereby the auxiliary amplifier turns on earlier, which limits the efficiency boost and the whole reason for using a Doherty amplifier. Also for this problem, the adaptive bias offers a solution. It is then possible to freely choose the back-off level with a smaller γ, but still delivering the I_2 current needed for linear output power. To investigate the effect of the adaptive bias, the equations for impedance levels and output powers were modified taking a dynamic bias level into account, changing both I_{max} and the conduction angle. Using the equations in Section II combined with the the possibility to freely choose the overdrive voltage for the auxiliary amplifier via the use of adaptive bias, plots for output currents, powers, and voltages using size ratio $\gamma = 2.56$ and $\gamma = 1$ with and without ideal adaptive bias, are shown in Fig. 6 below. As can be seen, solid curves, for ideal adaptive bias, give ideal Doherty amplifier behaviour, regardless of size ratio. Without adaptive bias, it is clear that using a large sized auxiliary amplifier $\gamma = 2.56$ compared to $\gamma = 1$ mitigates the deviation from ideal Doherty behaviour. For instance, the saturated power for $\gamma = 1$ is reduced by more than 3 dB, but for $\gamma = 2.56$ the saturated output power is equal to the ideal Doherty case. However, this disregards all practical difficulties with having an oversized auxiliary amplifier.

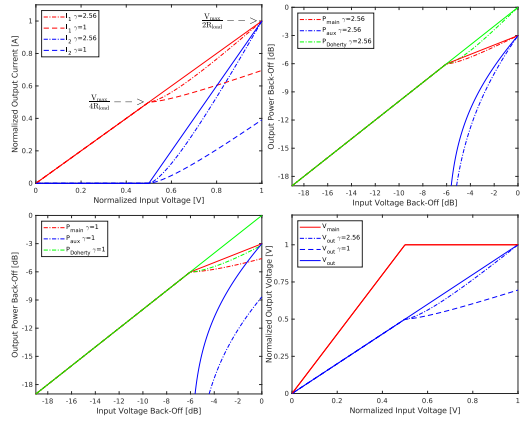

Fig. 6. Dashed curves: no adaptive bias. Solid curves: ideal adaptive bias. Upper Left: Currents I_1 and I_2 for $\gamma = 2.56$ and $\gamma = 1$. Upper Right: Output powers for $\gamma = 2.56$. Output powers. Lower Left: Output powers for $\gamma = 1$. Lower Right: Voltage levels for $\gamma = 2.56$ and $\gamma = 1$.

VI. Efficiency of Doherty Amplifier

Ignoring losses in the output combination network, the efficiency of a Doherty amplifier only depends on the efficiency

of the main and auxiliary amplifiers. For output powers below back-off, the auxiliary amplifier is turned off and the efficiency is simply determined by the main amplifier, which here is assumed to be biased in class B. For output power above back-off, the main amplifier is load modulated by the auxiliary so that it has constant and maximum output voltage swing, which means that it operates at maximum class B efficiency of $\pi/4 = 78.5\%$. The auxiliary amplifier is biased in class C and its efficiency is determined by its conduction angle [32]. For conduction angle $2\pi/3$ and with full output voltage swing the efficiency of the auxiliary amplifier becomes 89.7%. When backing down the output signal from the auxiliary PA its efficiency will scale linearly with the ratio of the output voltage and the supply voltage as expressed in (30).

$$\eta = \frac{2\Phi - sin\,2\Phi}{4(sin\,\Phi - \Phi cos\,\Phi)} \frac{v_{aux}}{v_{dd}} \tag{30}$$

For a linear Doherty amplifier the output voltage is proportional to the input voltage, see Fig. 6 lower right. Even so, the conduction angle at back-off is not straight forward, since the adaptive bias was used to adjust the conduction angle for linear output. It is therefore necessary to find the conduction angle of the auxiliary amplifier, for ideal Doherty performance, in the region between back-off and maximum output power. The power consumption is obtained from [32], combined with (23) for $\mathrm{I_{max}}$

$$P_{dc} = I_{dc}V_{dd} = \frac{g_{m_{aux}} v_{od}}{2\pi} \frac{2sin\,\Phi - 2\Phi cos\,\Phi}{1 - cos\,\Phi} V_{dd} \tag{31}$$

And the power in the fundamental harmonic is acquired from (24), combined with the load impedance of the auxiliary amplifier, and again (23) for $\mathrm{I_{max}}$.

$$P_{aux} = \frac{I_{fund}^2 R_{Load}}{2} = (\frac{g_{m_{aux}} v_{od}}{2\pi} \frac{2\Phi - sin\,2\Phi}{1 - cos\,\Phi})^2 \frac{R_{aux}}{2} \tag{32}$$

Which would indicate that the auxiliary output power is a 2-dimensional function depending on both conduction angle and input voltage. By introducing the adaptive bias we can control the conduction angle. However, for any output power that the auxiliary amplifier produces at a given input voltage, only one conduction angle exists, or the other way around. The adaptive bias will alter both conduction angle and overdrive voltage at the same time, not independently. Fig. 7 shows the efficiency of the Doherty amplifier for auxiliary/main size ratios $\gamma = 2.56$ and $\gamma = 1$ with and without adaptive bias.

VII. Distortion in Doherty Amplifier

A. Amplitude Distortion for Single Tone Input Signal

Without the use of adaptive bias to produce the desired $\mathrm{I_2}$ current from the auxiliary amplifier, the Doherty amplifier will, even with ideal transistor models, deviate from linear amplification. On a theoretical level the distortion can be largely mitigated by the method of increasing the size of the auxiliary amplifier versus the main amplifier. Since the analysis carried out in this paper uses ideal transistor models,

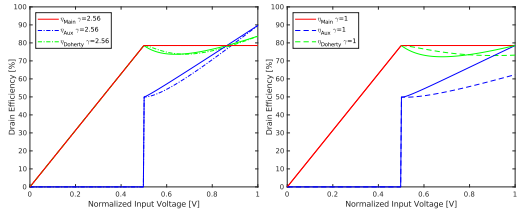

Fig. 7. Dashed curves: no adaptive bias. Solid curves: ideal adaptive bias. Left: Efficiency for main, auxiliary and combined Doherty amplifier $\gamma = 2.56$. Right: Efficiency for main, auxiliary and combined Doherty amplifier $\gamma = 1$.

no phase distortion will be generated and only amplitude-to-amplitude (AM-AM) distortion is considered. Fig. 8 shows AM-AM, which is directly derived from the output power of Fig. 6 upper right and lower left, for two different γ-ratios with and without ideal adaptive bias. As expected, the use of ideal

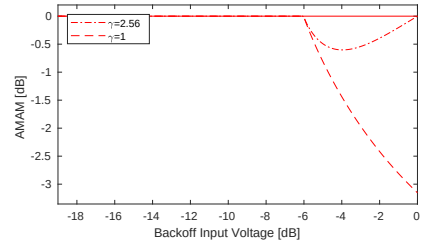

Fig. 8. AM-AM for $\gamma = 2.56$ and $\gamma = 1$. Dashed curves: no adaptive bias. Solid curves: ideal adaptive bias.

adaptive bias provides a linear output. It is also evident that a large size ratio reduces the AM-AM, keeping compression below 0.6 dB, for $\gamma = 2.56$, whereas for equal size main and auxiliary ($\gamma = 1$) the compression exceeds 3 dB.

