SUPPLY MODULATED GaN HEMT
POWER AMPLIFIERS - FROM TRANSISTOR TO SYSTEM

A thesis submitted to Cardiff University
in candidature for the degree of
Doctor of Philosophy

Alexander Alt MSc BEng (Hons)
Student ID 1561746

11.07.2021

Supervisors at Cardiff University:
Dr Jonathan Lees PhD MSc BEng MIET
Professor Paul J Tasker BSc PhD FIET FIEEE FLSW
Abstract

Power amplifiers (PAs) for mobile communication applications are required to fulfil stringent requirements concerning linearity while keeping a high efficiency over a wide power range and bandwidth. To achieve this, a number of advanced PA topologies have been developed, mostly based on either load-modulation, such as Doherty PAs or load modulation balanced PAs, or on supply modulation such as envelope tracking or envelope elimination and restoration. Supply modulation has an advantage over other topologies as the power range of high efficiency can be realised over arbitrary bandwidths, only limited by the bandwidth of the PA. This does, however, come at the cost of a significantly more complicated voltage supply. Instead of a static supply voltage, the PA needs to be provided with one which is rapidly changing, requiring a supply modulator capable of powering the PA while modulating its supply voltage. This thesis investigates a number of challenges in supply modulated power amplifiers, ranging from the transistor itself to circuit design and system level considerations and focusses on power levels up to 10 W and frequencies between 1 GHz and 4 GHz.

Transistors, as the centre-piece of a PA, determine how well the PA reacts to a changing supply voltage. In this work, the traits that make GaN HEMTs suitable for supply modulated PAs were investigated, and gain variation with changing supply voltage was established as an important parameter. This gain variation is described in detail and its impacts on PA performance are discussed. By comparing transistors in literature, gain variation has been demonstrated to be a prevalent characteristic in transistors with GaN HEMTs showing a very wide range of gain variation. Using a small-signal model based on measurements, the voltage dependent behaviour of the feedback capacitance $C_{GD}$ is, for the first time, identified as the origin of small-signal gain variation. This is traced down to the gate field plate which is commonly used to combat surface trapping effects in GaN HEMTs. With this in mind, two different ways of changing the transistor geometry to reduce the impact of gain variation and thus optimise the transistor for operation in supply modulated PAs are discussed and demonstrated using a 250 nm GaN HEMT.

As a result of the non-linearity of the feedback and gate-source capacitances, the input impedance of GaN HEMTs changes with supply voltage and drive power. This prevents the transistor from being matched at all supply voltages and input powers and introduces phase distortion. Using simulation and measurement, the impact of input impedance on linearity and efficiency of supply modulated power amplifiers is demonstrated on a 2.9 GHz 10 W PA. Careful selection of the input impedance allows improvement of AM/PM distortion of a supply modulated PA with little cost in terms of AM/AM and PAE.
Supply modulators have a significant impact on efficiency and linearity of the ET system. One supply modulator topology with the potential to generate a supply voltage with a high modulation bandwidth is the RF modulator in which the input DC voltage is turned into an RF signal and rectified, resulting in an output voltage which depends on the excitation of the PA. While PAs are well understood in every detail, there are gaps in the understanding of RF rectifiers. Using active load-pull/source-pull measurements, intrinsic gate and drain waveforms of a GaN HEMT operated as a rectifier are demonstrated for the first time. This allows in-detail evaluation of the impact of the gate termination in self-synchronous rectifiers. It also allows detailed analysis of the loss mechanisms in rectifiers and formulation of the required impedances to realise efficient self-synchronous operation, resulting in efficiencies exceeding 90% over wide power ranges. Using waveform engineering, a new type of RF modulator, with potentially very high bandwidths, based on even harmonic generation/injection is proposed. The necessary operating conditions of the rectifier part of the modulator are emulated using an active load-pull/source-pull system to successfully demonstrate that the rectifier behaves as predicted. Using a simple demonstrator, preliminary measurements were conducted and the RF modulator was shown to work, reaching efficiencies up to 78%.

As PA and supply modulator are combined, they present impedances to each other. These impedances have a significant impact on the behaviour of both sub-systems. A simple way to characterise both the impedance presented to the PA by the modulator and the impedance presented to the modulator by the PA is described. Using a state-of-the-art modulator, these impedances are measured, the modulator impedance is demonstrated to be close to the simulated value. These measurements also demonstrate that the impedances change significantly with the operating conditions.
Publications

First Author


Co-Author


Acknowledgements

“You can trust us to stick with you through thick and thin...But you cannot trust us to let you face trouble alone.”

J.R.R. Tolkien, The Fellowship of the Ring

Firstly, I would like to thank my supervisors, Jonathan Lees for giving me free reign over my research while providing guidance and support and Prof Paul Tasker for sharing his accumulated knowledge on everything semiconductor, they embody to me what university should be. I was lucky to be part of the InGaNET consortium which not only gave me insights into all steps of compound semiconductor fabrication and access to experts in these steps, it also lead to fruitful collaborations. I would like to extend my gratitude to all my collaborators, especially Kean Boon Lee at the University of Sheffield, Alejandro Villarruel-Parra and Andrew Forsyth at the University of Manchester and Abdalla Eblabla at Cardiff University, their ability to show me different perspectives was much appreciated. I owe thanks to Cardiff University for funding the PhD project, and allowing me to go on this wonderful academic journey. Thank you also to my colleagues at Cardiff university, Roberto Quaglia, Peng Chen and Sattam Alsahali for the discussions and to Michael Casbon for his invaluable support with everything measuring-related.

Special thanks for the Pen & Wig crew, Sam, Mike, Jerome, Heungjae, Nyle, Alek, Derek and Roberto, for providing a reason to leave earlier on Friday afternoons, I enjoyed the silly ideas and nerdy discussions in that garden, rain or shine.

During my time in Cardiff, I was very lucky to find a wonderful group of friends that made my PhD journey better in many ways, providing emotional support and excuses to go on adventures. I owe a big thank you for that to Mary, Katie, Jon, Rachael, Liad, Kate, Cleo, Matt, Katya, Anna and Becky. My heartfelt thanks to Anna and Becky for not only making the house feel like a home but also for accepting me in your families, it meant a lot to me! Over the last years, I met many wonderful people in the hiking club, caving club and mountaineering club, thank you all. I also want to thank my friends abroad for being there for me despite us meeting only a few times a year. Some of you I could welcome in Cardiff, others I could not but I appreciate you always making time for me. Danke!

Finally, I would like to thank my parents, Rolf and Gabriele, for their constant, unwavering support and their encouragement over all the years and my siblings Konstantin and Katharina for always being there for me and supporting me.
# List of Acronyms and Symbols

## Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2DEG</td>
<td>2-Dimensional Electron Gas</td>
</tr>
<tr>
<td>AC</td>
<td>Alternating Current</td>
</tr>
<tr>
<td>ACLR</td>
<td>Adjacent Channel Leakage Ratio</td>
</tr>
<tr>
<td>ACPR</td>
<td>Adjacent Channel Power Ratio</td>
</tr>
<tr>
<td>AM</td>
<td>Amplitude Modulation</td>
</tr>
<tr>
<td>APSK</td>
<td>Amplitude and Phase Shift Keying</td>
</tr>
<tr>
<td>AWG</td>
<td>Arbitrary Waveform Generator</td>
</tr>
<tr>
<td>BJT</td>
<td>Bipolar Junction Transistor</td>
</tr>
<tr>
<td>BS</td>
<td>Base Stations</td>
</tr>
<tr>
<td>CGP</td>
<td>Current Generator Plane</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal Oxide Field Effect Transistor</td>
</tr>
<tr>
<td>CPU</td>
<td>Central Processing Unit</td>
</tr>
<tr>
<td>CW</td>
<td>Continuous Wave</td>
</tr>
<tr>
<td>DC</td>
<td>Direct Current</td>
</tr>
<tr>
<td>DPA</td>
<td>Doherty Power Amplifier</td>
</tr>
<tr>
<td>DPD</td>
<td>Digital Pre Distortion</td>
</tr>
<tr>
<td>DPS</td>
<td>Dynamic Power Supply</td>
</tr>
<tr>
<td>DUT</td>
<td>Device Under Test</td>
</tr>
<tr>
<td>EDGE</td>
<td>Enhanced textbf{Data rates for} GSM Evolution</td>
</tr>
<tr>
<td>EER</td>
<td>Envelope Elimination and Restoration</td>
</tr>
<tr>
<td>ET</td>
<td>Envelope Tracking</td>
</tr>
<tr>
<td>EVM</td>
<td>Error Vector Magnitude</td>
</tr>
<tr>
<td>FET</td>
<td>Field-Effect Transistor</td>
</tr>
<tr>
<td>FPGA</td>
<td>Field-Programmable Gate Array</td>
</tr>
<tr>
<td>FP</td>
<td>Field Plate</td>
</tr>
<tr>
<td>GaN</td>
<td>Gallium Nitride</td>
</tr>
<tr>
<td>GFP</td>
<td>Gate Field Plate</td>
</tr>
<tr>
<td>GMP</td>
<td>Generalised Memory Polynomial</td>
</tr>
<tr>
<td>GMSK</td>
<td>Gaussian Minimum Shift Keying</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>GND</td>
<td>Ground</td>
</tr>
<tr>
<td>GPRS</td>
<td>General Packet Radio Service</td>
</tr>
<tr>
<td>GSG</td>
<td>Ground Signal Ground</td>
</tr>
<tr>
<td>GV</td>
<td>Gain Variation</td>
</tr>
<tr>
<td>HBT</td>
<td>Heterojunction Bipolar Transistor</td>
</tr>
<tr>
<td>HEMT</td>
<td>High-Electron-Mobility Transistor</td>
</tr>
<tr>
<td>HF</td>
<td>High Frequency</td>
</tr>
<tr>
<td>HV</td>
<td>High Voltage</td>
</tr>
<tr>
<td>IBO</td>
<td>Input Power Back-Off</td>
</tr>
<tr>
<td>IC</td>
<td>Integrated Circuit</td>
</tr>
<tr>
<td>IF</td>
<td>Intermediate Frequency</td>
</tr>
<tr>
<td>IMD</td>
<td>Inter Modulation Distortion</td>
</tr>
<tr>
<td>IMN</td>
<td>Input Matching Network</td>
</tr>
<tr>
<td>IQ</td>
<td>In-phase Quadrature</td>
</tr>
<tr>
<td>ISMN</td>
<td>Inter Stage Matching Network</td>
</tr>
<tr>
<td>IV</td>
<td>Voltage Current</td>
</tr>
<tr>
<td>JFET</td>
<td>Junction Field-Effect Transistor</td>
</tr>
<tr>
<td>LA</td>
<td>Linear Amplifier</td>
</tr>
<tr>
<td>LDMOS</td>
<td>Lateral-Diffusion Metal-Oxide-Semiconductor Field Effect Transistor</td>
</tr>
<tr>
<td>LINC</td>
<td>Linear amplification using Non-linear Components</td>
</tr>
<tr>
<td>LMBA</td>
<td>Load Modulated Balanced Amplifier</td>
</tr>
<tr>
<td>LTI</td>
<td>Linear Time-Invariant</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term-Evolution</td>
</tr>
<tr>
<td>LUT</td>
<td>Look Up Table</td>
</tr>
<tr>
<td>MAG</td>
<td>Maximum Available Gain</td>
</tr>
<tr>
<td>MESFET</td>
<td>Metal-Semiconductor Field-Effect Transistor</td>
</tr>
<tr>
<td>MISFET</td>
<td>Metal-Insulator-Semiconductor Field-Effect Transistor</td>
</tr>
<tr>
<td>MLC</td>
<td>Multi Level Converter</td>
</tr>
<tr>
<td>MLM</td>
<td>Multi Level Modulator</td>
</tr>
<tr>
<td>MMIC</td>
<td>Monolithic Microwave Integrated Circuit</td>
</tr>
<tr>
<td>MODFET</td>
<td>Modulation-Doped Field-Effect Transistor</td>
</tr>
<tr>
<td>MOSFET</td>
<td>Metal-Oxide-Semiconductor Field-Effect Transistor</td>
</tr>
<tr>
<td>MP SMM</td>
<td>Multi-Phase Switch Mode Modulator</td>
</tr>
<tr>
<td>MSG</td>
<td>Maximum Stable Gain</td>
</tr>
<tr>
<td>MTTF</td>
<td>Mean Time To Failure</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>OBO</td>
<td>Output Power Back-Off</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal Frequency Division Multiplexing</td>
</tr>
<tr>
<td>OMN</td>
<td>Output Matching Network</td>
</tr>
<tr>
<td>OPA</td>
<td>Out-phasing Power Amplifier</td>
</tr>
<tr>
<td>PA</td>
<td>Power Amplifier</td>
</tr>
<tr>
<td>PAE</td>
<td>Power Added Efficiency</td>
</tr>
<tr>
<td>PAPR</td>
<td>Peak-to-Average Power Ratio</td>
</tr>
<tr>
<td>PCB</td>
<td>Printed Circuit Board</td>
</tr>
<tr>
<td>PDF</td>
<td>Power Density Function</td>
</tr>
<tr>
<td>PEP</td>
<td>Peak Envelope Power</td>
</tr>
<tr>
<td>PET</td>
<td>Power Envelope PTracking</td>
</tr>
<tr>
<td>PM</td>
<td>Phase Modulation</td>
</tr>
<tr>
<td>PN</td>
<td>Positive-Negative junction</td>
</tr>
<tr>
<td>PSK</td>
<td>Phase Shift Keying</td>
</tr>
<tr>
<td>PWM</td>
<td>Pulse Width Modulation</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>RFM</td>
<td>Radio Frequency Modulator</td>
</tr>
<tr>
<td>RMS</td>
<td>Root Mean Square</td>
</tr>
<tr>
<td>SFP</td>
<td>Source Field Plate</td>
</tr>
<tr>
<td>SiC</td>
<td>Silicon Carbide</td>
</tr>
<tr>
<td>SiGe</td>
<td>Silicon Germanium</td>
</tr>
<tr>
<td>SM</td>
<td>Supply Modulation</td>
</tr>
<tr>
<td>SM PS</td>
<td>Switch Mode Power Supply</td>
</tr>
<tr>
<td>SMA</td>
<td>SubMiniature version A</td>
</tr>
<tr>
<td>SMD</td>
<td>Surface Mount Device</td>
</tr>
<tr>
<td>SMM</td>
<td>Switch Mode Modulator</td>
</tr>
<tr>
<td>SPA</td>
<td>Sequential Power Amplifier</td>
</tr>
<tr>
<td>UE</td>
<td>User Equipment</td>
</tr>
<tr>
<td>UHF</td>
<td>Ultra High Frequency</td>
</tr>
<tr>
<td>VNA</td>
<td>Vector Network Analyzer</td>
</tr>
<tr>
<td>VSWR</td>
<td>Voltage Standing Wave Ratio</td>
</tr>
<tr>
<td>VST</td>
<td>Vector Signal Tranceiver</td>
</tr>
<tr>
<td>WCDMA</td>
<td>Wideband Code Division Multiple Access</td>
</tr>
<tr>
<td>ZVS</td>
<td>Zero Voltage Switching</td>
</tr>
</tbody>
</table>
Electrical Units and Sizes

\[ V \quad \text{Voltage} \quad [V] = \left[ \frac{\text{kg} \cdot \text{m}^2}{\text{A} \cdot \text{s}^2} \right] \]

\[ I \quad \text{Current} \quad [\text{A}] \]

\[ P \quad \text{Power} \quad [\text{W}] = \left[ \frac{\text{kg} \cdot \text{m}^2}{\text{A}^2 \cdot \text{s}^3} \right] \]

\[ R \quad \text{Resistance} \quad [\Omega] = \left[ \frac{\text{kg} \cdot \text{m}^2}{\text{A} \cdot \text{s}^2} \right] \]

\[ L \quad \text{Inductance} \quad [\text{H}] = \left[ \frac{\text{kg} \cdot \text{m}^2}{\text{A}^2 \cdot \text{s}^2} \right] \]

\[ C \quad \text{Capacitance} \quad [\text{F}] = \left[ \frac{\text{A}^2 \cdot \text{s}^4}{\text{kg} \cdot \text{m}^2} \right] \]

\[ t \quad \text{Time} \quad [\text{s}] \]

\[ T \quad \text{Period} \quad [\text{s}] \]

\[ f \quad \text{Frequency} \quad [\text{Hz}] = \frac{1}{s} \]

\[ G \quad \text{Gain} \quad [\text{dB}] \]

\[ Z_{\text{opt}} \quad \text{Optimum load impedance} \quad [\Omega] \]

\[ V_{\text{DD}} \quad \text{Drain supply voltage} \quad [\text{V}] \]

\[ V_{\text{GG}} \quad \text{Gate supply voltage} \quad [\text{V}] \]

\[ \hat{V} \quad \text{Voltage amplitude} \quad [\text{V}] \]

\[ Z_0 \quad \text{Characteristic impedance} \quad [\Omega] \]

\[ Z \quad \text{Complex impedance} \quad [\Omega] \]

\[ |Z| = Z \quad \text{Absolute value of complex impedance} \quad [\Omega] \]

\[ Z_{\text{L}} \quad \text{Complex load impedance} \quad [\Omega] \]

Greek Characters

\[ \Delta \quad \text{Difference} \quad [1] \]

\[ \varphi \quad \text{Phase shift} \quad [^\circ] \text{ or } [\text{rad}] \]

\[ \varphi_c \quad \text{Conduction angle} \quad [^\circ] \]

\[ \Gamma \quad \text{Reflection coefficient} \quad [\text{dB}] \]

\[ \Gamma \quad \text{Complex reflection coefficient} \quad [1] \]

\[ \Sigma \quad \text{Sum} \quad [1] \]

\[ \omega \quad \text{Angular frequency} \quad [\frac{1}{s}] \]
# Contents

Abstract

Publications

Acknowledgements

List of Acronyms and Symbols

Contents

1 Introduction

1.1 Motivation

1.2 Research Questions

1.3 Original Contributions

1.4 Overview of the Thesis

References

2 Power Amplifiers for Mobile Communications

2.1 Mobile Communication Standards

2.2 GaN HEMTs in Power amplifiers

2.3 Transistor Characterisation

2.4 Power Amplifiers

2.5 Integrated Power Amplifiers

2.6 Non-linearities and Linearisation

2.7 Conclusions

References

3 Developments in Supply Modulation

3.1 Supply Modulation at System Level

3.2 Supply Modulators

3.3 Supply Modulated Power Amplifiers

3.4 Linearity and Linearisation

3.5 Conclusions

References
## Contents

### 4 GaN-on-Si HEMTs in Supply Modulated PAs

- 4.1 Characterisation ............................................. 123
- 4.2 Gain Variation in GaN HEMTs ................................. 137
- 4.3 Optimising GaN HEMT Geometry for Supply Modulated PAs .................. 151
- 4.4 Conclusions ...................................................... 154

References .............................................................. 154

### 5 Input matching Supply Modulated PAs

- 5.1 Concept .......................................................... 159
- 5.2 Source-pull simulations ........................................... 160
- 5.3 Verification ......................................................... 163
- 5.4 Conclusions ......................................................... 167

References .............................................................. 168

### 6 RF Supply Modulators

- 6.1 Characterising Transistors as Rectifiers ........................ 171
- 6.2 Gate Voltage in Transistor-based RF Rectifiers ................ 173
- 6.3 Class F RF Rectifiers ............................................... 180
- 6.4 Class B* RF DC/DC Converters ................................... 185
- 6.5 Conclusions ......................................................... 197

References .............................................................. 198

### 7 Integrating PA and Supply Modulator

- 7.1 Baseband Impedance ............................................... 200
- 7.2 Power Amplifier Impedance ....................................... 210
- 7.3 Supply Modulator Ripple ......................................... 214
- 7.4 Conclusions ......................................................... 219

References .............................................................. 220

### 8 Conclusions and further work

- Future Work ........................................................ 225

References .............................................................. 227

Bibliography ........................................................... 229

List of Figures ......................................................... 252

List of Tables ......................................................... 263

Appendix ................................................................. 264
Chapter 1

Introduction

Contents

1.1 Motivation ......................................................... 2
1.2 Research Questions ........................................... 5
1.3 Original Contributions ......................................... 6
1.4 Overview of the Thesis ......................................... 6
References .......................................................... 8

You see, wire telegraph is a kind of a very, very long cat. You pull his tail in New York and his head is meowing in Los Angeles...And radio operates exactly the same way: you send signals here, they receive them there.

Albert Einstein

Mobile communication has overturned our economy and has massively influenced our society in the last decades. It has become possible to be more connected than ever before and the data traffic resulting from this is expected to grow further [1]. This further growth is partially due to an expected increase in data transmitted to and between end users. New fields such as machine-to-machine communications, e.g. between cars, are expected to play a bigger role in the future of mobile communications. These extensions in data traffic lead to multiple challenges. An obvious one is available spectrum. Most of the frequency bands of interest for mobile communications are tied up in existing services. Regulatory authorities such as Ofcom or FCC relocated spectrum from legacy services such as analogue terrestrial television to mobile communications in preparation for 5G [2],[3]. At the same time, other ways to transmit more data within the existing spectral bands were evaluated. One of the trends associated with the 5G standard is the reduction of cell sizes [4], allowing the network operators to increase the frequency reuse and thus increase the spectral efficiency [5]. It is also a step away from the cabinet sized base stations, that consume multiple kilowatts [6] and require sophisticated infrastructure, to smaller base stations. Base station energy consumption is the major contributor of mobile communication system power consumption as a whole [7] and, unfortunately, reducing cell sizes does little to alleviate the overall system power consumption. The power saved by removing the cooling system is at least partially spend by by the signal processing equipment which doesn’t scale...
with RF output power [6]. Independent of the size of the base station, the RF power amplifier is one of the main consumers of electrical energy both directly and indirectly, as the linearity of the PA determines the energy required by the pre-distorter part of the signal processing.

1.1 Motivation

Power amplifiers are subject to increasingly challenging requirements in terms of efficiency, linearity and bandwidth. Carrier aggregation is expected to increase the signal bandwidth to at least 400 MHz[8] in 5G communication systems. This also increases the requirements in terms of linearity as the third order intermodulation now encompasses a bandwidth of 800 MHz, and the fifth order intermodulation 1200 MHz. To avoid interference with other channels or even other services, emissions in the adjacent and alternate band need to be minimized. The modulation bandwidth is not the only bandwidth that is increasing however, and amplifiers are expected to be able to operate efficiently over wide bandwidths and in multiple band ranges. Continuous mode power amplifiers are able to fulfill this requirement [9].

The third challenge is power amplifier efficiency, with the peak to average power ratio (PAPR) keeps increasing with new communication standards [10], see Fig. 1.1.1. This means that the amplifier spends more and more time in output power back-off (OBO), i.e. at reduced power levels. In this region, single amplifiers are inherently less efficient [12] as the example in Fig. 1.1.2 demonstrates. Over the years, several topologies have been developed to address this problem. In an out-phasing amplifier for example, two amplifiers are driven to a level where they are operating efficiently, and by adjusting the phase shift between them, the output power can varied while achieving a high efficiency over a wide power range. In Doherty power amplifiers, one amplifier is used to modulate the load of the other in order to keep it in an efficient state. Supply modulation techniques solve the problem by adjusting the supply voltage of the amplifier to keep it efficient. Envelope tracking is one implementation of

![Figure 1.1.1: PAPR for mobile communication standards after [11]](image-url)
Figure 1.1.2: Power density function (PDF) for an LTE signal and efficiency of an ideal class B amplifier over normalised output voltage

a supply modulation technique and has been used, so far, predominantly in hand-held devices. The main obstacle in applying it in base station amplifiers is the generation of the required modulated supply voltages at high powers, esp. for high modulation frequencies. In this context, the step towards smaller base stations opens up new possibilities.

**Supply Modulation**

Power amplifiers are at their most efficient if they are operating close to their maximum output power. In this mode of operation, their drain voltage goes down to the minimum voltage, giving the highest voltage swing. If the output power is reduced, the voltage swing reduces with it. This leads to more current-voltage overlap and thus more power dissipated in the transistor, reducing the efficiency. Reducing the supply voltage will allow the drain voltage to reach the minimum voltage again, as demonstrated in Fig. 1.1.3. If the supply voltage is reduced, neither voltage nor current swing change which means that the RF output power stays the same. The DC power, however, is reduced due to the reduced supply voltage and this reduction at a constant RF output power results in an increased PA efficiency. In modulated signals, the amplitude of the signal changes constantly and so does the supply voltage, as demonstrated in Fig. 1.1.3. The supply voltage now needs to change synchronously with the envelope of the modulated signal and this reduction in supply voltage drastically reduces the DC power consumption of the power amplifier, increasing its efficiency. The reduced dissipated power in the transistors also reduces the transistor’s junction temperature. As a high junction temperature is one of the causes of failures in transistors [13], supply modulation can also increase the mean time to failure (MTTF) and thus help reduce maintenance costs. Changing the supply voltage dynamically is a task that requires a dynamic power supply (DPS) or supply modulator, that is able to track the modulation. This speed requirement results in reduced efficiencies, the efficiency of the whole ET
Figure 1.1.3: Voltages, powers and efficiencies of an ideal class B PA with and without envelope tracking system now has to take the efficiency of the DPS into consideration, see equation (1.1).

\[ \eta = \eta_{PA} \cdot \eta_{DPS} \]  

(1.1)

As a consequence, both PA efficiency and DPS efficiency have to be considered in an supply modulated system, and might have to be traded-off against each other to achieve the highest overall efficiency. Supply modulation and envelope tracking will be treated in more depth in section 2.4.7, a review on past and current developments will be presented chapter 3.

Integration of Modulator and Power Amplifier in GaN on Si Technology

Integrating circuits has many advantages over packaged devices on circuit boards. The integration increases repeatability due to very controllable process steps, allows higher frequencies due to reduced parasitics and has the potential to reduce cost. Current and future generation MMICs will also be able to integrate different materials and technologies, allowing the integration of complex structures such as complete transmitter/receiver chains [14]. The interface between power amplifier and supply modulator has been shown to be important for the linearity of the power amplifier [15]. The ability to control this interface might allow the design of more linear envelope tracking systems. The reduced parasitics will facilitate the design of a broad-band power amplifier and increase the efficiency of both
power amplifier and modulator. To fully leverage the cost benefit of a MMIC, it needs to be fabricated on an affordable material with good RF performance, in this work, the material of choice is GaN on Silicon.

**Role of Supply Modulated PAs in Future Mobile Networks**

While the upcoming mobile communication standards like 6G or even the some features proposed for 5G are far from being introduced, a few trends have emerged that will have far reaching ramifications on future base stations and the power amplifiers within them. To meet the demands for higher data rates and lower latencies as well more reliable communication, future mobile communication networks will become more heterogeneous with a large number of additional small base stations at a multitude of frequencies from currently used sub 6 GHz bands to frequencies above 26 GHz [8] with predictions that operating frequencies will even exceed the 275 - 325 GHz band currently developed by the IEEE [16]. This heterogeneity results in a wide number of applications for power amplifiers, some more suitable for supply modulation than others. For the larger macro BS PAs in the sub-6 GHz range, supply modulation is a good option to improve system efficiency. As the maximum output power drops, the additional overhead of signal processing in supply modulation can become significant and reduce the overall system efficiency; this is offset by the fact that supply modulators are more easily realised at lower power levels as can be seen by their adoption in mobile devices. At the mmWave frequencies between 26 GHz and 50 GHz, the power levels are low while the modulation bandwidths are significant which makes it challenging to modulate the supply voltage. This is likely to slow down the adoption of supply modulation in power amplifiers at mmWave frequencies. As frequencies are pushed towards THz, things get considerably worse. A significant challenge in tapping into that spectrum is the signal to noise ratio which worsens as the atmospheric attenuation and the modulation bandwidth increase, requiring simpler modulation standards. Double-digit GHz bandwidths make tracking the supply voltage challenging while modulation standards with lower PAPR reduce the achievable efficiency improvements, making supply modulation an unlikely feature in amplifiers above the 50 GHz bands in the foreseeable future.

1.2 Research Questions

Envelope tracking has gained a lot of traction and has thus attracted researchers from both industry and academia resulting in a rapidly growing body of knowledge. Extending that body of knowledge therefore necessitates focusing on specific parts of an envelope tracking system. The research conducted in the course of this PhD can be roughly separated into four parts, asking the questions:
• What makes a GaN-on-Si HEMT suitable for supply modulation and how can we optimise it?
• How can we optimise the circuit design in supply-modulated PAs?
• Can we achieve very high bandwidths for use in characterisation using RF DC/DC type supply modulators?
• How do the supply modulator and the PA interact?

1.3 Original Contributions

Several original contributions were made as part of the research conducted and will be discussed in this thesis:

• Described in detail and analysed the impact of gain variation in GaN HEMTs on ET PAs [17]
• Identified the origin of gain variation in GaN HEMTs [17]
• Established ways to optimise GaN HEMTs for application in supply modulated PAs [17], [18]
• Proposed a novel way of using the input matching network of a GaN HEMT based PA to optimise the linearity of a supply modulated PA [19]
• For the first time presented directly measured intrinsic drain and gate voltage and current waveforms of a RF rectifier [20], [21]
• Used the measured waveforms to establish the loss mechanisms in rectifiers [20], [21]
• Derived the required impedances for efficient self-synchronous rectifiers [20], [21]
• Proposed a novel type of RF DC/DC converter based on harmonic generation/injection [21]
• Demonstrated a working prototype of this novel converter [21]
• Devised a simplified one-shot measurement method for the baseband impedance of supply modulators [22]

1.4 Overview of the Thesis

Excluding this introduction, this thesis consists of seven chapters. The first two discuss the literature with different focuses, followed by four chapters that show in detail the research which was conducted,
and finally, one concluding chapter.

In chapter 2, the first literature review chapter, power amplifiers for mobile communication are considered in detail, starting with the requirements of mobile communications and the consequences this application has on the PA. Transistors, the core elements of a PA, are described in detail, including ways to characterise them. Following this, PAs themselves are discussed, including ways to increase their efficiency over both bandwidth and power range. As they play a significant role in communications, a short summary on the non-linearities in PAs and ways to reduce them concludes the chapter.

Chapter 3, the second literature review chapter, focuses on supply modulated systems. It starts by examining them at the system level, considering the interfaces between the different parts both physically and in terms of signals. Following this, different supply modulators are described in detail, and compared in terms of their performance. The way transistors and PAs react if they are subjected to supply modulation is then discussed, concluding the chapter with linearity and linearisation in supply modulated systems.

The four chapters which present the conducted research and its results are aligned with the research questions, resulting in one chapter per question.

The first research chapter, chapter 4, aims to show the characteristics a GaN-on-Si HEMT needs to have to be suitable for supply modulation. It therefore discusses characterisation to measure the transistor, non-linear deembedding to be able to examine the inner workings of the transistor, and small-signal modelling to describe the transistor. One significant result of the characterisation measurements is that the gain varies significantly with supply voltage, this motivated the following sections of the chapter, showing the impact of gain variation, a way to compare it for different devices and its intrinsic origin. The chapter concludes by proposing ways of reducing a transistors gain variation.

Working on the second research question, chapter 5 shows how, by exploiting the non-linear input capacitance, the input matching network can be used to increase the linearity of a supply modulated PA, also discussing the trade-off necessary for this to work.

Chapter 6 considers RF DC/DC converters, focussing mainly on the rectifier. It discusses the impact of the gate voltage on the rectifier performance based on intrinsic measurements and uses this information in characterising self-synchronous rectifiers, starting with a class F rectifier and focusing on establishing the loss mechanisms. A novel combination of PA and rectifier which uses the second harmonic generated within the rectifier to increase the PA efficiency is then proposed, with the rectifier successfully measured under these conditions. A proof-of-concept of this type of DC/DC converter
concludes this chapter.

The integration of PA and supply modulator is discussed in chapter 7. The baseband impedance of the modulator, its measurement and its impact on the PA are discussed, as is the other relevant impedance, that of the PA. As many supply modulators have a significant amount of output ripple, its impact and ways to mitigate that impact are also discussed.

The last chapter, chapter 8, concludes the conducted research and critically analyses the achieved results. After that, it highlights promising directions for future research and open research questions that developed as part of the research.

References

Introduction - Overview of the Thesis


Chapter 2

Power Amplifiers for Mobile Communications

Contents

2.1 Mobile Communication Standards .................................. 12
2.2 GaN HEMTs in Power amplifiers ............................... 15
  2.2.1 Basic operation of Field Effect Transistors .......... 15
  2.2.2 Operation of AlGaN/GaN HEMTs .................. 17
  2.2.3 Structure of GaN HEMTs ............................ 18
  2.2.4 State of the Art GaN HEMTs ...................... 27
  2.2.5 Emerging Technologies ................................ 28
2.3 Transistor Characterisation .................................... 30
  2.3.1 IV-Measurements ..................................................... 30
  2.3.2 S-parameters ......................................................... 31
  2.3.3 Load-Pull ............................................................. 31
  2.3.4 Modelling ............................................................ 33
2.4 Power Amplifiers ................................................. 35
  2.4.1 Efficiency ............................................................. 35
  2.4.2 Power Amplifier Bias Points ......................... 36
  2.4.3 Impact of the Knee Region ........................... 40
  2.4.4 Matching ............................................................. 40
  2.4.5 Efficiency Enhancement .......................... 41
  2.4.6 Efficient Operation over Wide Bandwidths .... 43
  2.4.7 Efficient Operation over High Dynamic Ranges .. 44
2.5 Integrated Power Amplifiers .................................... 57
2.6 Non-linearities and Linearisation ................................ 58
  2.6.1 Non-linearities in Power Amplifiers ............... 59
  2.6.2 Linearisation .......................................................... 63
2.7 Conclusions ........................................................... 64
References ............................................................ 65

Mobile communication has increased dramatically over the last few decades in terms of bandwidth, subscribers and traffic and is expected to grow further. Another factor that is expected to grow is network complexity, supplementing the macro cells with a high number of smaller cells [6] that need to be connected to the main network using either fibre or back-haul wireless transmission. Additionally, mobile devices are connected to the internet using not only the mobile network but also using WiFi, switching seamlessly between the two. Mobile devices can also communicate with multiple base
stations simultaneously at different frequencies. All these communications require the signals being send back and forth between the mobile device and some sort of hub, usually a base station or a WiFi router, see Fig. 2.0.1(a). The figure also shows device-to-device connections that will allow direct connection of devices with the network only allocating their communication channel. For any of this communication channels to work reliably, the sender needs to make sure the signal strength at the receiver is high enough to turn it back into data. Typical communication devices thus need a power amplifier on the sending side and a low noise amplifier on the receiving side, see Fig. 2.0.1(b). This thesis will focus on the power amplifier and ways to optimise it for specific requirements such as efficiency, bandwidth and linearity. This chapter is the first of two that will set the ground work, starting by discussing mobile communication standards and their requirements in section 2.1. After that, transistor used in mobile communications will be discussed as they form the centre piece of a PA, first focusing on the technology in section 2.2, followed by section 2.3 which discusses their
characterisation. Following that, power amplifiers will be discussed in section 2.4, with a particular focus on high efficiency architectures. Integrating PAs and the challenges therein are discussed in section 2.5, the chapter closes with linearity in PAs in section 2.6.