B. Ideal Adaptive Bias Signal for Modulated Inputs

The source of the adaptive bias signal is the envelope of the modulated signal, which is equal to $\sqrt{(\mathrm{I^2 + Q^2})}$, where I and Q are the two base-band signals representing the real and imaginary part of the complex modulated signal. Each of the I and Q signals has a bandwidth that is half of the instantaneous bandwidth at the carrier frequency (RF-IBW). The envelope has bandwidth expansion due to the nonlinear function from IQ-data to envelope. In addition to this, the adaptive bias signal is a nonlinear function of the envelope, which is shown for the two cases $\gamma = 2.56$ and $\gamma = 1$ in Fig. 9. To provide a time domain example of an adaptive bias signal, used to linearise equal sized main and auxiliary amplifiers, i.e. $\gamma = 1$, a short segment of the envelope and corresponding adaptive bias signal for a 7.5 dB peak-to-average (PAR) OFDM signal is depicted in Fig. 10. As can be seen some sharp, i.e. high frequency, corners are present in the adaptive bias signal, and a fast Fourier transform (FFT) was used to investigate

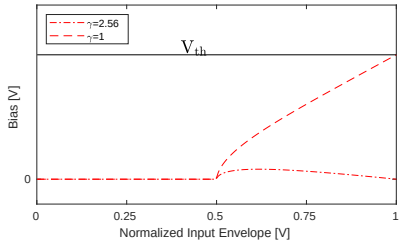

Fig. 9. Ideal adaptive bias signal for $\gamma = 2.56$ and $\gamma = 1$. Peak value of V_{th} for envelope equal to 1 comes from requirement of conduction angle equal to $2\Phi = \pi$ to match the output power from the main amplifier.

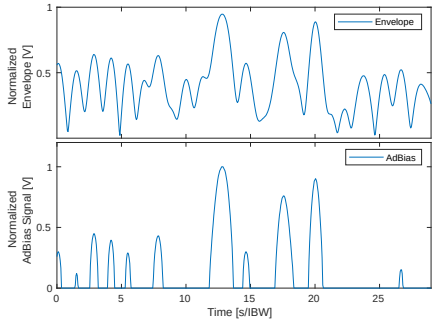

Fig. 10. Upper: Short segment of time domain envelope of OFDM signal. Lower: Corresponding ideal adaptive bias signal with $\gamma = 1$.

the frequency content of the adaptive bias signal, see Fig. 11. As can be seen energy spreads over wider bandwidth than

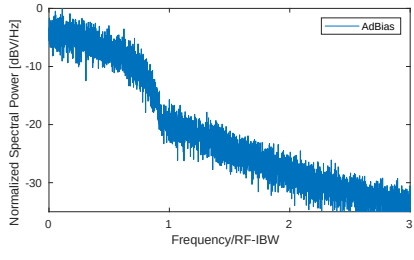

Fig. 11. Frequency content of ideal adaptive bias signal for $\gamma = 1$

occupied by the original I-Q signal, but most energy is still below the RF-IBW of the output signal.

C. Amplitude Distortion With Modulated Non-ideal Adaptive Bias

Without adaptive bias the Doherty amplifier suffers from amplitude distortion due to the non-ideal output current I_2 from the auxiliary amplifier, leading to compression in the main amplifier regardless of size ratio γ. But the problem is much more prominent for low size ratios, which could already be seen in Fig. 8 for single-tone AM-AM distortion. A measure for total compressive distortion (TCD), being very similar, but not identical, to error vector magnitude (EVM), is defined in (33) to evaluate the impact the compressive behaviour has on a modulated signal.

$$TCD = \frac{(V_{out} - G_{lin}V_{in})^2}{G_{lin}^2 V_{in}^2} \tag{33}$$

Where G_{lin} is the linear non compressive gain, i.e. the small signal gain. TCD should be interpreted as the (compressive) power of the deviation from the wanted linear signal, normalized to the power of the wanted linear signal. Differences between TCD and EVM is that EVM is only calculated at the symbol points, and with a linear gain adjustment that minimizes EVM. EVM is also limited to inband distortion, whereas TCD also takes out of band distortion into account. In this context where we want to estimate how much imperfections a PA adds to a signal, the TCD is thus better suited. Fig. 12 shows how TCD depends on the size ratio γ without any adaptive bias. The expected result with a more linear output signal for higher γ is obtained. In addition, as the PA compresses more at large input signals, the average output power (P_{avg}) also decreases as γ is reduced. However, since the high peaks in the OFDM signal occur quite rarely the effect on the average output power is rather limited.

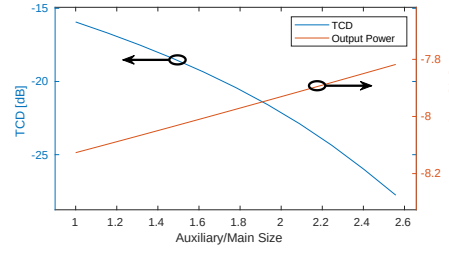

Fig. 12. TCD and P_{avg} versus auxiliary/main ratio for an OFDM modulated signal with 7.5 dB PAR.