2.1 Mobile Communication Standards

The first communication standards were devised for audio communications only at a time when the internet was far from being widely used. With 2G, digital transmission was introduced into the world of mobile communications. Its simple and amplifier-friendly modulation, minimum-shift keying, with additional Gaussian filtering (GMSK), limits the power consumption of the amplifier by always operating it in compression where it is the most efficient. General packet radio service (GPRS), for the first time, allowed transmission of internet protocol (IP) packets and thus internet access. The modulation and available bandwidth limited the data rate, however. As the demand for data rate increased, EDGE was introduced with a modulation scheme that was adapted to 8 phase-shift keying (8PSK). This allowed higher data-rates by transmitting more symbols per second but also lead to a variation of output power increasing peak-to-average power ratio (PAPR), see Fig. 1.1.1 with the known effect of lower efficiency due to power back-off operation. In contrast to 2G, 3G was designed with internet connectivity in mind from the start. This is reflected in the modulation scheme employed and the baseband bandwidth, see Fig. 2.1.1. To accommodate a growing number of users, new frequency bands had to be made available by regulatory instances, see Fig. 2.1.2. As data consumption continued to rise and new, data-heavy, services emerged, 4G or LTE was designed to provide very high data rates of up to 100Mbit/s. Advances in signal processing allowed complex modulation and encodings such as orthogonal frequency division multiplexing (OFDM) could be introduced. This

![Baseband Bandwidth in kHz](image)

**Figure 2.1.1:** Development of baseband bandwidth in mobile communication systems
multi-carrier approach also lead to a higher PAPR. That, combined with an increased baseband bandwidth, lead to new challenges in PA design. LTE-Advanced with carrier aggregation and based-band frequencies of up to 100 MHz further complicated PA design. In addition to the modulation, a second step was taken to increase data rates: the reduction of cell size to increase the frequency reuse. This cell size reduction also reduced the power levels necessary for communications, leading to micro and nano cells and thus a shift in base station power consumption [6]. The last step was the allocation of a significantly larger portion of spectrum, now comprising frequencies from 450 MHz to 3.5 GHz. This spectrum was taken from older services made obsolete or more efficient by digital transmissions such as communication or terrestrial broadcasting channels.

**5G** continues the trend of allocating additional spectrum for transmission. The most obvious addition is that of a new frequency range from 24.25 GHz to 52.6 GHz [8], although so far the only allocated bands in that range are 24.25 - 29.5 GHz and 37 - 40 GHz, hinting at efforts to extend the bands to higher frequencies. The same can be seen in the lower frequency band: While the bands do not exceed 5.925 GHz, the low frequency range covers a range from 410 MHz to 7.125 GHz [8], [23]. Additionally, the standard now allows further channel aggregation, resulting in aggregated bandwidths of up to 400 MHz [8].

**Modulation**

To transmit data in modern, digital communication systems, digital signals have to be transformed into an analogue signals. A common way of modulating the carrier is by allocating values to a combination of amplitude and phase. By plotting amplitude and phase as a vector in a Cartesian coordinate system, each value can be displayed as point in that plane, see Fig. 2.1.3. The axis of this plane are usually called the in-phase (I) and quadrature (Q) axis where in-phase refers to a cosine, and quadrature to the sine with a phase shift of 90° between the two. Analogous to complex numbers, the vectors can be described using amplitude and phase or, comparable to real and imaginary part, using sine and cosine component. There are multiple ways to arrange the symbols in the IQ plane, the most...
common are APSK and QAM. In amplitude and phase-shift keying (APSK), the symbols are arranged in circles around the origin, the number of concentric circles can be increased to increase the number of points. Additionally, the points can be moved closer together on the circle [24]. Quadrature amplitude modulation (QAM), arranges the symbols on a regular, quadratic grid, such as the one in Fig. 2.6.1 with 16 values, hence the name 16QAM. As with APSK, the number of points can be increased arbitrarily. Even in its early stages, 5G is standardised to support 1024QAM [23]. QAM has the advantage that the symbols are equidistant, see Fig. 2.1.3 while for APSK, the symbols move closer together for the inner circles [24]. Especially for low numbers of points, APSK has the advantage that there are less power levels and less phase levels which results in the PA operating at high power, where it is most efficient half of the time while for QAM, the PA is at its highest efficiency only in the four corners, as shown in Fig. 2.1.3, resulting in an increased PAPR for 16QAM [25]. Independent on the arrangement of the symbols, an increase of the number of points requires the system to be able to reproduce the symbols more accurately as they move closer together but still need to be distinguished. The impact
amplifier non-linearities have in this will be discussed in more detail in section 2.6. When evaluating signals, the peak to average power ratio is a good starting point but it does not provide of information about the probability of the different power levels. Two measures to provide that information are the probability density function (PDF) and the complementary cumulative distribution function (CCDF). The PDF plots probability over power, showing how much time the signal spends at which output power. The CCDF offers a different perspective on the same data, showing the probability of the signal to be below a certain power level. This is demonstrated in Fig. 2.1.4 which shows the PDF and CCDF of two signals with PAPRs of 5.66 dB and 5.67 dB. While their PAPR is almost the same, their power distribution looks significantly different. To be able to transmit at higher data rates, OFDM uses several carriers modulated with QAM or APSK at the same time. The transmitted signal therefore consists of the sum of several modulated carriers which results in a significantly increased PAPR [11]. In addition to the linearity, the large PAPR and the wide bandwidth present a wide range of interesting challenges in PA design in terms of efficiency and bandwidth.

2.2 GaN HEMTs in Power amplifiers

This section is going to focus on transistors, starting by a short discussion on how field effect transistor operate in general in section 2.2.1, followed by an introduction to the principles behind GaN HEMTs section 2.2.2. This is followed by a breakdown of the layers in the epitaxy using the example of a GaN-on-Si HEMT and describing the relevant parameters in transistor design in section 2.2.3. The characteristics of transistors relevant in PA design are discussed and different technologies are compared in terms of their performance in section 2.2.3, followed by a comparison of state-of-the-art GaN HEMTs in section 2.2.4 and emerging transistor technologies in section 2.2.5.

2.2.1 Basic operation of Field Effect Transistors

Most of the transistors in our modern, connected world are field effect transistors (FETs) which use electric fields applied via a gate to change the amount of carriers in a channel between drain and source. Different technologies use pn-junctions (JFET), Schottky junctions (MESFET, HEMT) or insulating layers such as silicon-oxide (MISFET,MOSFET) to separate the gate from the channel. This separation has a significant impact on which gate source voltage needs to be applied for the channel to change between conducting and isolating, this gate source voltage is called the threshold voltage \( V_{th} \). In the example in Fig. 2.2.1, the threshold voltage is around -5.5 V. The figure shows how in the saturation region, the current is predominantly determined by the applied gate source
Drain
Source
Gate
Isolation
R
D
CDS
CGD
GDS
IDS
CGS
R
S

Channel
Depletion Region

Figure 2.2.2: Simple model of a FET showing capacitances and resistances

effect of the knee region on PA behaviour will be discussed in more detail in section 2.4.3. A third region not shown in Fig. 2.2.1 is the breakdown region which occurs at high voltages. Here, the voltage is high enough to cause impact ionisation [26, p.315] in the transistor, leading to massive current flow which heats the device and can result in the destruction of the device. The simple, generalised FET model Fig. 2.2.2 shows how, due to the separation of gate and channel, all FETs have a capacitance between the gate and the channel which can be separated into one part that forms a capacitance between gate and source and one between gate and drain. The ratio of that split is determined by the applied drain source voltage [27] and the transistor technology. As the gate source voltage, $V_{GS}$, is used to deplete the channel of carriers, the capacitors between gate and channel additionally depend on the gate voltage, $V_{GS}$. In the linear region, the current depends on both applied voltages. In the field of RF PA engineering, the linear region is usually referred to as knee region and avoided as much as possible. Operation in this region results in non-linearities due to the drain current in this region strongly depending on the drain voltage and gate voltage. Additionally, the figure shows that moving the drain-source voltage closer to 0 V reduces the maximum current the transistor can conduct. The

drained

Figure 2.2.1: Simulated characteristics of a Qorvo GaN-on-SiC HEMT: (a) DCIV and (b) transfer characteristics, drain current vs. gate voltage (solid lines) and transconductance vs. gate voltage (dotted lines)
2.2.2 Operation of AlGaN/GaN HEMTs

In HEMTs, the channel is formed by a sheet of carriers that is confined by the band structure of the materials around it. In a junction between a wide bandgap material and a material with a narrower bandgap, electrons diffuse from the wide bandgap to the narrow bandgap material until equilibrium is reached, resulting in a charge dipole at the interface, as shown in Fig. 2.2.3. In the material with the narrower bandgap, the carriers form a thin layer of charges called the 2D electron gas (2DEG). The number of carriers in the 2DEG, typically expressed as carrier per area in the metric carrier concentration, $n_s$, limits the achievable current; strategies to ensure a high carrier concentration in the 2DEG depend on the technology. In AlGaAs/GaAs HEMTs, the AlGaAs barrier can be doped to increase $n_s$, while in AlGaN/GaN HEMTs, $n_s$ is a result of spontaneous and piezoelectric polarisation [29]. As the carriers of the 2DEG travel in the channel, the characteristics of the channel material determine important characteristics such as transconductance, gain and linearity of the transistor [30]. Tab. 2.1 shows that GaN has a higher saturation velocity, energy bandgap and breakdown field than

| Table 2.1: Comparison of electrical properties of semi-conductors |
|-------------------|---|---|---|---|
| Energy bandgap $E_g$ [eV] | Si | 4H-SiC | GaAs | GaN |
| Saturation velocity $v_{sat}$ [$10^7$ cm/s] | 1 | 2 | 1 | 2.5 |
| Maximum Mobility $\mu_n$ [cm$^2$/Vs] | 1350 | 700 | 8500 | 2000 |
| Breakdown field $E_{bk}$ [V/$\mu$m] | 30 | 300 | 40 | 330 |

Data from [31]
other semi-conductors commonly used in communication systems, only being second in the maximum mobility. This demonstrates why the last two decades have seen a significant interest in GaN HEMTs, both commercially and in terms of research.

2.2.3 Structure of GaN HEMTs

Manufacturing a GaN HEMT can be split into two main parts, epitaxy, i.e. building the wafer to the required specifications, a planar process which ends in an unstructured wafer, and the processing in which the transistor is build using the wafer. This section will first discuss the basic epitaxy of a GaN-on-Si wafer and the trapping effects that result from different parts of it, and then discuss the transistor design, relevant design parameters and process steps.

Epitaxy and trapping

Figure 2.2.4 shows the layer stack of a GaN-on-Si wafer, most wafers for GaN HEMTs for RF PA applications are using a comparable structure. This section will discuss the individual layers and their roles, as well as the trapping behaviour they cause. The number of layers depends on the technology, transistors with only 4 epitaxial layers have been demonstrated [32], most published technologies are closer to the one in Fig. 2.2.4. Additionally, the thickness of the individual layers can differ considerably from the one in the figure, there seems to be little consensus on the optimum thickness of the layers apart from the barrier thickness [32]–[35].

The substrate has a few important roles to fulfil. It is the layer on which all other layers are deposited and therefore provides structural support. As it is the main thermal pathway between transistor and heat sink, it needs to conduct heat way away from the transistor, through the substrate into a heat-sink of some sort. To reduce the thermal impact of the substrate, wafers can be thinned
after processing, reducing the thermal resistance of the substrate significantly. As the temperature impacts the sheet resistance of the 2DEG \[36\], achievable power and gain are impacted by the channel temperature, making a good thermal pathway between transistor and heat sink desirable. Table 2.2 shows that diamond is superior, followed by SiC, AlN, GaN and Si. The thermal conductivity of all substrates decreases with temperature, their ranking stays the same, however \[37\], \[38\]. A

Table 2.2: Comparison of substrate materials for GaN HEMTs

<table>
<thead>
<tr>
<th></th>
<th>Si</th>
<th>SiC</th>
<th>AlN</th>
<th>GaN</th>
<th>Diamond</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lattice mismatch GaN/substrate (1)</td>
<td>-16.9 %</td>
<td>3.5 %</td>
<td>2.4 %</td>
<td>–</td>
<td>13 %</td>
</tr>
<tr>
<td>Thermal conductivity(\ast) [W/(mK)] (2)</td>
<td>149</td>
<td>400</td>
<td>285</td>
<td>160</td>
<td>1600</td>
</tr>
<tr>
<td>Resistivity(\ast)† [Ohm/cm] (3)</td>
<td>(0.01 - 10^4)</td>
<td>(&gt; 10^6)</td>
<td>(&gt; 10^8)</td>
<td>(3.8 - &gt; 10^{12})†</td>
<td>(0.03 - &gt; 10^{15})†</td>
</tr>
<tr>
<td>Cost [€/cm²] (4)</td>
<td>0.1</td>
<td>10</td>
<td>?</td>
<td>100</td>
<td>?</td>
</tr>
</tbody>
</table>

Data from \(1\) \[39\], \[40\], \(2\)\[34\], \[37\], \(3\) \[41\]–\[45\], \(4\) \[46\], \(\ast\) at room temperature † depending on doping

second important parameter is the resistivity of the substrate. As the transistor is separated from the substrate by merely a few micrometers, there is a capacitive coupling between the transistor and the substrate. While this has little impact in the case of an isolator, e.g. SiC, it becomes significant as the resistivity decreases \[41\]. This results in series RC circuits between any two elements, with the capacitance consisting of the two respective capacitances to the substrate, as shown in Fig. 2.2.5. As small elements have a smaller capacitance, the substrate coupling between the gate and other elements will be negligible while large elements such as the drain and source areas or the 2DEG as a whole will have a larger coupling. This RC circuit resistively loads any structure that couples to the substrate, as demonstrated using coplanar wave-guides in \[41\]. While the resistivity changes with temperature and frequency, the change is smallest in very low-resistivity Si \[41\]. Having a very low-resistivity substrate very close to the transistor will not result in losses; it will, however, results in significantly increased device capacitances and thus limit the performance \[47\]. Table 2.2 shows the resistivity of relevant substrate materials, the values depend significantly on the doping of the substrate. Additionally, the resistivity may change during processing \[48\], \[49\]. Not only technical parameters play important roles

![Figure 2.2.5: Schematic showing substrate coupling](image-url)
in selecting the substrate, however, the manufacturing needs to be viable from a financial point of view as well. Here, Si substrates outperform all other substrates, making them the obvious choice if cost is more important than performance. Transistors can be separated from the substrate [14], [37] and placed on a different one. This allows fabrication on cheap, large substrates, such as Si, and operation on more expensive substrates with better thermal conductivity and resistivity which might be challenging to produce at larger sizes in the required quality. By only replacing the substrate for a subset of the devices, manufacturers could serve the high performance and the low-cost market with the same devices. One example of such a high performance devices is obtained by removing the Si substrate of a fabricated HEMT and replacing it with single-crystalline diamond resulting in an improved gain and efficiency of the HEMT [37], although even with the high thermal conductivity of the diamond substrate, the transistors shows significantly better performance in pulsed conditions.

The nucleation layer plays multiple important roles as it provides an interface between the lattice of the wafer and that of the GaN buffer, necessitated by the lattice mismatch which is more or less pronounced, as shown in Tab. 2.2. As the nucleation layer is basically working as a matching element, the surface of the nucleation layer has a significant impact on the strain in the wafer. By optimising the growth conditions of the nucleation layer, 2DEG carrier density and mobility can be improved significantly [50]. As the nucleation layer connects the substrate to the GaN layers, it not only needs to provide an lattice interface, it also needs to provide a thermal interface, conducting heat into the substrate effectively. This requires particular attention to not introduce a thermal boundary that stops heat from flowing into the substrate [51].