Adding adaptive bias can mitigate the amplitude distortion, but the effect will depend on bandwidth limitations. At the very least, it will suffer from a $1^{st} - order$ low-pass filtering by the output impedance of the adaptive bias generation circuit and the capacitive load it must drive. A popular technique to reduce the input capacitance and improve the reverse isolation, at the carrier frequency, of a differential amplifier, is to use neutralizing capacitors to cancel out C_{gd} of the common source (CS) transistors, as depicted in Fig. 13. For common mode signals, however, like the adaptive bias, the input capacitance is increased by the neutralizing capacitors rather than decreased. Ignoring all capacitors, except for C_{gs} and C_{gd}, the common mode input capacitance can be expressed as

$$C_{input_{CM}} = 2C_{gs} + 4C_{gd}(1 - G_{BB}) \tag{34}$$

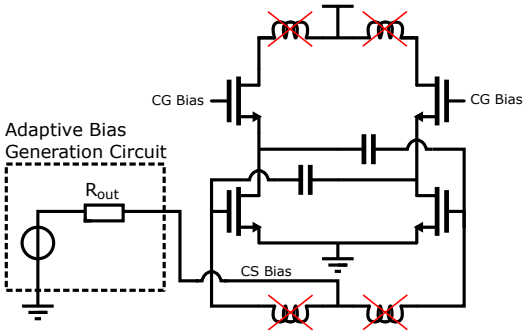

Fig. 13. An adaptive bias generation circuit with an output impedance drives the bias of a differential PA using neutralizing capacitors. Red crosses over the inductors symbolizes that the inductors present very small impedance at base-band frequencies

G_{BB} is the gain at base-band frequencies, here estimated to $G_{BB} \approx -g_{m_{CS}}/g_{m_{CG}} = -1$. This gives an input capacitance of $C_{input_{CM}} = 2C_{gs} + 8C_{gd}$, which results in a rather large capacitive load for the adaptive bias circuit. A low bandwidth adaptive bias signal will make the Doherty amplifier deviate from linear amplification and thereby create distortion. To assess the severity of the problem, an ideal adaptive bias signal for linear amplification of a high PAPR modulated signal, for various $\gamma - \text{ratios}$, was filtered through $1^{st} - \text{order}$ low-pass filters with different bandwidths. Fig. 14 shows the time domain adaptive bias signal from lower Fig. 10 compared to its filtered version. This is a rather realistic filter scenario

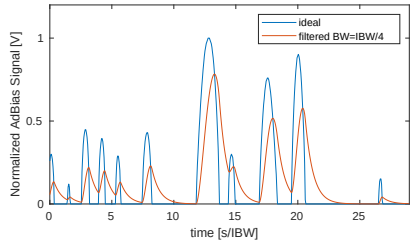

Fig. 14. Ideal adaptive bias signal and filtered adaptive bias signal, both for $\gamma = 1$

considering that an adaptive bias circuit will have an output impedance driving a capacitive load. The actual bandwidth (BW) of the low pass filter depends on design parameters like amount of current spent in the adaptive bias generation circuit and the size and design of the PA affecting the capacitive load. Incorrect bias for the auxiliary amplifier, which is the outcome of non-ideal adaptive bias, can result in two things. Firstly, when the adaptive bias level is lower than it should be, the Doherty output power compresses. Secondly, when the adaptive bias level is higher than it should be, it load modulates the main amplifier more than necessary causing loss in efficiency. Fig. 15 shows how TCD increases with reduced low-pass filter cut-off frequency.

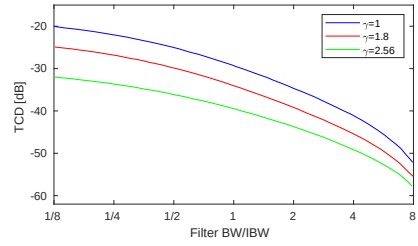

Fig. 15. TCD approaches zero for wide bandwith adaptive bias.

VIII. Adaptive Bias Test Circuit

A. Adaptive Bias Test Circuit System Architecture

To validate the theory, a Doherty amplifier integrated in a 5G mmW transceiver circuit in a 22 nm FD-SOI CMOS process was designed, fabricated, and measured. The details about the circuit, measurement setup, and measurement results can be found in [3]. For convenience, the circuit architecture, die photo, and PCB photo are reproduced here, see Fig 16. The transmitter uses two IQ-mixers for frequency upconversion and generation of the required 90-degree phase shift between the input signals to the main and auxiliary amplifiers. To boost the gain, both input signals are then amplified by pre-power amplifiers (PPAs). A current combined Doherty PA, consisting of a main and auxiliary amplifier and an output combining network, drives a 50 Ω load via a balun, which provides impedance transformation and differential to single ended signal conversion. A transmit receive switch (TRX-switch) is also included in the architecture.

B. Adaptive Bias Test Circuit Design

The Doherty amplifier has identical size (g_m) of the main and auxiliary amplifiers, i.e. $\gamma = 1$, and a highly programmable adaptive bias circuit is designed and connected to the gate of the CS transistor of the auxiliary amplifier. The adaptive bias circuit is designed to fulfill the four requirements needed to effectively linearise the Doherty amplifier as stated in [3]. These four requirements are, 1) control of the DC level, i.e. the small-signal operating point of the input transistors, 2) effective control of the starting point for the bias level increase w.r.t. the input signal level, 3) control the increase rate after the starting point, and 4) ability to drive the output load, in this case the common mode input impedance of the CS-stage of the auxiliary amplifier, such that the bias signal tracks the envelope of the signal with little phase lag and amplitude variation over the bandwidth of the signal envelope.

Fig. 17 shows a detailed schematic of the adaptive bias circuit designed. The target is to independently control the parameters according to requirements 1) - 3) above while still supporting the bandwidth requirement. A digitally controlled

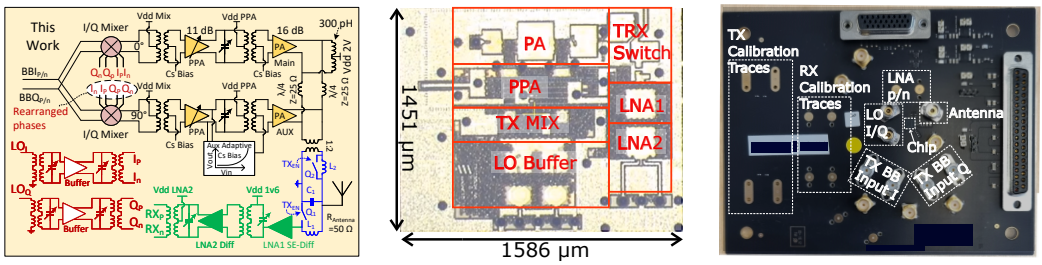

Fig. 16. Left: Architecture of the complete transceiver front-end, Middle: Chip photo, Right: PCB photo.

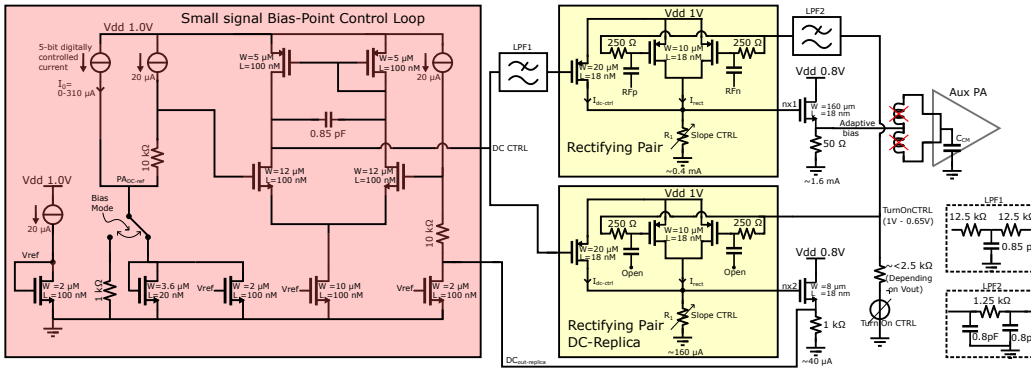

Fig. 17. Schematic of the adaptive bias generation circuit.

current source I_0, combined with either a current mirror or a 1 k Ω resistor produces a DC level used to set the small-signal operating point of the differential input transistors of the auxiliary PA. To operate the adaptive bias in constant GM mode for linearisation of an amplifier biased in class A, AB, or B the current mirror is selected and for class C mode operation, which is the case for the auxiliary amplifier of a Doherty PA, the 1 k Ω resistor is selected. Since the small-signal operation point, suitable for class C mode operation, should be well below V_{th} an additional 20 μA combined with a 10 k Ω resistor produces a voltage off-setted by 200 mV for a voltage level suitable to use as input to an operational amplifier (OP-amp).

Two PMOS transistors with combined output signal implements a rectifying pair and produces a current I_{rect} that depends on the RF signal amplitude. Controlling the gate bias level of the rectifying pair controls at what RF amplitude I_{rect} will start to increase i.e. the turn on point of the adaptive bias. Current I_{rect} is then transformed into a voltage through a tunable resistor R_1, which is used to control the slope of the increase of the adaptive bias voltage after the turn on. A straightforward method to combine the small-signal DC level with the amplitude dependent part of the adaptive bias signal that is produced by the rectifying pair is proposed in [21] which also is from the same first and last author. However, when adjusting the large signal settings, i.e. the turn on point and slope settings, an undesired effect is that the small-signal DC level changes, which makes it more difficult to find optimum settings for a particular use case and/or power amplifier design. Settings for the adaptive bias, i.e. the small and large signal settings, must be adjusted for optimal performance for each sample, and also when changes in operating frequency or temperature occur.

C. Orthogonalizing Large and Small Signal Settings

The unwanted impact on the DC-level is due to that when changing the settings for turn on point and slope it has an affect on the DC-voltage output from the rectifying pair, even for zero or low RF signal amplitudes. Ideally the DC-level should be unaffected by the large signal settings. In [21] changing the large signal adaptive bias setting results in about 75 mV of small-signal DC level changes. However, these changes were at a bias level above the threshold voltage. In the proposed design the aim is to have a small signal DC-level quite far below the threshold voltage to properly turn-off the class C biased auxiliary amplifier, which with a similar design as in [21] would result in a relatively larger DC-level change when the output current from the rectifying pair changes or resistor R_1 is changed.

To orthogonalize the large and small signal settings the small signal DC-level is controlled by a low frequency neg-

ative feedback loop, which regulates the output voltage of a common drain (CD) stage (node $DC_{out-replica}$), identical to the CD stage driving the output load, but downscaled in size and driven by an identical replica of the rectifying pair, to the same voltage as the desired DC-level (node PA_{DC-ref}). The replica has no RF input signal and therefore it captures the undesired effects that the large signal settings would have on the small-signal DC level. An OP-amp controls a transistor with output current $I_{dc-ctrl}$, that compensates for the undesired DC-level changes. A copy of the $I_{dc-ctrl}$ current is then fed to the rectifying pair which has the RF-input signal connected to it and is used to drive the CS input transistors of the auxiliary amplifier. As long as the rectifying pair is well matched to its replica the proposed method is expected to largely suppress output DC-level variations when adjusting the large signal settings.

To guarantee stability of the DC-level negative feedback loop, a rather large capacitor (0.85 pF) is added, realizing a capacitive narrowbanding compensation. Another problem that arises is that when a large RF signal is present, a quite large 2^{nd}-harmonic is present in nx1 and the turn on control nodes, and to avoid that it leaks back and affects the output signal from the DC-replica, two low pass filters (LPF) are added. One LPF is added between the output signal of the OP-amp and the current source transistor, which produces $I_{dc-ctrl}$ in the rectifying pair, and one LPF is added between the rectifying pair and the resistive ladder (indicated as an controllable voltage source in Fig. 17) producing the voltage reference for the turn on control of the rectifying pair.

D. Adaptive Bias High Bandwidth Design Considerations

For effective linearisation, the adaptive bias signal must be able to follow the envelope of the modulated signal and must drive the common mode input capacitance which is estimated in (34). The auxiliary amplifier used here has a simulated common mode input capacitance of $1.4 - 1.8$ pF including parasitic capacitances, depending on the bias level.

The rectifying pair drives a resistively loaded common drain (CD) buffer, wide enough to provide a low output impedance, to be capable of driving the common mode input capacitance with a sufficiently large bandwidth. Since the circuit does not offer any possibility to measure a high bandwidth adaptive bias signal, the performance is verified by simulations of the adaptive bias circuit and measurements on the quality of the modulated PA output signal. A two tone simulation, which gives an envelope with the beat frequency of the two RF input tones, that the adaptive bias should follow, was conducted with varying tone separation. Compared to the OFDM-signal the envelope of the two tone simulation is rather extreme, since all the energy is located at the edges of the band, whereas for the OFDM-signal the energy is evenly distributed over the bandwidth. At 1.6 GHz tone separation the magnitude of the adaptive bias signal has dropped by < 1.3 dB and the phase lags by 41 degrees, compared to the magnitude and phase of the adaptive bias signal for a low tone separation frequency.