The buffer provides isolation between substrate and channel, it is therefore not necessarily required in cases where the substrate has a very high resistivity [32], [52]. In cases with limited substrate resistivity, a highly resistive buffer can ensure that no vertical breakthrough takes place if the buffer thickness is chosen appropriately [53]. Doping the GaN buffer with iron atoms increases the buffer resistivity, reducing buffer leakage and increasing the breakdown voltage; its effect on the knee-walkout are disputed [52], [54]. To reduce the impact of the Fe doping on the transistors performance while keeping the high vertical breakthrough voltages, the buffer can be doped gradually with a high dopant concentration in the bottom of the buffer with the concentration reducing as the buffer transitions into the channel [28]. While carbon also reduces the buffer leakage by making the GaN layer highly resistive [55], [56], the carrier density and velocity decrease, decreasing the maximum current [56]. In addition to that, the knee-walkout is substantially increased [52], [57]. This is problematic as carbon atoms are present in the growth reactors [57], by adjusting growth temperature and reactor pressure, the carbon concentration in the buffer can be controlled. During operation of the transistor, defects in the buffer can trap carriers with energy levels of $E_C - 0.57$ eV and $E_C - 0.75$ eV, resulting in output power degradation and increased knee walkout [58].
The **channel** accommodates the 2DEG [59], it needs to be thick enough to do so. While the 2DEG is assumed to only extend in an infinitesimally thin plane parallel to the layer interfaces, in reality the carriers spread in the third dimension as well [30], [60]. By introducing a very thin sheet of AlN between channel and carrier, the 2DEG can be contained in the channel and prevented from spreading into the barrier, resulting in an improved performance in terms of carrier density and drift mobility [30], [59]. On the bottom side, the sheet resistance, gate leakage and output conductance can be improved by keeping the channel above 200 nm thick [61], a channel which is thick enough can maintain a good RF power and PA efficiency [62]. The carrier velocity of the carriers in the 2DEG is a non-linear function that depends on the electric field and the mobility [63], as well as the temperature [64]. In the areas of relatively low electric field, the drain and source access regions which are relatively far from the gate, this temperature dependence is significantly more pronounced than it is in the area around the gate where the electric fields are highest [64], the relatively low electric fields also results in relatively low electron velocities. The losses in the channel also depend on the electric field and increase with it, as a result, the region around the gate is the main source of heat in a HEMT [64].

As discussed in section 2.2.2, the **barrier** is responsible for the generation of the 2DEG, in the case of an AlGaN/GaN HEMT by using spontaneous and piezoelectric polarisation [29]. As the piezoelectric polarisation is a result of the strain in the lattice, the thickness of the barrier is critical. If the barrier is too thick, the internal strain reduces, reducing the piezoelectric polarisation and, with it, the carrier concentration [65], while a barrier which is too thin does not allow a 2DEG to form [65]. While the maximum 2DEG density has been reported to be around 14 nm in Al$_{0.34}$Ga$_{0.66}$N [65], dropping with a further increased barrier thickness, other papers report a 2DEG density which continues to rise with thickness at the cost of mobility in Al$_{0.32}$Ga$_{0.68}$N [28]. This discrepancy is possibly a result of different epitaxial processes that complicate direct comparisons. As the crystal structure, and thus the strain, depends on the ratio $x$ between Al and Ga in the Al$_x$Ga$_{1-x}$N barrier, the Al content has a significant impact on carrier concentration and mobility [55], [66]. While the carrier concentration increases with $x$, the mobility decreases, necessitating a trade-off.

The gate can be deposited straight onto the barrier, alternatively, depositing thin GaN layer, or **cap**, between barrier and gate allows trading off carrier concentration for mobility [28] and reduces the electric field magnitude at the gate edge while resulting in a reduced surface roughness [67]. Independent of the cap, electrons can get trapped on the surface of the of barrier or cap, forming a virtual gate which prevents the transistor fully turning on by presenting a persistent negative charge [68], [69]. In addition to the trapping at the surface of the epitaxial stack and the trapping in the buffer, trapping can also happen in the barrier [68], [70], the origin of that has been demonstrated to be largely tunnelling [70].
Figure 2.2.6: Geometry of a two finger GaN HEMT: (a) top view and (b) lateral view

Processing

On the wafer with all of the layers, the transistor is then fabricated. Fig. 2.2.6 shows a typical geometry and make-up of a two finger GaN HEMT, this section will discuss relevant parts of it. The figure shows how the active region, i.e. everything constituting to the 2DEG, is removed in most areas, with all connections being realised on the isolating layers, usually the buffer. In the width axis, there is no change in the transistor itself, i.e. the geometry is constant in the active region, as shown in Fig. 2.2.6. The design takes place in the length axis with the main parameters being the length of the gate and of the distances between the different contacts. The gate length, $L_G$, has a direct impact on the frequencies the transistor can operate at since it determines the gate capacitance and the gate resistance. The gate needs to be long enough to reliably deplete the channel, the gate length to achieve this depends on the distance of the 2DEG from the gate. If the gate is too short, the drain current depends on the applied drain source voltage [71], resulting in a $g_m$ that depends significantly on the drain source voltage. Additionally, the voltage required to pinch-off the transistor increases with applied drain-source voltage, leading to potential pinch-off issues. As keeping the gate short is important to achieving high frequency operation, establishing the minimum aspect ratio that provides a suitably small short channel effect is an important step in technology development. For GaN HEMTs with an aspect ratio of 10, i.e. a gate which is ten times longer than the distance between gate metal and 2DEG, the short channel effect is small, at an aspect ratio of 15, short channel effects all but disappear [72]. The gate length also has an impact on $g_m$ and the threshold voltage, as well as the maximum current [73]. In GaN HEMTs, gate lengths down to 60 nm have been presented [74], achieving $f_{max}$ values of up to 300 GHz while maintaining good current density and pinch-off and little short channel effect [74], reducing the gate length further will improve the maximum operating frequency, the move from $L_G = 60$ nm to $L_G = 20$ nm only increases $f_{max}$ from 300 GHz to 444 GHz, at a significantly higher short channel effect and a considerably more complex fabrication. The drain-source length, $L_{DS}$, also has a significant impact on the performance of the transistor, as it is directly
related to the conduction losses in the channel, it forms a part of the on-resistance of the transistor. The **source-gate length**, $L_{GS}$, is also part of that, its resistance, $R_S$, has an additional feedback effect as RF PAs are almost exclusively realised as common-source circuits, reducing the gain and the achievable output current [73], [75]. Reducing $L_{GS}$ too much has a negative effect on the breakdown voltage, however, necessitating a trade-off [75]. Instead of reducing the distance, a n+ source ledge has been proposed, showing improved maximum current, $g_m$ and $R_S$, albeit without further elaboration as to the reason of this improvement [76]. Another important factor in the geometry is the **gate pitch**, the distance between the gates, i.e. between the region where the most heat is generated. By measuring temperature rise at different distances from a heat source, it was shown that thermal coupling is an effect that works over significant distances even on good thermal conductors such as SiC [77].

The role of the passivation is to reliably insulate the different conducting parts of the transistor and to prevent surface traps from forming [68], [69]. An additional effect of SiN passivation is an increase in carrier density [69]. While a thin sheet of passivation is sufficient to prevent surface traps from forming, the passivation needs to be thicker in most GaN HEMTs as the field plates are deposited on the passivation, an increasing number of field plates will require more layers of passivation as well. SiN is the state of the art [66], but there are alternatives. AlSiN has been shown to increase both breakdown voltage and gain and thus PAE and power density [78].

By introducing **field plates**, the magnitude of the electric field under and around the gate can be reduced significantly [79] which improves the breakdown voltage. The reduced field strength also has a beneficial effect on the trapping behaviour, reducing the knee-walkout significantly [80]. This field plate is usually connected to the gate or the source of the HEMT, while connections to the drain has been proposed [81], there have been no reported measurement results of a device with drain field plates so far. Having multiple field plates allows more complex manipulations of the electric field distribution, having one gate field plate (GFP) and one source field plate (SFP) is a standard approach across manufacturers, as shown in Tab. 2.4. The length of this field plate has an impact on the electric field distribution [82] and thus the knee-walkout [83] and the intrinsic capacitances of the transistor. Adding a second SFP can be used to further optimise the field distribution [80]. The field plates can be modelled as additional, parasitic transistors with gates that are connected to the intrinsic gate in the case of the GFP and source in the case of a SFP [84], the field plate in this case forms a parasitic metal insulator semiconductor (MISFET) with the passivation being the insulator [85]. This MISFET has a threshold voltage significantly larger than the transistor itself due to the increased distance to the 2DEG, this increases the depletion region depending on the applied drain voltage [86]. The source field plate has a similar effect, as the distance between 2DEG and source field plate is usually significantly higher than that of the gate field plate, its impact is significantly lower [85]. As processes in compound semiconductor usually stack planar structures, the field plates are
Figure 2.2.7: Connecting multi-finger transistors using (a) planar gate and drain, (b) planar drain and source, and (c) source via holes

parallel to the channel in any standard process. By realising the field plate slanted instead of parallel to the channel, the threshold voltage of the parasitic MISFET becomes less clear, especially for longer field plates which results in a smoother $C_{GD}$ curve [87].

In multi-finger transistors, the transistor needs additional metal layers to connect all gate, all drain and all source contacts. Ideally, these connections would have no impact on the transistor performance, this is not the case in real transistors. A common way of realising the connections is keeping gate and drain connections planar with a distribution bar and connecting the source using air bridges over the drain fingers [33], as shown in Fig. 2.2.7. This allows an easy realisation of the transistor as a coplanar device at the cost of inductance between the source pads and the added capacitance between the bridges and the drain pads, it additionally is the most compact design. Another way keeps source and drain connection planar and connects the gate using bridges over or tunnels under the source bar [88] which results in less drain source capacitance but is more complex to fabricate since the gates are the smallest features, any crossing needs to be considered very carefully to not compromise the performance and reliability of the transistor. Instead of connecting all source pads on the top, the source pads can also be connected on the back of the substrate, using via holes. This requires the gap between the fingers to be wide enough to realise the vias and requires the substrate to be thinned to keep the source inductance low. Due to the size of the vias, this approach increases the size of the transistor considerably. It also limits the transistor to applications where the source of the transistor can be connected to the heat-sink, circuits such as buck-converters are not easily realisable. While vias are often used to conduct heat in PCBs, they have little thermal impact on the transistor unless they are filled with a metal such as gold [89], the fact that they require the gates to be further apart does improve the thermal behaviour, however.
Depositing Au on the barrier will lead to an Schottky contact, which is desired in case of the gate. The drain and source however need to be directly connected to the 2DEG, this is established using an **Ohmic contact** which can be manufactured using processes such as rapid thermal annealing or low temperature microwave annealing [90]. As this Ohmic contact has a resistance which is in series with resistance of the channel, it has a direct impact on the performance of the transistor. Especially for short gate-length transistors, this contact resistance can be larger than the resistance in the channel, dominating the on-resistance of the transistor. Using regrowth, the contact resistance can be reduced significantly [91], [92] at the cost of a additional processing steps.

**Transistors in power amplifiers**

In mobile communications, there are three major applications that use PAs, user equipment (UE) such as mobile phones or tablets, base-stations (BS) and wireless backhaul that connects the base stations to each other and to central nodes, as shown in Fig. 2.0.1a. While all of these three applications need PAs with high efficiencies, their requirements differ significantly. In UE, the power levels are significantly lower which allows realisation of the PAs in the very cheap CMOS [93] or SiGe [94] processes. These technologies are not able to output the power necessary in backhaul applications and base-stations. Depending on the size of the base station, the power levels can be higher by orders of magnitude [6], requiring other transistor technologies. At higher power levels, laterally-diffused metal-oxide semiconductor (LDMOS) devices, GaAs HEMTs, GaAs high voltage heterojunction bipolar transistor (HV HBT) and GaN HEMTs are the prevalent transistor technologies. This section will discuss these high power devices, their properties and how well they operate in a PA.

The main criteria PAs have to fulfil are output power, efficiency, and linearity. These criteria can be translated in transistor characteristics. The output power is determined by the maximum current and the maximum voltage. The efficiency is influenced by the knee region, lossy elements in the transistor, and gain. The linearity depends on the transconductance, the knee region and electrical and thermal memory effects in the transistor. Additionally, all main criteria depend on the temperature of the transistor and thus on its thermal behaviour and the way it is connected to a heat sink. Transistors are scaled according to the desired maximum current, this scaling has no impact on any of the other parameters such as the maximum voltage, knee region and memory effects which are determined by the transistor technology and epitaxy. The transistor scaling does, however, scale up the intrinsic losses and the intrinsic capacitances. As a result, larger transistors have higher intrinsic capacitances. To be able to compare devices of different sizes, the capacitances and the achievable output power are usually normalised to the width of the transistor. From an application point of view, this is not a particularly relevant figure of merit as a low normalised capacitance has no advantage if the achievable current and maximum voltage require the transistor to be significantly larger. To
take that into account, the capacitances are normalised to the achievable output power instead of the transistor width. The intrinsic capacitances are particularly relevant in broadband PAs where the capacitances need to be incorporated into the input and output matching networks, the larger the capacitances, the harder it becomes to achieve a broadband match [95].

Table 2.3: Comparison of intrinsic capacitances for different transistor technologies

<table>
<thead>
<tr>
<th></th>
<th>LDMOS</th>
<th>GaN HEMT</th>
<th>GaAs pHEMT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average $C_{GS}$</td>
<td>1.6 pF/W [96]</td>
<td>0.3 - 0.5 pF/W [97], [98]</td>
<td>1.6 pF/W [99]</td>
</tr>
<tr>
<td>Average $C_{DS}$</td>
<td>0.6 pF/W [96]</td>
<td>70 fF/W [97], [98]</td>
<td>0.3 pF/W [99]</td>
</tr>
<tr>
<td>Average $C_{GD}$</td>
<td>9 fF/W [96]</td>
<td>7 fF/W [97], [98]</td>
<td>65 fF/W [99]</td>
</tr>
</tbody>
</table>

To achieve a broadband input match, it is desirable for transistors to have a small input capacitance $C_{GS}$, with the achievable value determined by the geometry of the transistor, as discussed in section 2.2.3. Table 2.3 shows that in GaN HEMTs, the average $C_{GS}$ normalised to output power is significantly lower than it is in LDMOS and GaAs devices. On the output side of the transistor, the drain-source capacitance $C_{DS}$ and drain-gate capacitance $C_{GD}$ need to be incorporated into the output matching network (OMN). While $C_{GD}$ is often ignored, it can have a significant impact as table 2.3 shows. In GaN and GaAs devices, $C_{GD}$ is only 5 to 10 times smaller than $C_{DS}$, in LDMOS devices, this ratio is significantly larger. As these capacitances change with applied voltages, operation with varying supply voltage requires a more detailed investigation of these capacitances, this will be discussed in section 3.3.1. Ways to take these variations into consideration during the OMN design will be discussed in section 3.3.2. The achievable bandwidth is not only limited by the intrinsic capacitance but also by the optimum load-line of the transistor and thus the maximum voltage and current [12]. The further from the system impedance of usually 50 Ω, the harder it becomes to match the transistor over a wide bandwidth. Assuming an optimum load-line of 50 Ω, the power resulting at that load depends exclusively on the voltage. The load line can be assumed to extend from zero to the breakdown voltage or symmetrically around the recommended supply voltage, the latter one will be used in this case as it removes assumptions as to how close to the breakdown voltage the drain voltage is allowed to be. Assuming a sinusoidal drain voltage, the relationship between power, load impedance and supply voltage can be calculated, as shown in Fig. 2.2.8. The figure also shows common supply voltages of transistors for RF and microwave applications; 8 V, 12 V, 28 V and 50 V. The figure shows that, depending on the targeted power level, transistor technologies with different supply voltages are advantageous. GaAs devices have supply voltages up to 8 V which means that they have an output power of 0.64 W at the 50 Ω load-line. A 10 W, 8 V GaAs PA would have the intrinsic load-line to $3 Ω = 50 Ω/17$, complicating broadband matching significantly. GaN devices are
available with supply voltages between 12 V and 50 V, making them a contender for a wide range of power levels, while LDMOS devices are available with supply voltages from 28 V to 65 V which marks them as a good choice for very high power levels. While it is possible to operate a transistor at a reduced supply voltage to reduce its load impedance is possible, it is rarely done as not fully utilising transistors is a costly way of facilitating matching, the fact that the knee region limits the achievable efficiency at lower supply voltages more than at high supply voltages is another factor that needs to be taken into consideration. Moving in the other direction, i.e. increasing the supply voltage, is not feasible due to the hard limitation of the breakdown voltage. GaN devices are available with a wide range of supply voltages, Fig. 2.2.9 shows how the frequency at which the current gain reaches one, $f_T$, decreases with increasing supply voltage. Combining the data from Fig. 2.2.8 and Fig. 2.2.9 shows that with increasing frequencies, it becomes increasingly hard to realise high power broadband PAs.