However, a perhaps more illustrative result of how well the adaptive bias follows the envelope of high bandwidth modulated signals is shown in Fig. 18, which compares post layout simulation results of the adaptive bias circuit output signal with the theoretically derived ideal adaptive bias signal, using the envelope of an OFDM signal of 0.4 - 3.2 GHz modulation BW, when driving the auxiliary PA. The ideal adaptive bias signal is derived using $\gamma = 1$ and is identical to the results shown in Fig. 9 and 10. As can be seen in Fig. 18, for large signal levels and low bandwidths the simulated adaptive bias signal tracks the desired theoretical adaptive bias signal closely. For low signal levels, i.e. for envelopes below 0.5 V, where the adaptive bias signal ideally is zero there is a larger deviation. This, however, is not a a severe issue and is not expected to cause any linearity degradation. Normalizing the time axis makes it possible to visualize how the amplitude and phase depend on the bandwidth, and as can be seen, even for a very high bandwidth (3.2 GHz) only minor deviations occur.

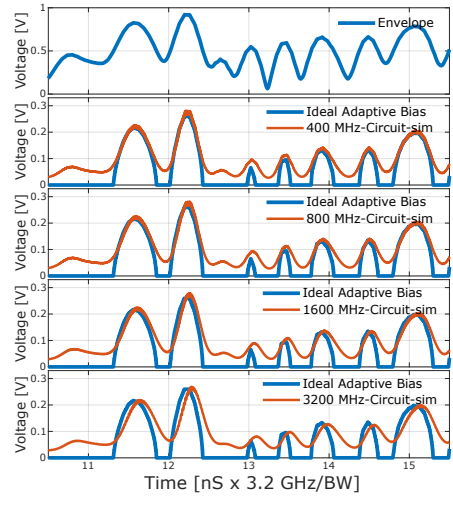

Fig. 18. Upper: Normalized envelope of OFDM signal with 0.4 - 3.2 GHz BW. Second from top to bottom: Ideal adaptive bias signal for upper envelope and simulated adaptive bias signal with increasing modulation bandwidth. Scale on time-axis is normalized to visualize the effect of increasing signal BW.

E. Adaptive Bias Power Consumption

The absolute majority of the power consumption of the adaptive bias circuit comes from the CD buffer driving the auxiliary PA, and to reduce its power consumption it is supplied by 0.8 V, while the other parts of the adaptive bias circuit are supplied by 1 V. When the adaptive bias circuit is operating with a high PAR signal at full input power, the CD buffer power consumption will be set by the average voltage level of the CS input stage ($< 200\,mV$), which can be seen in Fig. 18, and the 50 Ω resistor. In average the complete power consumption becomes about 4 mW of the adaptive bias circuit.

IX. Measurement Results

The test circuit was measured in detail in [3] and optimal adaptive bias settings were identified. Identical settings were used for both continuous wave (CW) and modulated signal measurements. In this section, the theoretical results, as presented in this paper, derived using a simplified ideal transistor model suitable for hand calculations, is compared with measurement results from two different samples. The test circuit does not offer any possibility to measure the output current, output power, or power consumption of the individual main and auxiliary amplifiers, instead the combined output power and power consumption of the Doherty amplifier were analysed.

A. CW-Tone Measurements

Fig. 19 shows measured output power, AM-AM, and adaptive CS gate bias voltage to the auxiliary amplifier, when stimulated with a CW tone of increasing input power, for two different samples. The measurement was repeated with a static auxiliary class C bias, i.e. turned off adaptive bias. The theoretical predictions from Fig. 6, 8, and 9 for $\gamma = 1$ are then normalised to the same output power level and plotted in the same figure, both with and without adaptive bias. Considering that the theoretical models use highly simplified frequency agnostic transistor models with zero output current below V_{th} and an output current I_{out} equal to $g_m(V_{gs} - V_{th})$ as long as $I_{out}R_{Load} < V_{dd}$, the predictions show good agreement with the presented measurement results at 26.5 GHz, both for the output power, see Fig. 19 Upper, and for AM-AM, see Fig. 19 Middle. The test circuit is equipped with a low frequency output to measure the adaptive CS gate bias voltage to the auxiliary amplifier when the PA is operating with constant envelope signals, and this measured output voltage from the adaptive bias circuit is shown in Fig. 19 Lower.

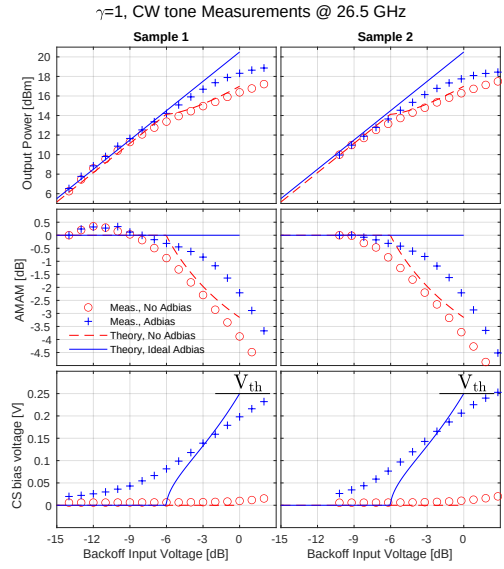

Fig. 19. Measurements compared with theoretical predictions using ideal transistor model for two samples, with and without adaptive bias vs input signal amplitude. Upper: Output power. Middle: AM-AM. Lower: Adaptive bias voltage. Unfortunately some low output power measurements are missing for Sample 2.

To validate that the negative feedback loop effectively controls the small-signal bias level such that the bias level does not change too much when changing the large signal settings (turn-on point and slope), the measured small signal bias level is plotted in Fig. 20. The result shows a variation of about 26 mV when changing the two registers, which is about one third of the reported small-signal bias level variation in [21].

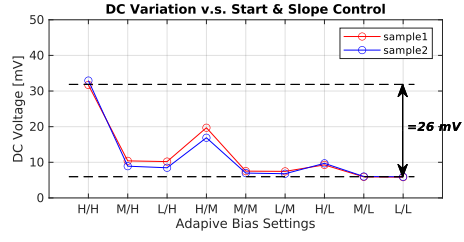

Fig. 20. Small-signal bias level (DC) v.s adaptive bias large signal control settings. H, M, L stands for High, Mid, Low. First letter is slope control and second letter is turn-on point control

Fig. 21 illustrates orthogonality of the two large signal controls, slope and turn-on point. When changing the turn-on point the slope should ideally be unaffected, and vice versa. The turn-on point is defined as when the adaptive bias level has increased by 5 mV, and the slope is defined as the slope of the adaptive bias signal at the point in the middle (on a logarithmic scale) between the turn-on point and the maximum input signal. For completely orthogonal control, the measurements should result in straight vertical lines.