### 2.2.4 State of the Art GaN HEMTs

Figure 2.2.9 compared a wide range transistors in terms of supply voltage, $f_T$ and power density, this section will look in more detail at a slice of this figure, transistors with a gate length of 250 nm. As Fig. 2.2.9 showed, this is one of the most common gate lengths for GaN HEMTs. The reason behind this is possibly that in the frequency range they operate at, GaN HEMTs have the output power edge over GaAs devices, while providing significantly higher gain and better efficiencies than LDMOS devices with no other technology to fill the gap. Reducing the scope to one specific gate length facilitates comparison of the different technologies. As publication of key parameters differs significantly between manufacturers, they can not be directly compared in all cases, resulting in empty cells in Tab. 2.4. The table shows that most manufacturers use SiC substrates, apart from MACOM which uses a Si substrate. While direct comparison is complicated by the fact that different manufacturers present their transistor under different conditions, all technologies show similar performance in terms of most
Table 2.4: Comparison of 250 nm technologies

<table>
<thead>
<tr>
<th>metrics</th>
<th>Wolfspeed</th>
<th>Qorvo</th>
<th>IAF</th>
<th>UMS</th>
<th>WIN</th>
<th>MACOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substrate</td>
<td>SiC</td>
<td>SiC</td>
<td>SiC</td>
<td>SiC</td>
<td>SiC</td>
<td>Si</td>
</tr>
<tr>
<td>Field plates</td>
<td>Gate and Source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current density [A/mm]</td>
<td>0.8</td>
<td>1.2</td>
<td>0.88</td>
<td>1.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$g_m$ [mS/mm]</td>
<td>350</td>
<td>290</td>
<td>340</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_T$ [GHz]</td>
<td>30</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breakdown voltage [V]</td>
<td>120</td>
<td>145</td>
<td>100</td>
<td>&gt;100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Power density [W/mm]</td>
<td>5 - 8</td>
<td>7</td>
<td>5.9</td>
<td>4.5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Supply voltage [V]</td>
<td>28, 40</td>
<td>50</td>
<td>30</td>
<td>30</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>PAE [%] at 4 GHz</td>
<td>65</td>
<td>73</td>
<td>&gt;50</td>
<td>&gt;65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAE [%] at 1 GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAE [%] at 10 GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data from</td>
<td>[104]</td>
<td>[100]</td>
<td>[105]</td>
<td>[102], [103]</td>
<td>[106], [107]</td>
<td>[98]</td>
</tr>
</tbody>
</table>

metrics, different technologies result in different breakdown voltages and thus different supply voltages however.

2.2.5 Emerging Technologies

While most research aims to improve current technology by optimising the layer structure in terms of thickness, doping, and quality, or the transistor geometry in terms of field plates, transistor dimensions, and fabrication as a whole, there is also research outside of this. This section looks at emerging technology that show promising behaviour or interesting concepts. As the publications in this area describe relatively early research or research in a sparsely populated research field, the results are typically not surpassing current technology in all relevant areas; whether the technical challenges will
be overcome and the emerging technologies will enter mainstream research is a question time will answer.

Double-heterojunction field-effect transistors with very thin barrier layers and regrown ohmics offer very low on-resistances, the small aspect ratio additionally results in very high operating frequencies at high current densities and breakdown voltages [108]. A enhancement mode transistor using this technology has a pinch-off voltage of around 0 V, which seems to increase with supply voltage however. Additionally, the device seems limited in terms of usable supply voltage [108]. A depletion mode transistor using a similar technology seems to share none of these weaknesses, however [109]. By using novel combinations of materials, e.g. using Scandium in ScAlN/GaN transistors [110], low sheet resistances [111] and thus very high current densities of 2.4 A/mm can be reached with small on-resistances [110], there has been no reported performance in terms of power density or efficiency, however. While in conventional devices, $f_{\text{max}}$ is significantly larger than $f_T$, $f_{\text{max}}$ is equal [110] to significantly smaller [112] than $f_T$, indicating little voltage gain to voltage attenuation in the transistor, which means that the gate voltage has to be equal or larger than the drain voltage. This presents an issue since they need to be out of phase, significantly improving the voltage at the gate edge.

So far, all transistors have been realised as planar structures built up using parallel layers. In addition to planar technologies, there are approaches that make use of the third dimension to improve the transistor behaviour. In tri-gate GaN-HEMT, the channel is split width-wise by trenches into smaller transistors. The gate metal is deposited in these trenches, wrapping around the channel which reduces short channel effects [113] while increasing gain, transconductance and current density. Whether the gate width used for the calculation incorporates the trenches is not specified, neither is the transistor’s susceptibility to trapping. By replacing the standard barrier-channel structure with a stack of multiple layers of 2DEG and surrounding this structure with the gate, the knee-region can be reduced significantly while keeping the leakage current low in so-called super-lattice castellated FETs [114]. The transconductance is relatively low and the knee-walkout are significant however, the latter might be a result of the lack of gate field plate in the trenches. A different way of three dimensional structures is taking the V groove approach used extensively in silicon power applications to improve the breakdown voltages, to retain the advantages of the 2DEG while improving the breakdown voltage significantly [115]. This approach is still in its infancy, however, with no report on its electrical behaviour so far.
2.3 Transistor Characterisation

Before a transistor can be used in a circuit, it needs to be characterised. Characterisations can serve multiple purposes: They are necessary to establish that the transistor is working as intended, to give the circuit designer critical information about optimum operating conditions and to generate a model that allows simulations in conditions that have not been covered in measurements. The characteristics that are evaluated depend on the application, if the goal is designing a low noise amplifier, noise measurements will be critical to establish the suitability of the transistor and give the designer the necessary impedances, possibly bias-dependent, but no large signal measurement will have to be conducted. If the transistor is to be evaluated for use in a PA however, small-signal measurements are not sufficient and will need to be supplemented using large signal measurements.

2.3.1 IV-Measurements

The transistor characterisation that is the easiest to conduct, at least in a basic way, is the IV measurement. It works by sweeping gate and drain voltage of the transistor and measuring the resulting drain current, see Fig. 2.3.1. This simple measurement yields the transconductance $g_m$, the on-resistance and pinch-off behaviour and maximum voltages and currents. The limits of this measurement are breakdown voltages at gate and drain and thermal restriction. As all of the DC power is dissipated in the transistor, characterising the behaviour for high voltages and high currents may result in damaged devices. This can be prevented by conducting pulsed IV measurement, operating the transistor only for short amounts of time, thus protecting the transistor by limiting the power.
dissipated in the transistor, keeping it from heating up and influencing the measurement.

Pulsed IV measurements have other benefits as well. The trapping and de-trapping mechanisms described in section 2.2.3 can only be observed in an dynamic environment. Therefore, pulsed IV measurements can be used to characterise transistors for switching applications under realistic conditions [116]. It can also be used to predict how a PA performs for different supply voltages and thus under supply modulation [117]. By changing the pulse durations and introducing pre-pulses separated from the measurement pulse by a varying amount of time, a comprehensive characterisation of the device’s trapping behaviour can be established [118]. Pulsed IV measurements can be combined with RF measurements to allow the measurement of trapping under conditions typically encountered in radar applications [119].

2.3.2 S-parameters

Scattering parameters or s-parameters are the most common RF measurements due to the fact that they can be conducted with a fixed, well-defined load and do not require short or open circuits, both of which are hard to realise at high frequency. Additionally, transistors and amplifiers might run into stability issues if presented with highly reflective loads. In a s-parameter measurement, one port of an n-port network is excited at a single frequency and the magnitude and phase of the resulting signal are measured at all n ports, the ports are terminated with the system impedance during this process; this measurement can then be repeated for multiple frequencies to cover a range of frequencies, in the case of networks with more than one port, the measurement needs to be conducted for each port. This measurement yields information about input and output impedances as well as magnitude and phase of transmission between ports. However, it only measures linear behaviour, non-linearities such as the generation of harmonic frequencies will thus not be measured limiting its use in PA measurements. S-parameters are widely used in modelling however, establishing the transistor’s behaviour while it is operating linearly which allows easy extraction of the intrinsic components such as capacitances, resistances and inductances, enabling the small-signal modelling which can also be the starting point of a large-signal model.

2.3.3 Load-Pull

Output power, efficiency, gain and linearity of a power amplifier depend on the impedance the transistor is presented with. While IV measurements allow calculations yielding suitable impedances, they are not sufficient for PA design if the frequency of operation is high enough for the impact of the transistor’s intrinsic elements to be non-negligible. Once this happens, linear and non-linear capacitances complicate the PA design as the impedance the transistor has to be presented at the current generator plane now differs from that that needs to be presented at the device plane. This
Figure 2.3.2: Load-pull measurement results of a Wolfspeed CGH40100F at 2 GHz

is further exacerbated if the transistor is packaged, making conclusion about the impedance at the current generator plane only possible after significant characterisation effort. As a result, an empirical approach is proposed [120], varying the load impedance systematically to cover an impedance region, measuring the transistor’s behaviour at each measured impedance. As the load is “pulled” to different impedances, this measurement is called load-pull. Depending on the measurement system, measurement parameters such as output power, efficiency, gain and linearity can be characterised for the individual impedances. Using interpolating algorithms, contours of the measured values can be calculated, as shown in Fig. 2.3.2. These contours are often refereed to as circles. Plotting multiple superimposed circles provides an easy way to determine impedances that should be targeted during the design process as they graphically show for example areas of high power, high efficiency and where they overlap. In Fig. 2.3.2, the impedance 17+j*10Ω would offer a good trade-off between output power and efficiency while the impedance of maximum power would be in the centre of the 40 dBm circle, around 17+j*4Ω. Load-pull is not limited to the fundamental frequency; depending on the setup, it also allows putting the transistor into arbitrary multi-harmonic impedance environments, e.g. the ones necessary for high efficiency classes such as class BJ or class F [121]. Load-pull can also be used to characterise some of the trapping effects present in GaN HEMTs [122] using a measurement called fan-diagram in which the load impedance is swept along the real axis for different supply voltages, driving the transistor into saturation at each point establishes the limit of the IV curves at the respective bias voltage, as shown in Fig. 2.3.3. The first load-pull setups varied the load mechanically and mechanical load tuners are still in use today, they have been replaced by active load tuner in
many applications, however, as active tuners can be tuned significantly more quickly due to their lack of moving parts \[123\], \[124\]. They use amplifiers to inject amplitude and phase controlled signals into the transistor, allowing the generation of arbitrary load impedances, including load impedances on the edge and outside the Smith-chart, areas unreachable by passive load pull. Fig. 2.3.4 demonstrates such a commercial system, allowing the acquisition of multi-harmonic load-pull data. Here the impedances are measured by measuring the incident and reflected waves using a Keysight PNA-S series network analyser. As the network analyser is only able to measure one frequency at a time, the system includes additional comb generators phase references, enabling the individual measurements at all the harmonics to be phase-aligned.

### 2.3.4 Modelling

Creating models from measured IV, s-parameters and load-pull data is very advantageous. While the measurement is only conducted at a fixed number of points, a transistor model can be used at arbitrary points, especially useful as the probability of the measurement points coinciding with the optimum for a specific goal is very small. Models ideally allow full interpolation and, depending on the model, a certain amount of extrapolation.
Small-signal modelling

Small-signal models represent the transistor if it is excited with small-signals, hence the name. Under small-signal excitation, the transistor is linear, i.e. all capacitances, resistors, and current sources are of a fixed value [125]. Depending on the application, this might be either a suitable as the transistor is predominantly operated in that region, e.g. in gain blocks or low noise amplifiers, it can also be the starting point for a large-signal model. Small-signal models can be established by combining a number of measurements under different bias conditions [126], [127]. The small-signal model is static in that all of its intrinsic component values are fixed. In real transistors, however, the values of the intrinsic elements change with the applied voltages [128]. This can be either dealt with by having a number of small-signal models for different bias points or by modelling these non-linear intrinsic elements and using them as a starting point for a large-signal model.

Large-signal modelling

In PA design, large-signal models are necessary due to the significant amount of non-linearities that appear once the PA is driven into compression. Depending on the device technology, trapping effects need to be represented in a model as well. Often, large-signal models (LSM) are divided into behavioural models and physical models. Behavioural models are generated by fitting measurement data to equations, an example of this is the Cardiff model where load-pull data can be used to fit coefficients, resulting in a model that is able to interpolate accurately even with few measurement points [129]. Physical models extend the small-signal models, the degree of the extension depends on the desired model accuracy and application. The current source model needs to be changed to a more complex one which can take all applied voltages and the knee-region into account [130]. As the temperature has a significant influence on the behaviour of the transistor, a comprehensive large-signal model needs to consider the junction temperature, both statically and dynamically [33], [131], the trapping effects are another example of an additional effect that can be added to a physical model [132], [133]. The intrinsic capacitances [133] and resistances [134] also depend on the applied voltages and, partially, temperature [133], requiring their characterisation over these parameters. While some of these can be modelled using the equations that govern transistor behaviour, e.g. by utilising the surface potential [135], most of these effects are measured and fitted to equations. As a result, even physical models are, at least partially, behavioural, blurring the line between behavioural and physical model. As transistors are often routinely scaled into different periphery sizes by adjusting the gate width and paralleling multiple fingers, having models that are able to scale is desirable as it reduces the number of models that need to be generated. While paralleling models for the individual fingers and linearly scaling the gate width can deliver a first approximation of the behaviour of a device, not all interactions can be modelled in this simple approach. Scalable models allow use of one model for
2.4 Power Amplifiers

As signals typically cannot be generated at a power level sufficiently high for mobile communications, power amplifiers are used to increase the signal power levels to the level required for reliable communication. Mobile communication takes place on multiple bands so amplifiers in both base station and handset need to be able to operate in those. Especially with the multitude of frequency bands that have been introduced with 4G, and the ones proposed for 5G, see Fig. 2.1.2, devices have to have a number of amplifiers to cover the different bands, very wideband amplifiers or a trade-off between those two. Additionally, the PAs have to be able to cover the whole modulation bandwidth, a task made significantly harder by the increasing bandwidths Fig. 2.1.1. At the same time, PAs are expected to be as efficient as possible as they are one of the major sources, if not the dominant source, of dissipated power in communication equipment [6]. These losses are costly in terms of energy and have a significant environmental impact, making efficient power amplifiers a goal from multiple perspectives. As mentioned in section 1.1, higher efficiency will also serve to reduce the average junction temperature of the power transistors, increasing the amplifier’s mean life-time and thus the system reliability. Another result of high efficiency is its impact on communication equipment not connected to the grid such as handsets and off-the-grid base stations. Lower power consumption in these cases can lead to longer battery life and to off-the-grid base stations that are easier to realise, offering connectivity to previously isolated communities and areas.

2.4.1 Efficiency

In mobile communications, efficiency is an interesting metric in a number of ways. Spectral efficiency describes how efficiently the spectrum available to a carrier is used. There are many different ways of quantifying this, most build onto the form of data rate per bandwidth [5]. Efforts to increase this efficiency influence the design of the modulation standard and have thus a direct impact on the power amplifier. In the field of power amplifiers, efficiency has another definition, however. It uses the general definition of energy efficiency, i.e. the ratio of utilised power and consumed power, and gives a metric for the proportion of power used for the intended purpose. In power amplifiers, there are multiple definitions for efficiency. The drain efficiency

\[ \eta_{\text{Drain}} = \frac{P_{\text{RF}}}{P_{\text{DC}}} \]  (2.1)
describes how much of the DC power that enters the amplifier is converted into RF power. As amplifiers by definition amplify, they need an input signal. Especially for RF amplifiers where the gain of the amplifier is low, this input signal can have a significant power. Power added efficiency (PAE) is an alternative definition of efficiency that takes this into account. It takes the input power into account, see (2.2), and therefore yields lower values than the drain efficiency; \( \text{Gain}_\text{Linear} \) refers to the gain measured linearly as opposed to in dB.

\[
PAE = \frac{P_{\text{RF, out}} - P_{\text{RF, in}}}{P_{\text{DC}}} = \frac{P_{\text{RF, out}}}{P_{\text{DC}}} \cdot (1 - \frac{1}{\text{Gain}_\text{Linear}})
\]

The gain term demonstrates that the PAE depends on the definition of the gain used for the calculation, using the power gain will always yield a higher PAE than using definitions that include input mismatch, such as the transducer gain \([95]\). In many parts of this work, both input matching and gain are of secondary importance, therefore the term efficiency will mean the drain efficiency unless stated otherwise.

### 2.4.2 Power Amplifier Bias Points

A major design choice in PA design is the bias point. By selecting the gate bias voltage, the quiescent current, the DC current flowing through the transistor with no signal applied, is chosen with large consequences in terms of PA behaviour. The impact of the gate voltage on a PA is often described in a simplified manner by reducing it to the conduction angle \( \varphi_c \). This angle expresses which portion of the input signal is above the threshold voltage of the transistor at full power with \( \varphi_c = 360^\circ \) meaning that the input signal is above the threshold voltage at all times, while at \( \varphi_c = 180^\circ \), the input signal is below the threshold voltage half of the time, resulting in a half-wave rectified sine wave. As mentioned in \([12]\), in conduction angle classes, all harmonics with the exception of the fundamental are assumed to be perfectly short-circuited leading to a sinusoidal drain voltage as shown in Fig. 2.4.1. The figure

![Figure 2.4.1: Comparison of PA classes of operation (a) loadlines at the current generator plane and (b) voltage and current waveforms at the current generator plane](image_url)
Figure 2.4.2: Output power, efficiency and gain over conduction angle $\varphi_c$ at maximum drain current

shows the normalised voltage and current waveforms of the different PA classes at the current generator plane. In class A, current is conducted at all times while the other classes have varying portions where no current is flowing. This time where the transistor is pinched-off increases as the conduction angle is reduced and the class is changed from class AB where the $\varphi_c$ is between 360° and 180°, to class B at which $\varphi_c$ equals 180°, to class C with $\varphi_c < 180°$. To further simply the analysis, the transistor is assumed to work as an ideal current source. Under these conditions, the complete PA behaviour can be described using simple equations as discussed in [12], these equations are derived and discussed further in [136, p. 10]. These equations show how output power, efficiency and gain change with conduction angle, assuming the PA is driven to the same maximum current. Class A has the highest gain and the lowest efficiency, so gain decreases and efficiency increases as the conduction angle is reduced. In class AB, the output power is increased slightly over class A at the mentioned cost of gain and efficiency. Class B has the same output power as class A at a 6 dB lower gain and a significantly improved efficiency. Class C PAs have a even higher efficiency, the cost however is output power and gain; at a conduction angle of 90°, the gain is reduced by almost 20 dB, the achievable output power is reduced by a third.

Effect of drive level

These simple equations used so far show how the PA behaves with maximum voltage and current. If the input power is reduced however, both the DC and RF current will decrease and, assuming a constant load impedance, so will the RF voltage with the DC voltage staying constant. As a result, the RF power decreases more rapidly than the DC power, resulting in a reduced efficiency as shown in Fig. 2.4.3. This reduced efficiency as a result of the constant DC voltage already hints at the idea of supply modulation, which will be treated further in section 2.4.7 and chapter 3. The figure also shows that in the case of an ideal class B PA, the gain, i.e. the ratio between input and output power,
stays constant, as is the case in an ideal class A PA. In class AB and class C PAs however, the gain changes with drive level which introduces a non-linear effect. Class AB PAs with their conduction angle ($\varphi_c$) between $180^\circ$ and $360^\circ$ have that angle only at full output power where both voltage and current swings are at their maximum. If the drive level is reduced, the current swing reduces as well, lifting the minimum of the current waveform which will also move the intersection points with the threshold voltage, decreasing $\varphi_c$ and increasing the gain until it reaches the gain of the class A PA. For small signals, the conduction angle and the gain of ideal class AB PAs and class A PAs are therefore identical. The gain changes for class C PAs as well. As the drive level is reduced, the portion of the signal above the threshold voltage gets smaller resulting in a reduced $\varphi_c$ and gain. For small signals, the gain of ideal class C PAs is therefore zero. If the drive signal is increased from the level required for full voltage and current swing, all classes start to move to a conduction angle of $180^\circ$ and a rectangular current waveform.

**Generation of harmonic frequencies**

As the conduction angle changes, the waveform changes with it. In class A, the current is perfectly sinusoidal as shown in Fig. 2.4.1(b), reducing the conduction angle means that a part of the sine wave is truncated. The truncated current waveform has now additional frequency components, the level of these harmonic components depends on the conduction angle, as Fig. 2.4.4 shows. In conventional PAs, the drive signal is the only origin of harmonic frequencies, if they are required, e.g. in high efficiency PA classes, the bias point needs to be selected in a way that allows the transistor to generate the required harmonics. As expected, there are no harmonic frequencies generated in a class A PA.

**Realising conduction angle PAs**

While the previous conduction angle analysis gives a good indication of PA behaviour, it is not directly applicable to real PAs. This is the result of the simplified current source model used to describe the transistor. While class A is technically easy to realise, its quiescent current of $I_{\text{max}}/2$
Figure 2.4.4: Magnitude of harmonic frequency components of the drain current over $\varphi_c$ of ideal conduction angle PAs, normalised to the maximum fundamental current at class A

and the resulting losses make it infeasible due to thermal constraints, which is rarely an issue as the low efficiency usually makes it impractical for most applications anyway. The threshold voltages in real transistors are not as well-defined as they are in theory, and the same applies for the non-linear transconductance. Additionally, the threshold voltage depends on temperature and supply voltage, making it impossible to set the conduction angle to an exact value. As class B is defined by an exact conduction angle, realised class B PAs are typically biased in class AB, although the bias point can be kept close to the ideal class B by optimising the bias voltage for the minimum third harmonic generation, one of the characteristics of class B, see Fig. 2.4.4.

Class C PAs are not only lacking in terms of output power and gain, they are also harder to realise. Due to manufacturing tolerances, the threshold voltage varies, even within transistors of the same type, an effect that can not be included in transistor models. In other classes, with the bias point above the threshold voltage, the simulated bias point can be reached in a realisation by ensuring that the quiescent current of simulation and realisation are equal, something impossible of the quiescent current is zero. As a result, getting a good agreement between simulation and measurement requires more trial and error. Another factor that makes class C PAs hard to realise is the gate voltage. As the average voltage, i.e. the bias voltage, in a class C PA is lower than it is in other classes, the amplitude of the gate signal needs to be higher to achieve the maximum current. As a result, the gate voltage can become significantly more negative in class C operation, an operating condition not supported by all transistors. That being said, class C PAs play an important role in amplifier topologies where two amplifiers need to operate over different power ranges such as the Doherty PA or the Sequential PA which will be discussed later in this section.
2.4.3 Impact of the Knee Region

The analysis so far assumed the transistors to have no knee region, which is not the case in reality, as discussed in section 2.2. The knee region limits the achievable minimum voltage which means that there will always be an overlap of voltage and current, reducing the achievable efficiency. Additionally, this inaccessible voltage range limits the achievable voltage swing, in turn reducing the achievable output power of the PA. Figure Fig. 2.4.5 demonstrates how this effect impacts power and efficiency for different conduction angles. It demonstrates that achievable output power and efficiency can be easily calculated by multiplying the ideal values with 1 minus the size of the knee region relative to the maximum voltage, i.e. 0.9 in the case of the knee region being 10% of $V_{\text{max}}$, 0.8 in the case of a knee region of 20% $V_{\text{max}}$. Even for complex high efficiency modes, this simple way of calculating maximum output power and efficiency holds true.

2.4.4 Matching

As discussed in section 2.3.3, gain, output power and efficiency of the PA depend on the impedance the PA is presented with, depending on the requirements of the PA, a trade-off between these parameters is made to pick a suitable load impedance. In a load-pull measurement, this impedance is generated and presented by the load-pull system, in a PA however, this impedance needs to be presented to the PA passively. Typically, the output in Fig. 2.4.6 is assumed to be connected to a broadband 50 $\Omega$ load.

The role of the matching network is to transform the load to the $Z_L$ established as a suitable load impedance. The same procedure is required at the input of the PA. The system impedance needs to be transformed to the $Z_S$ the transistor requires for suitable operation. Matching the input and output at one frequency is trivial [95], matching it over a wider bandwidth, however, is not. Additionally, the matching network needs to not only work in the design bandwidth, its behaviour at the harmonics...
of the design bandwidth is important as well. One example for this is the conduction angle PAs, in which the voltage is assumed to be sinusoidal as a result of a perfect short circuit at all harmonics higher than the fundamental. Other examples will be presented in the following sections on efficiency enhancement. As the impedance trajectory the matching network needs to follow with frequency becomes more complex, several strategies have been developed to solve this problem and provide ways to generate wideband matching networks. Different approaches such as the real frequency technique [137], with its variant, the simplified real frequency technique [138], bounded performance technique [139], Bayesian optimisation and stochastic reduced order models can be used to generate the OMN once the target impedances are established [140]. All of these algorithms have one thing in common; they require the use of simulation tools such as Keysight ADS.

2.4.5 Efficiency Enhancement

Independent of amplifier class, topology or technology, the achievable efficiency depends on the ability to reduce losses in the circuit and in PAs, the main losses are in the transistor. Whenever there is an overlap between voltage and current i.e. the voltage waveform and the current waveform are both non-zero at the same time, power is dissipated. As discussed in section 2.4.3, the knee region will unavoidably lead to losses as the voltage can not reach zero, making voltage and current overlapping at all times. There are however ways to reduce that overlap and increase the efficiency. If a class B PA with its efficiency of ideally 78.5% is evaluated, the overlap can be easily seen in the waveforms, shown as red area in Fig. 2.4.7. The voltage waveform in the class B PA is a sine wave which only touches zero in one point. By replacing that sine wave with a square wave, as shown in Fig. 2.4.8, the PA is turned into a class F PA [12]. Here, the overlap is reduced to zero and the PA is ideally 100% efficient. Turning the sine wave of the class B PA into a square wave requires an infinite number of frequency components to be in the right phase [141]. However, even controlling a limited number of frequency components improves the efficiency, controlling frequencies up to the 5th harmonic will give up to 92%, controlling up to the 3rd harmonic 88% efficiency in the case of the maximally flat waveforms [142] that are shown in Fig. 2.4.8. As real transistors have a non-zero knee region, the realisable class F PAs differ slightly from the theoretical ones [143]. Class F is only one example for a combination of voltage and current waveforms that provide high efficiencies, there is a wide range
of solutions that provide a voltage/current combination with no overlap, some of them established as their own amplifier class, e.g. class D [144] or class E [145], [146], with distinct waveforms.

Having no overlap between voltage and current is not necessarily enough however. By analysing the spectrum and calculating the voltage and current components at the different harmonics, it is possible to establish the impedances the OMN needs to present to the transistor to allow the voltage to take the desired shape. If the harmonic impedances are highly reflective, no power is dissipated. Some of the solutions which have no overlap require resistive impedances, resulting in the PA dissipating power at higher frequencies, reducing the achievable efficiency; one such case is that of both voltage and current waveforms being rectangular. If the calculated harmonic impedances are negative, power needs to be injected into the PA to achieve the desired waveforms. In one example both the voltage and current waveform look like a half-wave rectified sine waves, resulting in required negative even harmonic impedances, i.e. energy needs to be inserted into the PA on the drain side at these frequencies in a scheme called harmonic insertion [147]. While the voltage waveform can be shaped by terminating the harmonics, the current waveform only depends on the drive signal and, in real PAs, on the interaction with the boundaries of the IV characteristics, i.e. knee region and maximum current. This leaves shaping the input voltage as a tool to shape the current waveform. This shaping of the gate
voltage can be done by terminating the gate impedance using a specially designed input matching network [148], in this case, the phase of this termination is critical in achieving high efficiency [149], [150]. Additionally, injecting second harmonic at the input has been shown to not be beneficial in all amplifier classes; while class F benefits [151], class $F^{-1}$ shows a reduced output power at a comparable efficiency [151]. Another way is using a feedback loop that works exclusively at the harmonics of interest [152]. In research, being able to modify the injected signal can be advantageous, so using a frequency triplexer to supply the gate with the different harmonics can lead to insights [153].

2.4.6 Efficient Operation over Wide Bandwidths

PAs are expected to cover wide frequency ranges as this allows them to be used to cover more than one of the bands shown in section 2.1. So far, all the analysis has assumed a single frequency of operation, however. Generating a matching network for a high efficiency PA mode, e.g. class F with its short circuits at the second harmonic, and open circuits at the third harmonic, is a task that becomes increasingly difficult as the bandwidth increases, realising for example a class F PA over a wider bandwidth is very challenging, depending on the bandwidth the PA needs to cover.

One way to solve this is by identifying optimum impedance regions over frequency in separate measurements and designing a matching network to track those. This can be done by adding matching elements and optimising their parameters, either manually or by an optimisation algorithm [153], [154], achieving bandwidths over an octave [155]. Continuous modes offer a more systematic approach to achieving wide bandwidths of high efficiency. The idea behind them is that there are multiple combinations of impedances that offer waveforms delivering the same output power and efficiency [156].

Using the example of a continuous class F PA, the waveforms in Fig. 2.4.9a show the voltage and current waveforms resulting from such a set of impedances. While the voltage waveform changes significantly with the factor $\gamma$, the output power and efficiency stay constant in this case. The impedances plotted in Fig. 2.4.9b show how both the fundamental and the second harmonic impedance move with $\gamma$. By assigning frequencies to $\gamma$ values, the necessary impedance trajectory can be described. These impedances need to be transformed to the ones at the current generator plane [138], [157]. In this step of considering the transistor, clipping contours can be used to predict impedance areas to avoid with the higher harmonics [158]. Another consideration in realising the PA is the bias point. In addition to output power and efficiency, the impedances in a continuous mode class F PA depend on the conduction angle [159].

Class F is only one example of a continuous mode, there are also examples of class $F^{-1}$ [160], [161], class E [162], [163], class J [164] and class BJ [165], [166]. As the bandwidth increases, the fundamental at the higher end of the band moves towards the second harmonic of the lower end of
the band. By having a PA that works in class F or class F\(^{-1}\), depending on the frequency, the PA can be designed to be efficient and cover more than an octave [167], [168]. While continuous mode PAs provide an impedance environment that offer flat output power and efficiency over frequency, the continuous mode impedances are not the only set of impedances providing that environment. A more general approach to waveform engineering gives the designer more freedom by using a wider set of equations and an iterative algorithm [169].

2.4.7 Efficient Operation over High Dynamic Ranges

As discussed in section 2.4.5, operation in output power back-off (OBO) leads to a reduced movement on the load-line and thus a voltage swing that does not reach the knee region, resulting in a reduced efficiency. To circumvent this, the operating conditions of the transistor need to change with the input signal. This input power dependence shows that to achieve high back-off efficiency, matching networks made of linear elements can not be sufficient on their own. Non-linear elements can achieve this,
changing the operating conditions depending on the input power. There are two main approaches to attain the full voltage swing [170] and thus the full efficiency at reduced output powers, see Fig. 2.4.10. Taking the example of 6 dB OBO, both voltage and current amplitude are half of their maximum, resulting in a quarter of the maximum power, i.e. 6 dB. With the load impedance presented to the transistor increased by a factor of two, the voltage swing will reach its maximum, reaching full efficiency, and the current will be a quarter of the maximum, resulting in the same output power at higher efficiency. Another way of achieving high efficiency is by reducing the supply voltage to make sure the voltage swing reaches 0 or, in a real PA, the knee voltage.

### Shifting load-line at constant load impedance

In conventional power amplifiers the load impedance is fixed by the load and the output matching network. Shifting the load-line by changing the supply voltage can restore the efficiency of the PA in OBO by moving the minimum voltage close the knee region, see Fig. 2.4.11. This is achieved by replacing the simple fixed DC drain bias voltage with a variable DC supply, see Fig. 2.4.12. This supply modulator changes the drain bias voltage of the PA dynamically, in response to an input signal that relates the power amplifier’s input signal to the required drain bias voltage, the shaping function. There are two fundamentally different ways to operate a supply modulated PA.

Envelope elimination and restoration (EER) was first proposed in 1952 as a way to efficiently transmit single-sideband signals [171]. In this approach, the amplitude modulation is separated from
the phase modulation, allowing the RF amplifier to be driven with a constant amplitude, phase modulated signal and using the supply modulator to determine the amplitude, i.e. the instantaneous output power. This mode of operating the transistor in a compressed state has been called “C-mode” [11], see Fig. 2.4.13. In this region, the output power exclusively depends on the supply voltage. This means that the supply modulator has to operate very accurately to fulfil the stringent linearity measures required by modern and future communication systems, making it more practical in applications with reduced linearity requirements [172]. Additionally, the operation in deep compression requires high gain to keep the high gain compression from negatively impacting the PAE.

While envelope tracking (ET) uses a modulated supply voltage as well, it follows a fundamentally different approach. Instead of operating in compression, it keeps the transistor from reaching compression, operating it in the so called “L-mode”, where the output power solely depends on the input power, as shown in Fig. 2.4.13. As a result, the exact value of the supply voltage is irrelevant for the operation of the PA, the only impacted parameter is the efficiency. The “L-mode” region represents the saturation region of the transistor where the drain current exclusively depends on the gate voltage. The “L-mode” region ends when the PA starts to be efficient, it does not coincide with the region of high efficiency, limiting the achievable efficiencies. In a real transistor, the different modes are not as easily differentiated as they are in theory, if they are clearly defined at all. In the example
in Fig. 2.4.13, the “P-mode” region in which the output power depends on the supply voltage and the input power, includes all of the Booth chart, making operation as an ET PA according to the strict definition impossible [11]. Because of that and the necessity for a slight compression to achieve competitive efficiencies, realisations of envelope tracking PAs usually operate in a hybrid mode [173].