B. Modulated Signal Measurements

Modulated measurements were performed to validate the high frequency performance of the adaptive bias circuit. Fig. 22 shows measured results of the key parameters EVM, PA drain efficiency (DE), and adjacent channel leakage ratio (ACLR) for 12 dBm average output power for sample 1 and 10 dBm average output power for sample 2 with 1600 MHz modulation BW, with and without adaptive bias, i.e. with static class C bias.

Fig. 23 compares measurement results for signals with different modulation bandwidth of 100 MHz, 200 MHz, 400 MHz, 800 MHz, 1600 MHz. As can be seen, both inband and out of band linearity (EVM and ACLR) are significantly improved (about 2 dB) by the adaptive bias, at a very small expense of efficiency (about 0.3 %-units) for all bandwidths, which is in line with the predictions of the simulated perfor-

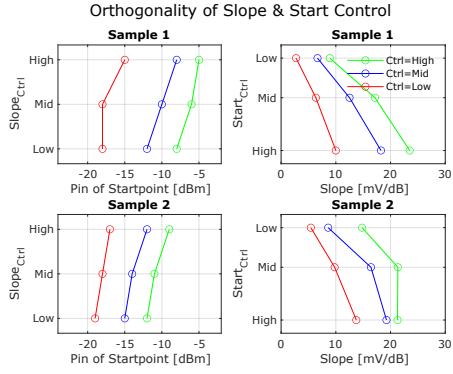

Fig. 21. Orthogonality of slope and turn-on point.

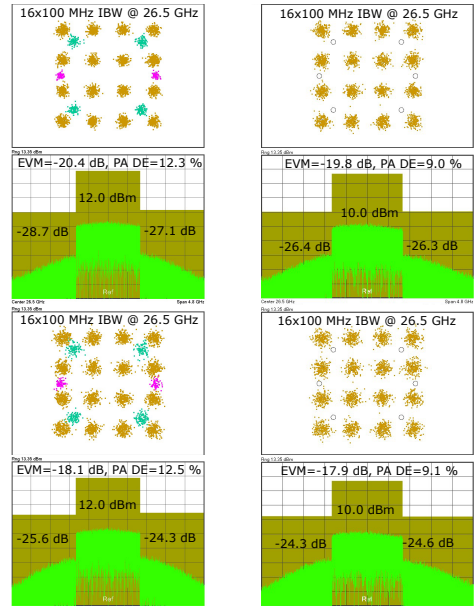

Fig. 22. Sample 1 (left column) and sample 2 (right column), measured performance at PA output with modulated 5G NR 16QAM-OFDM signals for 1600 MHz BW using optimized adaptive bias signal for the auxiliary amplifier (upper row) and compared with static class C bias for the auxiliary amplifier (lower row). Pilot symbols are shown for sample 1 in cyan and purple color. Average value for the modulated signal EVM, PA DE, ACLR, and output power are stated in the figure.

mance of the adaptive bias generation circuit in [3].

To investigate the amplitude dependency of the performance improvement that the adaptive bias brings, a 400 MHz modulated 16-QAM signal was measured, in Fig. 24 both for sample 1 and sample 2, at different output power levels. The results clearly show that for low output power levels the adaptive bias does not affect the results, but as the peaks of the high PAR input signal starts to exceed the B.O. level, the adaptive bias starts to linearise the output signal.

Fig. 25 shows PA DE vs transmitter EVM for different signal BW using adaptive bias, and compares this to a static bias with different levels for a fixed BW (400 MHz) signal, in both cases with OFDM modulation. As expected, the static bias increase improves EVM at the expense of PA DE, whereas the adaptive bias improves EVM without reducing PA DE as the signal BW is reduced.

X. Conclusion

Analytical derivations show that for ideal Doherty operation the fundamental frequency output current from the auxiliary amplifier must increase linearly for input amplitudes above the back-off level. Even with ideal transistor models, the Doherty amplifier will therefore produce a nonlinear output signal or a reduced efficiency, due to the nonlinear increase of the output current from the auxiliary amplifier. Two mitigation techniques are identified. First, increase the transconductance, i.e. g_m, of the auxiliary amplifier compared to the main amplifier. Second, control the auxiliary amplifier by dynamically adjusting the bias level so that the desired output current level is achieved. Ideally, this makes it possible to adjust the back-off level, produce a linear auxiliary output current, and avoid to increase the size of the auxiliary amplifier. The adaptive bias signal, however, suffers from bandwidth expansion both due to its origin from the envelope signal and the nonlinear function that produces it from the envelope. Therefore, generating the adaptive bias signal can be challenging. Analytical calculations in combination with simulated bias signals, reveal that even with rather non-ideal adaptive bias signals, however, significant improvements in AM-AM, saturated output power, and efficiency, can be achieved. For instance, an ideal adaptive bias signal filtered through a 1^{st}-order low pass filter with a cut-off frequency that is just 25% of the modulated RF signal bandwidth, suppresses amplitude distortion by more than 6 dB of a 7.5 dB PAR OFDM modulated signal. Similar linearity improvement can be achieved by doubling the size, i.e. g_m, of the auxiliary amplifier. The design of a highly configurable adaptive bias circuit is presented and analysed in detail, for CW-tone as well as for modulated signals. Simulations show that it is possible to generate an adaptive bias signal that quite closely tracks an ideal theoretically calculated adaptive bias signal consuming as little as 4 mW power. Measurements also confirm that design strategies for orthogonalizing the control settings for small-signal bias level, turn-on point, and slope for the adaptive bias where successful. CW-tone and modulated signal measurements on two samples of a Doherty PA with equally sized main and auxuiliary amplifiers confirm the presented theory. The adaptive bias increases the 1 dB compression point by >3dB. EVM and ACLR are improved by about 2 dB at a very small expense in efficiency. Signal measurements with different BW confirm that the integrated adaptive bias generating circuit can successfully track the

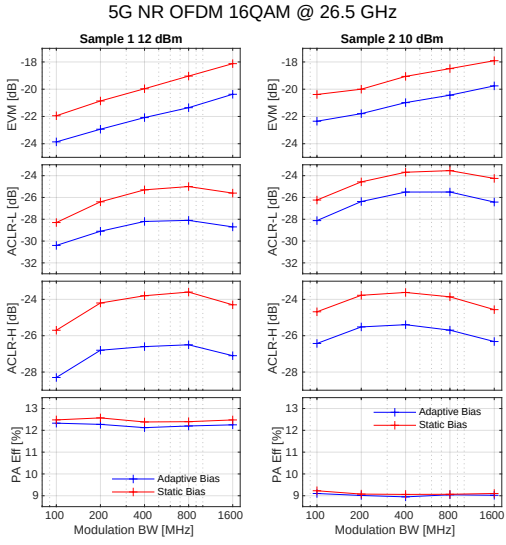

Fig. 23. Comparison of measured EVM, PA DE, and ACLR v.s. modulation BW, both with and without adaptive bias