While supply modulation is a conceptually very simple approach, there a several limitations and challenges. One limitation lies within the transistor itself. The previous discussion assumed that the achievable efficiency stays constant when reducing the supply voltage. This is not necessarily the case however, as with a constant knee-region, the ratio between knee-voltage and supply voltage determines the achievable PAE, as discussed in section 2.4.3. Additionally, the non-linear intrinsic capacitances lead to a change in optimum load impedance. Load-pull measurements demonstrate this as both output power and efficiency contours move with supply voltage. Knee region and non-linear
capacitances therefore limit the efficiency that is achievable by supply modulation. Another limitation is the supply modulator, which needs to be able to supply a precise voltage waveform at high efficiencies [174]. Additionally, the supply modulator has to be able to react to the rapidly changing power levels that result in modern communication signals, leading to growing bandwidth requirements [174]. A limitation that is more relevant to the system surrounding the supply modulated PA is the generation of the input signal of the supply modulator. This stops most supply modulated PAs from being a high efficiency drop-in replacement for conventional PAs at a system level.

Developments in supply modulated PAs will be treated in more detail in chapter 3.

Modulating Load Impedance

The drain current imposed by the input signal will result in a voltage $V_D = R_L \cdot I_D$. As described in subsection 2.4.5, the voltage swing determines the efficiency of a class B amplifier. If the load impedance $R_L$ is kept to $R_L = \frac{V_{\text{max}}}{I_D}$, the voltage swing will stay constant and the efficiency will stay high and, in the ideal case of no knee region, constant. Fig. 2.4.15 shows how the load impedances of a class B amplifier have to look for high efficiency at different power levels, demonstrating how the voltage swing stays nearly constant for a wide range of current swings if the load impedance is changed. There are multiple ways to change the dynamic load impedance of an amplifier, passively, using elements such as varactors or switches, and actively, using other transistors.

In passive load modulation, the impedance presented to the transistor by the passive output matching network is changed. This can be achieved by changing inductance and capacitance values in the OMN circuit. There are multiple ways of doing this. One way is using switches that connect and disconnect components [175]. Switching between impedances depending on the input power range will introduce distortion at the switching thresholds with the level of distortion inversely proportional to the number of states. Continuously adjustable components, such as BST varactors [176], [177], or varactor diodes [178] allow a step-less control that does not suffer from impedance discontinuities.

Figure 2.4.15: Class B load-lines for different normalised load impedances, 0.25, 0.5, 0.75 and 1
They can keep the efficiency plateau steady for 3 - 5 dB [176], [179] and may additionally increase the bandwidth of the power amplifier [179], [180]. By using non-linear elements such as diodes, the load of the PA can be modulated dynamically [181]. The presence of the non-linear elements might come at the cost of linearity, however.

The amplifier can also be load modulated actively, keeping the load impedance in the desired range and therefore the efficiency high over a high output power range. In such cases multiple amplifiers share a single load. The well-known architectures that are used in this context are the Doherty power amplifier, the out-phasing transmitter and the load modulated balanced amplifier (LMBA). While the premise of a common load is the same for these architectures, they differ significantly in approach.

**Doherty Power Amplifier**

The Doherty PA (DPA) is a active load modulation architecture first proposed in 1936 to increase the efficiency of amplitude modulated signals [182]. In it, a main, or carrier, amplifier is kept in an efficient mode of operation by a second amplifier, the peaking or auxiliary amplifier, see Fig. 2.4.16 The load modulation in the DPA only takes place during a part of the output power range, see Fig. 2.4.17, in a traditional DPA, this high efficiency range, or plateau, spans a power range of 6 dB [183], in asymmetric DPAs, this can extend further [184]. This plateau is the result of the DPA’s operation: For low power levels, the peaking amplifier is turned off, e.g. by being biased in class C, and the main amplifier operates as a single class AB or class B PA. As the main PA approaches saturation, the peaking PA starts to operate and the two output powers add in phase at the load. As the two PAs are not isolated, they interact. While the physical load impedance stays fixed, the load impedances presented to the two amplifiers change, and with them, the relation between voltage and current, see Fig. 2.4.18. The current of the main PA continues to rise linearly with input voltage but as the power injected by the peaking amplifier increases, the load impedance keeps reducing, keeping the voltage swing at the maximum level and the efficiency at it’s highest. As the efficiency of the DPA as a whole
is determined by both main and peaking amplifiers, the efficiency reduces once the peaking amplifier starts operating, as it is not operating at it’s maximum voltage swing. This leads to a 'dip' in efficiency right after the peaking amplifier starts working, at the transition point.

The analysis so far assumed that main and peaking PAs are of equal size and have thus the same maximum output power. To be able to amplify the high PAPR signals of modern communication systems efficiently, a high efficiency range of 6 dB is insufficient. By increasing the size of the peaking device, the transition point can be shifted towards lower powers, extending the high efficiency plateau of the Doherty PA [185]. This comes at the cost of a deeper efficiency dip, however: The reduced efficiency of the peaking PA now dominates since the peaking PA is larger, see Fig. 2.4.17, and the peaking amplifier starts of with a lower efficiency as its drain voltage is now reduced. This can be mitigated by modulating the supply voltage of the peaking amplifier [186], [187] or changing the phase between the main and peaking amplifier dynamically [188]. Another way of extending the efficiency
plateau is extending the Doherty PA by more branches to form a multi-stage DPA. In a three-stage DPA, the main PA can be replaced by a traditional DPA, extending the efficiency plateau [189], [190], theoretically by 6 dB to 12 dB [184].

There are a number of issues that have to be overcome in a practical realisation of a DPA. To set the direction of the change of impedance presented to the main PA, there has to be an impedance transformer between main PA and load. While this can be realised in different ways, [191], [192], all of these are band-limited. While this limits the obtainable RF bandwidth of the DPA, there has been a wealth of publications working on mitigating this limitation, resulting in increasingly wide bandwidths [193]–[196] exceeding one octave [197], [198].

As the peaking amplifier has to be turned off in the low power range, before the transition point is reached, it is usually biased in class C. This bias point results in a reduced gain and reduced maximum output power as discussed in section 2.4.2. This means that the transistor is not used to its full capacity while the reduced gain leads to a reduced PAE. The problem can also be solved by using an adaptive gate bias for the peaking amplifier [199]. This allows the transistor of the peaking amplifier to be turned off completely before the transition point is reached and to be biased in a way that provides high gain and output power for higher output powers. The biasing issue can also be solved by driving the two inputs separately [188], [200], allowing both transistors to be biased in class B and therefore achieving full output power and class B gain. Another approach is using multi-stage amplifiers for both main and peaking PA and operating only one of the drivers in class C [201]. This will allow the last stage to be biased in class B with the aforementioned benefits.

Another limitation lies within the individual PAs. The previous analysis assumed that the efficiencies stays constant over the impedance range covered by the load modulation, i.e. the main PA has to operate as efficiently at half the power as it does at full power, in asymmetric Doherty PAs, the high efficiency has to be maintained over an even wider power and impedance range. The requirements for the peaking amplifier are even more demanding with impedance that starts from theoretically infinite, see Fig. 2.4.18. Achieving a high efficiency over this impedance range requires design strategies that differ from the usual PA strategies [202], [203], the ultimate limit of achievable efficiency however is the transistor itself, as shown in section 2.3.3. If there is no impedance at which the desired back-off output power coincides with a high efficiency, the transistor will not be able to work efficiently over the whole power range in a DPA.

Despite these limitations, the DPA is the dominant PA architecture in base-stations due to its relative simplicity of looking like a single PA to the system, having only one input and one output, and the ability to accommodate high power levels up to 600 W PEP [204]–[206], the power level required in macro cells [6].
Out-phasing Transmitter

In out-phasing transmitters, two amplifiers driving a common load with the phase difference between them depending on the desired output power. They are each driven into compression which results in high individual efficiencies and combined with an in-phase combiner. There is a phase shift between them where \( \varphi_1(t) = -\varphi_2(t) \) which results in the individual output voltages \( S_1 \) and \( S_2 \) being:

\[
S_1 = A_0(t) \cos(\omega t + \phi(t) + \varphi_1(t))
\]

\[
S_2 = A_0(t) \cos(\omega t + \phi(t) + \varphi_2(t))
\]

The addition of the two vectors in the combiner yields the output voltage, the angle \( \varphi \) between the vectors determines which portion of the power is available at the output of the transmitter, as shown in Fig. 2.4.19. Energy conservation means that the power not available at the output needs to be reflected back into the circuit. In the out-phasing transmitters first presented by Chireix in 1936 [208], the power is reflected into the amplifiers. To the transistor, this reflection is indistinguishable from a change in load impedance. By increasing the load presented to the amplifiers by increasing the phase \( \varphi \), the amplifiers can operate efficiently over a wide output power range. Practical implementations suffer from the effect of reactive elements of the power devices, Cripps describes in [183, p.62] how the load impedances are modulated in a Chireix architecture. Fig. 2.4.20a shows one way to deal with them, using shunt reactances at the output of the amplifiers as proposed by Chireix [208]. Their values can be used to optimise the transmitter for signals with different PAPRs [209]. As discussed in [136], phase accuracy is a possible issue to achieve low power levels, [210] demonstrates that state of the art out-phasing transmitters are capable of delivering highly linear signals with PAPRs of 10 dB.

In linear amplification using non-linear components architectures (LINC) the amplifiers are combined using a combiner with isolation. The load impedance of the amplifier stays therefore constant and they operate into a fixed load. Energy conservation means that the isolation resistor of the com-
The out-phasing PA has limitations that have so far prevented its commercial breakthrough, some of those limitations have already been hinted at. Due to the reactive combiner, the out-phasing PA works only over a narrow RF bandwidth [184], [213] although recent publications have demonstrated out-phasing transmitters with 30% fractional bandwidth [214]. The second issue is the signal generation as an out-phasing PA requires two coherent RF channels that need to be controlled in amplitude and phase. Especially for low power levels, the phase needs to be controlled very tightly. By combining out-phasing and a Doherty PA, a single input self-outphasing Doherty-Chireix can be designed, its bandwidth however is not yet reported and might be a limiting factor [215]. A way around this is operating the outphasing PA in so-called ‘Mixed Mode’, in which the drive level is reduced for low power levels, significantly increasing efficiency at high OBO [216]. Another limitation of out-phasing PAs is shared with Doherty PAs and other load modulation techniques, the transistors and the fact that the efficiency contours do not extend to very low power levels as discussed in section 2.4.7. That means that even if the ideal load trajectory along the real axis was realisable, the transistors would not operate efficiently over the whole dynamic power range. This can be mitigated by using the load-pull data of the transistors to design the impedance trajectories [214] or by establishing how the impedance trajectories can be manipulated to follow continuous mode class F trajectories [217]. If the out-phasing PA is combined with a supply modulation scheme, the previous limitations are extended as the supply modulation means that the modulation bandwidth is additionally limited by the supply modulator, this reduces however the stringent phase accuracy requirements.
Load-Modulated Balanced Amplifier

Another way of actively modulating the load impedance of an amplifier is the load modulated balanced amplifier (LMBA) proposed in [218]. Instead of terminating the isolation port of the output 90° coupler of the balanced amplifier, that port is used to inject a signal into the two power amplifiers, see Fig. 2.4.21. This control signal is added to the reflected power coming from the load, which leads to a change in the load presented to the devices. As both 90° couplers and efficient power amplifiers can be realised over wide bandwidths, this potentially allows a very wideband power amplifier with high efficiencies at OBO. The control signal has to have a sufficient amplitude in order to be able to modulate the load of the two PAs, depending on the power levels, this might require an additional amplifier. While the first examples used separate drive signals for the balanced PA and the control PA [218], [219], there has been a drive for a single input PA structure since. This can be achieved using an unequal power splitter [220] and allows the LMBA to be characterised as a simple PA, albeit at a reduced broadband performance [221] and back-off performance [222]. This can be partially corrected by using a filter to to compensate the different phases required at different frequencies [223]. As with any other load modulation technique, the LMBA relies on the transistors being as efficient at lower power impedances. If the control PA is controlled externally though, the impedance trajectory can be chosen almost arbitrarily [221], which allows the optimisation of the trajectory for any transistor at the cost of a second coherent RF channel or a controllable phase shifter and attenuator.
Other Techniques

While the techniques presented so far are the dominant ones in achieving high efficiencies over wider output power ranges, there are other techniques that can be used to reach this goal.

Switching between two amplifiers

By having a number of amplifiers optimised for different power levels, a wide power range can be covered efficiently by switching between them [224]. If the two amplifiers use one of the previously mentioned efficiency enhancement techniques, the overall efficiency of the system can be kept high over very wide power ranges, e.g. 17 dB in a PA consisting of two PAs that are based on a GaAs HEMT and a GaN HEMT [224]. This comes at the cost of a higher complexity in both signal generation, combination and generation of the modulated signal. The switch at the input introduces further limitations, it could, however, be replaced if the discrete PAs had a separate input signal, further increasing complexity.

Sequential Power Amplifier

Like the Doherty PA, the sequential PA (SPA) consists of one main PA and one [225] or more [226] peaking PAs. In the SPA however, there is no load modulation, the main and the peaking paths are combined using a coupler that isolates the two PAs [12], [227]. As with the DPA, the main PA is the only part active at low power levels. As the input power increases, and the main PA starts to reach saturation, the peaking PA starts to work. As there is no load modulation, the main PA is driven more and more into saturation, reducing its linearity. The reduction in gain is partially compensated by the peaking PA. As the peaking PA is usually biased in class C, its gain increases with input power. As the two PAs operate independently, the bandwidth of the SPA is only limited by the bandwidth of the PAs and that of the input and output couplers. As a result of the isolating property, the combiner can only work without losses at the point where the power ratio of the main and the peaking PA is equal to the coupling ratio of the combiner [228]. At any other power level, the
combiner dissipates power, resulting in significant power losses. This power loss in the coupler can be reduced by dynamically adapting the coupler [225], e.g. by switching between different [228] or by using capacitors in a quadrature coupler [229].

**Digital PAs**

A very different approach to achieve high efficiency is by using a digital PA, using two switches that switch between a supply voltage and ground [230]. As the switches are at their most efficient when they conduct, the efficiency is high as long as the transition between low and high voltage is short relative to the length of the on-time. This approach requires switching speeds that significantly exceed the desired operating frequency however, by a factor of 4 in [230]. Additional limitations are the signal generation as two channels need to be perfectly synchronised and provide the drivers with very short pulses with a high resolution, especially if a modulated signal is to be transmitted.

**Comparison of high efficiency architectures**

Tab. 2.5 compares the different high efficiency PA topologies that were discussed in this section in a range of different relevant metrics. Allocating the value however is a subjective process as reported realisations rarely discuss all of these metrics and focus on a small subset, e.g. efficiency and RF bandwidth [192]. A result of this approach is that a well established technique, such as the Doherty PA, will perform better due to decades of research going into its improvement, while relatively new techniques such as the LMBA can be assumed to be far from reaching full potential. Allocating the values is done based on the limitations of the respective technology reported in literature, and while extrapolation tries to take future developments into account, this table is just a subjective snapshot at the time of writing. No PA topology is superior in all aspects, which means that the ideal topology depends on the requirements of the application. DPAs fare well in many metrics, a combination of a simple structure and and decades of research on improving it. This versatility has led to them being used heavily in many PAs in communication systems. Other topologies dominate in just one are,
e.g. outphasing and EER in efficiency, making them a good choice in applications where that metric is paramount and the disadvantages are acceptable. To decide on the suitability of a topology, the whole system has to be taken into consideration. Depending on the requirements and the available hardware, a complicated signal generation or linearisation could be implemented at little to no cost, making these metrics irrelevant in that case.

### 2.5 Integrated Power Amplifiers

Integrating circuits means that some or all of their constituent elements are fabricated on the same substrate. CPUs with their billions of transistor would be impossible to build if the individual transistors were discrete elements, due to the number of connections required and because these connections would limit the achievable performance. Additionally, such a discrete CPU would be significantly larger in size, making it infeasible for use in modern devices. The same reasons are the motivation to integrating PAs, facilitating manufacturing, increasing performance, and reducing the circuit size. As in many cases in engineering, there is a trade-off between these factors, e.g. shrinking a circuit from a discrete PA to a monolithic microwave integrated circuit (MMIC) PA will reduce the size but that does not necessarily increase the performance [192]. Having the ability to access the transistor almost at the current generator plane allows terminating baseband impedances to a degree that is hard to realise with additional components between the transistor and the [231] baseband termination. This access close to the transistor needs to be dealt with carefully, especially with a high number of transistor fingers. Managing the relative phase between their gate and drain nodes becomes challenging which can have a significant impact on the performance [232].

One of the limitations of MMIC PAs are inductors. Whether they are used for matching or as bias feeds, planar inductors have an inherently lower quality factor which results in losses, limiting the achievable efficiency [233]. This can be partially mitigated by using bondwires as part of the inductors [192]. Other techniques of increasing the quality factor include increasing the distance to the lossy substrate and elevating the inductor, further reducing the losses [234].