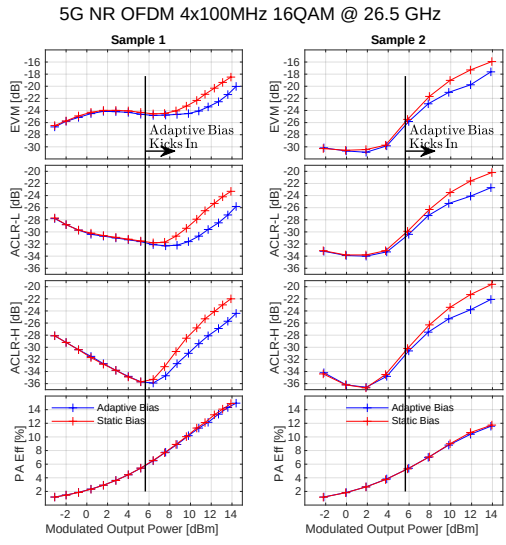

Fig. 24. Comparison of measured EVM, PA DE, and ACLR v.s. output power, both with and without adaptive bias

envelope of the signal, improving the linearity of the output signal, at a very low efficiency degradation. Finally, the impact of a static bias increase of the class C biased auxiliary amplifier is measured and compared with the gains brought by the adaptive bias for different bandwidths.

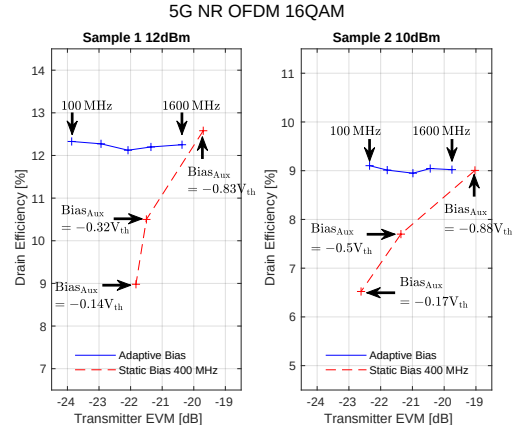

Fig. 25. Measured PA DE vs. Transmitter EVM using adaptive bias with varying signal BWs (1600, 800, 400, 200, and 100 MHz), and for static class C bias with different levels at 400 MHz signal BW.

REFERENCES

[1] W. Doherty, "A New High Efficiency Power Amplifier for Modulated Waves," *Proc. Inst. Radio Engineers*, vol. 24, no. 9, pp. 1163–1182, 1936.

[2] W.H. Doherty, "Amplifier," US Patent 2210028A, 1936.

[3] C. Elgaard, M. Özen, E. Westesson, A. Mahmoud, F. Torres, S.B. Reyaz, T. Forsberg, R. Akbar, H. Hagberg, and H. Sjöland, "Efficient Wideband mmW Transceiver Front End for 5G Base Stations in 22-nm FD-SOI CMOS," *IEEE J. Solid-State Circuits*, vol. 59, no. 2, pp. 321–336, 2024.

[4] J. Jung *et al.*, "A 39 GHz 2×16-Channel Phased-Array Transceiver IC With Compact, High-Efficiency Doherty Power Amplifiers," in *IEEE Radio Freq. Integr. Circuits Symp. (RFIC)*, 2023, pp. 273–276.

[5] S. Shakib, M. Elkholy, J. Dunworth, V. Aparin, and K. Entesari, "A Wideband 28-GHz Transmit–Receive Front-End for 5G Handset Phased Arrays in 40-nm CMOS," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 2946–2963, 2019.

[6] S. Mondal, R. Singh, and J. Paramesh, "21.3 A Reconfigurable Bidirectional 28/37/39GHz Front-End Supporting MIMO-TDD, Carrier Aggregation TDD and FDD/Full-Duplex with Self-Interference Cancellation in Digital and Fully Connected Hybrid Beamformers," in *IEEE Int. Solid-State Circuits Conf. (ISSCC)*, 2019, pp. 348–350.

[7] J. Park, S. Lee, D. Lee, and S. Hong, "A 28GHz 20.3%-Transmitter-Efficiency 1.5°-Phase-Error Beamforming Front-End IC with Embedded Switches and Dual-Vector Variable-Gain Phase Shifters," in *IEEE Int. Solid- State Circuits Conf. (ISSCC)*, 2019, pp. 176–178.

[8] L. Gao and G.M. Rebeiz, "A 22–44-GHz Phased-Array Receive Beamformer in 45-nm CMOS SOI for 5G Applications With 3–3.6-dB NF," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 11, pp. 4765–4774, 2020.

[9] J. Pang *et al.*, "A 28-GHz CMOS Phased-Array Beamformer Utilizing Neutralized Bi-Directional Technique Supporting Dual-Polarized MIMO for 5G NR," *IEEE J. Solid-State Circuits*, vol. 55, no. 9, pp. 2371–2386, 2020.

[10] X. Tang, Y. Liu, G. Mangraviti, Z. Zong, K. Khalaf, Y. Zhang, W.M. Wu, S.H. Chen, B. Debaillie, and P. Wambacq, "Design and Analysis of a 28 GHz T/R Front-End Module in 22-nm FD-SOI CMOS Technology," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 6, pp. 2841–2853, 2021.

[11] J. Pang *et al.*, "A CMOS Dual-Polarized Phased-Array Beamformer Utilizing Cross-Polarization Leakage Cancellation for 5G MIMO Systems," *IEEE J. Solid-State Circuits*, vol. 56, no. 4, pp. 1310–1326, 2021.

[12] Y. Yin, B. Ustundag, K. Kibaroglu, M. Sayginer, and G.M. Rebeiz, "Wideband 23.5–29.5-GHz Phased Arrays for Multistandard 5G Applications and Carrier Aggregation," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 1, pp. 235–247, 2021.

[13] W. Zhu, J. Wang, X. Zhang, W. Lv, B. Liao, Y. Zhu, and Y. Wang, "A 24–28-GHz Four-Element Phased-Array Transceiver Front End With 21.1%/16.6% Transmitter Peak/OP1dB PAE and Subdegree Phase Resolution Supporting 2.4 Gb/s in 256-QAM for 5-G Communications," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 6, pp. 2854–2869, 2021.

[14] Y. Yi, D. Zhao, J. Zhang, P. Gu, Y. Chai, H. Liu, and X. You, "A 24–29.5-GHz Highly Linear Phased-Array Transceiver Front-End in 65-nm CMOS Supporting 800-MHz 64-QAM and 400-MHz 256-QAM for 5G New Radio," *IEEE J. Solid-State Circuits*, vol. 57, no. 9, pp. 2702–2718, 2022.

[15] B. Sadhu, A. Paidimarri, W. Lee, M. Yeck, C. Ozdag, Y. Tojo, J.O. Plouchart, X. Gu, Y. Uemichi, S. Chakraborty, Y. Yamaguchi, N. Guan, and A. Valdes-Garcia, "A 24-to-30GHz 256-Element Dual-Polarized 5G Phased Array with Fast Beam-Switching Support for >30,000 Beams," in *IEEE Int. Solid- State Circuits Conf. (ISSCC)*, vol. 65, 2022, pp. 436–438.

[16] B. Sadhu, Y. Tousi, J. Hallin, S. Sahl, S. Reynolds, O. Renström, K. Sjögren, O. Haapalahti, N. Mazor, B. Bokinge, G. Weibull, H. Bengtsson, A. Carlinger, E. Westesson, J.E. Thillberg, L. Rexberg, M. Yeck, X. Gu, D. Friedman, and A. Valdes-Garcia, "7.2 A 28GHz 32-element phased-array transceiver IC with concurrent dual polarized beams and 1.4 degree beam-steering resolution for 5G communication," in *IEEE Int. Solid-State Circuits Conf. (ISSCC)*, 2017, pp. 128–129.

[17] J.D. Dunworth, A. Homayoun, B.H. Ku, Y.C. Ou, K. Chakraborty, G. Liu, T. Segoria, J. Lerdworatawee, J.W. Park, H.C. Park, H. Hedayati, D. Lu, P. Monat, K. Douglas, and V. Aparin, "A 28GHz Bulk-CMOS dual-polarization phased-array transceiver with 24 channels for 5G user and basestation equipment," in *IEEE Int. Solid - State Circuits Conf. (ISSCC)*, 2018, pp. 70–72.

[18] J. Pang *et al.*, "A 28-GHz CMOS Phased-Array Transceiver Based on LO Phase-Shifting Architecture With Gain Invariant Phase Tuning for 5G New Radio," *IEEE J. Solid-State Circuits*, vol. 54, no. 5, pp. 1228–1242, 2019.

[19] A. Verma *et al.*, "A 16-Channel, 28/39GHz Dual-Polarized 5G FR2 Phased-Array Transceiver IC with a Quad-Stream IF Transceiver Supporting Non-Contiguous Carrier Aggregation up to 1.6GHz BW," in *IEEE Int. Solid- State Circuits Conf. (ISSCC)*, vol. 65, 2022, pp. 1–3.

[20] J. Pang, Y. Zhang, Z. Li, M. Tang, Y. Liao, A.A. Fadila, A. Shirane, and K. Okada, "A Power-Efficient 24-to-71 GHz CMOS Phased-Array Receiver Utilizing Harmonic-Selection Technique Supporting 36dB Inter-Band Blocker Rejection for 5G NR," in *IEEE Int. Solid- State Circuits Conf. (ISSCC)*, vol. 65, 2022, pp. 434–436.

[21] C. Elgaard, S. Andersson, P. Caputa, E. Westesson, and H. Sjöland, "A 27 GHz Adaptive Bias Variable Gain Power Amplifier and T/R Switch in 22nm FD-SOI CMOS for 5G Antenna Arrays," in *IEEE Radio Freq. Integr. Circuits Symp. (RFIC)*, 2019, pp. 303–306.

[22] C.R. Chappidi, T. Sharma, Z. Liu, and K. Sengupta, "Load Modulated Balanced mm-Wave CMOS PA with Integrated Linearity Enhancement for 5G applications," in *IEEE Int. Microw. Symp. (IMS)*, 2020, pp. 1101–1104.

[23] Y. Jin and S. Hong, "A 24-GHz CMOS Power Amplifier With Dynamic Feedback and Adaptive Bias Controls," *IEEE Microw. Wireless Compon. Lett.*, vol. 31, no. 2, pp. 153–156, 2021.

[24] S. Chen, G. Wang, Z. Cheng, P. Qin, and Q. Xue, "Adaptively Biased 60-GHz Doherty Power Amplifier in 65-nm CMOS," *IEEE Microw. Wireless Compon. Lett.*, vol. 27, no. 3, pp. 296–298, 2017.

[25] H.T. Nguyen, T. Chi, S. Li, and H. Wang, "A 62-to-68GHz linear 6Gb/s 64QAM CMOS doherty radiator with 27.5leveraging high-efficiency multi-feed antenna-based active load modulation," in *IEEE Int. Solid-State Circuits Conf. (ISSCC)*, 2018, pp. 402–404.

[26] H.T. Nguyen and H. Wang, "A Coupler-Based Differential Doherty Power Amplifier with Built-In Baluns for High Mm-Wave Linear-Yet-Efficient Gbit/s Amplifications," in *IEEE Radio Freq. Integr. Circuits Symp. (RFIC)*, 2019, pp. 195–198.

[27] H. Zhang, R.Z. Zhan, Y.C. Li, and J. Mou, "High Efficiency Doherty Power Amplifier Using Dual-Adaptive Biases," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 8, pp. 2625–2634, 2020.

[28] Z. Zong, X. Tang, K. Khalaf, D. Yan, G. Mangraviti, J. Nguyen, Y. Liu, and P. Wambacq, "A 28-GHz SOI-CMOS Doherty Power Amplifier With a Compact Transformer-Based Output Combiner," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 6, pp. 2795–2808, 2021.

[29] M. Pashaeifar, L.C.N. de Vreede, and M.S. Alavi, "A Millimeter-Wave CMOS Series-Doherty Power Amplifier With Post-Silicon Inter-Stage Passive Validation," *IEEE J. Solid-State Circuits*, vol. 57, no. 10, pp. 2999–3013, 2022.

[30] P. Colantonio, F. Giannini, R. Giofrè, and L. Piazzon, "The AB-C Doherty power amplifier. Part I: Theory," in *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 19, no. 3, May 2009, pp. 293–306.

[31] ——, *High Efficiency RF and Microwave Solid State Power Amplifiers*. 1st ed. IntechOpen, 2009.

[32] S.C. Cripps, *RF Power Amplifiers for Wireless Communications*, 1st ed. Norwood, MA, USA: Artech House, 1999.