As the frequency is increased, the required transmission lines become short enough to become a viable alternative to inductors [235], [236]. In that case, realising low loss transmission lines becomes critical. The losses in transmission lines depend mainly on the width of the line and thus the resistance and the substrate. In the case of GaN-on-Si, the resistivity of the silicon and its temperature play an important role in the transmission losses [41].

The transistor can also be fabricated independently of the matching networks and connected in an additional assembly step. As the resulting circuit is not monolithic, this is called a quasi MMIC.
This approach has several advantages: The matching networks can be designed on cheaper material, commercial transistors can be used in conjunction with processes from other companies, and the separation of transistor and matching network can increase the yield as the resulting sub-circuits are significantly smaller. Using quasi MMICs instead of MMICs comes at little cost in terms of performance as long as the inductance of the bond-wires is either sufficiently small or can be integrated into the matching network [237]. The quasi MMIC assembly can be in a package [238] or on a carrier [237]. In addition to cost, another advantage of quasi MMICs is its flexibility, as it allows realisation of different parts of the matching networks on different substrates. By using substrates with a very high dielectric constant of $\epsilon_r = 90$, very low impedance transmission lines can be realised in a compact way [238]. As MMICs are often soldered onto PCBs, the matching networks can be partially realised on the same PCB without the need for other substrate materials [239]. A special case of Quasi MMIC is metal-embedded chip assembly (MECA) [240]. In this technique, the different parts of the IC are embedded in a metal substrate and connected using bond-wires. This allows integrating almost all technologies into a quasi MMIC, be it on GaAs, CMOS, GaN or Alumina. By growing the copper heat spreader on and around the components to be connected, a good thermal connection and RF grounding can be guaranteed [240].

### 2.6 Non-linearities and Linearisation

This section will discuss linearity in general before focusing on the non-linearities that occur in PAs in section 2.6.1. Following this, ways to improve the linearity will be discussed in section 2.6.2. To appreciate what non-linearities are, linearity needs to be treated first. Ideally, the PA would operate as a linear, time-invariant (LTI) system, i.e. the output signal $v_{out}$ would only differ from the input signal $v_{in}$ in amplitude and phase, as shown in (2.3).

$$v_{out}(f) = \hat{v}_{in}(f) \cdot A(f) \cdot \sin(\omega t - \phi_{in} - \phi(f))$$ \hspace{1cm} (2.3)

Additionally, the time-invariance would result in $A(f)$ and $\phi(f)$ being constant over time. In such a system, any input signal will result in a pre-determinable output signal. If $A(f)$ and $\phi(f)$ are not constant however, the output signal depends on their value. If $A(f)$ and $\phi(f)$ depend on the input signal, the system becomes non-linear, which results in $A(f)$ and $\phi(f)$ changing and additional frequencies being generated.
2.6.1 Non-linearities in Power Amplifiers

In communication systems, the input signal $v_{in}$ of the PA is generated by modulating one or more carriers with the baseband signal $v_{BB}$. The output signal $v_{out}$ of the PA is then transmitted, it is then demodulated on the receiving side. For this demodulation to be successful, the signal needs to be sufficiently linear. The encoding and the decoding on the transmitter resp. the receiver side include error correction algorithms, so a small amount of transmission errors can be dealt with. As this correction also needs to take distortion introduced by the channel into account, the PA is required to fulfil stringent linearity requirements. In mobile applications such as LTE or 5G, the linearity requirements are typically specified in the error vector magnitude (EVM) and the adjacent channel leakage ratio (ACLR), used interchangeably with adjacent channel power ratio (ACPR) [8]. The EVM describes the magnitude of the error vector $\vec{s}_{err}$ that is required to transform the original vector, $\vec{s}_0$, to the measured vector, $\vec{s}_m$, as demonstrated in Fig. 2.6.1. The closer the symbols are together in the IQ plane, the lower is the EVM that the system can tolerate. While a QPSK modulation only requires the EVM to be below 17.5%, a 256 QAM modulation, such as the one used in LTE or 5G, requires a significantly lower EVM of 3.5%, a further increase in symbols reduces the required EVM to 2.5 % at 1024 QAM [23]. The second important linearity specification is the ACLR, the amount of power in the frequency bands next to the one occupied by the signal itself. To ensure that communication in one band does not interfere with communication on the neighbouring channels, the signal level in the neighbouring channels needs to be sufficiently low. The limits that have been standardised as 45 dB below the carrier for most bands and cases in the base stations [23] and -10 dBm to -25 dBm in user equipment [8]. As the EVM and ACLR standards need to be adhered to, linearity is a significant parameter during the PA design. To be able to deal with these non-linearities, it is imperative to understand their origin in the PA.
In-band distortion

Fig. 2.6.1 demonstrates that the EVM depends on both the phase error, $\phi$, and the magnitude error, $\nu$. These two errors are direct results of the non-linearities of the PA. By using the power series in 2.4 to describe the behaviour of the PA [12, p.231], the introduction of additional frequency components can be modelled.

$$v_{\text{out}} = a_1 \cdot v_{\text{in}} + a_2 \cdot v_{\text{in}}^2 + a_3 \cdot v_{\text{in}}^3 + a_4 \cdot v_{\text{in}}^4 + a_5 \cdot v_{\text{in}}^5 + \ldots \tag{2.4}$$

Expanding this equation using the trigonometric identities shows that the output voltage includes a significant amount of additional frequency components. Of particular interest is here the fact that the third and fifth order terms result in a component at the fundamental frequency that counteracts the first order term, which means that the output signal can be described as $v_{\text{out},0} = a_1 \cdot v_{\text{in}} + a_3 \cdot \frac{3}{4} v_{\text{in}}^3 + a_5 \cdot \frac{10}{16} v_{\text{in}}^5 + \ldots$. The impact of these higher order terms depends significantly on the drive level. At low drive levels, the first order term dominates. The higher order terms start to become noticeable as the drive level increases, resulting in the output power deviating from that achievable with no higher order terms, as shown in Fig. 2.6.2. This reduces the gain, the gain is said to compress.

The point at which the difference between ideal and non-linear gain amounts to 1 dB is defined as the 1 dB compression point. PAs are often measured at $x$ dB compression point where $x$ depends on the application and the desired linearity. The phase distortion of PAs can be mainly attributed to the non-linearity of the intrinsic capacitances of the active device and the interaction with the knee region [243], [244]. These two effects, compression and phase change, directly result in the magnitude and phase errors, $\nu$ and $\phi$. In PA characterisations, the amplitude and phase distortion are usually specified as AM/AM and AM/PM distortion, they can be obtained from either CW power sweeps or modulated measurements.
Out-of-band distortion

The term out-of-band in this case only refers to a frequency range around the signal, commonly 3 to 7 times the signal bandwidth. While there are significant terms at higher frequencies, around the harmonics of the signal, they are usually assumed to be filtered out. As opposed to in-band distortion which can be calculated using single tones, out-of-band distortion can only be shown for modulated signals. To keep the analysis simple, a two-tone signal, $v_{2T}$, consisting of two static signals of equal amplitude, spaced apart by the frequency $f_{\text{mod}}$, which is significantly lower than $f_0$ is chosen, as defined in (2.5) where $\omega_1 = 2\pi(f_0 - f_{\text{mod}}/2)$ and $\omega_2 = 2\pi(f_0 + f_{\text{mod}}/2)$.

$$v_{2T} = v_i(\cos(\omega_1 t) + \cos(\omega_2 t)) \quad (2.5)$$

By solving (2.4) for the separate frequency component with $v_{in} = v_{2T}$, the intermodulation products can be calculated, Tab. 2.6 shows the terms up to the fifth order. As the order increases, the trigono-

<table>
<thead>
<tr>
<th>$a_1v_i$</th>
<th>$a_3v_i^3$</th>
<th>$a_5v_i^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>1</td>
<td>$\frac{9}{4}$</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>1</td>
<td>$\frac{9}{4}$</td>
</tr>
<tr>
<td>$2\omega_1 \pm \omega_2$</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{25}{16}$</td>
</tr>
<tr>
<td>$2\omega_1 \pm \omega_2$</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{25}{16}$</td>
</tr>
<tr>
<td>$3\omega_1 \pm 2\omega_2$</td>
<td>$\frac{5}{8}$</td>
<td>$\frac{25}{16}$</td>
</tr>
<tr>
<td>$3\omega_1 \pm 2\omega_2$</td>
<td>$\frac{5}{8}$</td>
<td>$\frac{25}{16}$</td>
</tr>
</tbody>
</table>

metric identities introduce more and more additional frequency components. The even order terms introduce frequencies around DC and in the even order harmonic zones, as shown in Fig. 2.6.3, so they can be filtered out with relative ease. The odd order terms also introduce frequencies in the harmonic zones which are easily filtered out, they additionally introduce frequencies in the immediate vicinity to the two original tones, however. These intermodulation products (IMPs) or intermodulation distortion (IMD) are in the fundamental zone and thus too close to the signal to be filtered out. These frequency components are often referred to by the order that introduces them, IMD3 for the ones closest to the two tones, IMD5 for the next set and so on. Typically, only IMD3 and IMD5 are considered as higher IMDS are too low in amplitude to be significant in most applications. The magnitude of the IMDS depends significantly on the selected bias point [183], [245] and the input power [183], allowing a certain amount of optimisation by aligning the PDF and the sweet spots.

Real modulation standards such as the ones used in mobile communications can be visualised as superposition of many two-tone signals [12, p.233]. As a result, the place of the discrete tones in
**Power Amplifiers for Mobile Communications - Non-linearities and Linearisation**

---

**Figure 2.6.3:** Frequency domain response of a non-linear system with a two-tone excitation after [242, p.34]

**Figure 2.6.4:** Frequency domain response of a non-linear amplifier with a modulated signal

---

Fig. 2.6.3 is now taken by frequency band, as shown in Fig. 2.6.4. The figure also shows how the compression now affects the band as a whole. Additional out-of-band distortion can be introduced by mixing products with frequency components present in the supply voltage, this will be discussed in detail in section 3.1.5.

**Memory Effects**

In addition to these effects, the PA also suffers from memory effects that can have effects both in-band and out-of-band. The term memory effect already implies that in the case of these effects, the behaviour of the PA is not only determined by the signals applied at that time but also by those applied in the past. The most common memory effects are electrical memory, thermal memory and trapping effects [12]. Trapping effects are common in HEMT devices as discussed in section 2.2, pushing the manufacturers to conduct extensive research into ways to reduce these effects, e.g. by improving passivation and buffer epitaxy and adding field plates. As $g_m$ and the maximum current $I_{\text{max}}$ of the transistor change with temperature [246], the thermal state of the transistor has a significant impact on the PA performance. As the time constant of typical transistors means that they do not reach a thermal steady state [12], [131], both $g_m$ and $I_{\text{max}}$ change over time, leading to additional distortion. The third memory effect is the electrical memory, in which the instantaneous supply voltage is not only...
a result of the applied voltage from the power supply but additionally has components as a result of the drain current being drawn. The electrical memory effect is a result of the baseband impedance being non-zero in the frequency band in which the PA draws current, it manifests itself as an asymmetry of the IMD bands [12]. It can be controlled by the bias network [12], [247] in conventional PAs, its impact on supply modulated PAs will be treated in more detail in section 3.1.6.

2.6.2 Linearisation

The non-linearities described in the previous chapter distort the output signal of a PA which poses an issue if the PA is required to comply with the linearity standards. If the PA can not fulfil the linearity requirements, it needs to be either backed-off, reducing the available output power and efficiency, or linearised, i.e. the non-linearities introduced by the PA need to be reduced to ensure that, as the signal reaches the antenna, it fulfils the requirements. Over the decades, three main types of linearisation have been established, feedback, feed-forward and pre-distortion. Feedback uses a small part of the output signal of the PA and feeds it back into the input of the PA. By adapting the processing between the measurement and the insertion into the input, the feedback loop can be optimised. This feedback loop can operate at baseband frequencies or at RF frequencies [183], it can also operate at multiple frequencies, e.g. at RF frequencies to linearise in-band and out-of-band distortion and, additionally, at other frequencies, e.g. to mitigate the electrical memory effect. The feed-forward approach splits the signal, processes it and reintroduces it at a later point in the transmission chain. Often, the signal is split at the input of the PA, passes through an error PA which introduces non-linearities which are the opposite to those of the PA, and added to the output of the PA where the non-linearities cancel. Pre-distortion uses knowledge of the PAs behaviour to craft a distorted input signal that, once amplified by the PA, turns into the desired, ideally linear signal. Fig. 2.6.5 shows a simple example of this using an arbitrary function. Here the PA characteristic (red) was defined, by replacing the linear input signal with the pre-distorted version of the input signal (green) and the PA non-linearities cancel out, here represented by the blue dotted line, and result in a perfectly linear relationship between input power and output power. As the pre-distorter needs to describe a non-linear input signal, it needs to be realised using either a non-linear device, such as a diode or a transistor [248], on the input, or digitally. As in modern mobile communication systems, the signal generation is realised digitally anyway, most of these system also use digital pre-distortion (DPD). To linearise the PA, the DPD needs access to all of the distortion it is to linearise. This means that the DPD needs a bandwidth significantly larger than the baseband bandwidth by a factor of 3 - 5 [170], if the available bandwidth is too low, the DPD will not be able to fully pre-distort the signal [249]. This is particularly problematic as baseband bandwidths are expected to increase with the introduction of mmWave communication.
systems. Analogue pre-distorters are not limited in that way, but by the circuit design as they need to exhibit the correct amplitude and phase distortion. RF bandwidths of 200 MHz at $K_u$-band \cite{250} and 250 MHz at $E$-band \cite{248} have been demonstrated. As pre-distorters can only pre-distort non-linearities present in characterisation, they can not take effects such as change in temperature or antenna mismatch. By adding a feedback loop and adding the information about the state at the PA output to the pre-distorter, it becomes possible to take these effects into account \cite{251}. To describe the PA behaviour, the DPD needs to build mathematical models, the stronger the non-linearities the more complex these models need to become. Elaborate pre-distortion algorithms are able to correct even strong memory effects \cite{252}. So far, only the RF input signal has been used to linearise the PA. Control over the supply voltage of the PA adds more options to linearise the PA, however, and these will be discussed in section 3.4.

\section{Conclusions}

As mobile communications developed from a niche, providing exclusively audio communication, to mainstream high speed data communications, power amplifiers keep being developed to be able to support those increasingly challenging applications. This requires a combination of a wide range of research areas, from epitaxy and crystallography over device physics to circuit design, resulting in a wealth of research all geared towards fulfilling the requirements of modern communication standards. This chapter provided an overview of the research conducted in the many areas relevant to the work, particularly focussing on GaN HEMTs and highly efficiency power amplifiers. GaN HEMTs have turned out to be a highly competitive technology for the higher frequencies at which the drive for higher data rates has mostly taken place, this is reflected in the wealth of research going into it on all levels which has improved device performance, manufacturing yield and cost. As the modulated
signals became more dynamic, the research into optimising power amplifiers for these signals took off. A sizeable portion of that research was focussed on Doherty PAs due to its relative simplicity with supply modulation slowly gaining traction. While GaN HEMTs have been used in this research, there is still a gap in terms of what makes GaN HEMTs suitable for use in supply modulated PAs.

References


Power Amplifiers for Mobile Communications - Conclusions


Power Amplifiers for Mobile Communications - Conclusions


Power Amplifiers for Mobile Communications - Conclusions


Power Amplifiers for Mobile Communications - Conclusions


Power Amplifiers for Mobile Communications - Conclusions


Power Amplifiers for Mobile Communications - Conclusions


Power Amplifiers for Mobile Communications - Conclusions


Power Amplifiers for Mobile Communications - Conclusions


As discussed in section 1.1 and section 2.4.7, supply modulation (SM) schemes such as envelope tracking work by modulating the supply voltage of a PA in response to instantaneous the power level. While this concept is incredibly simple, realising it is anything but. Unlike a simple PA with one RF input and one RF output, a SM system comprises not only the PA with RF input and output but also a supply modulator that converts a constant DC voltage to the required modulated supply voltage. Additionally, the digital signal processing now needs to generate an intermediate frequency (IF) signal as a second output signal in addition to the RF signal, see Fig. 3.0.1. These additional factors mean that the simple concept of supply modulation has not only generated a lot of interest over recent years...
Figure 3.0.1: Overview of a supply modulated PA system showing the internal power flow

but also a significant amount of publications. This chapter will present and discuss a selection of these papers to examine the different parts of a supply modulation PA. Section 3.1 starts by looking at supply modulation from a system perspective. Section 3.2 will review the different supply modulator topologies in detail, followed by section 3.3 which considers supply modulated power amplifiers and their components and section 3.4, which finally looks at linearisation approaches in supply modulated systems.

### 3.1 Supply Modulation at System Level

Figure 3.0.1 shows the different parts of a supply modulated PA system. These parts all need to work together for the system to achieve the linearity and efficiency required for reliable communication. The number of parts and the fact that the system is used to amplify real modulation signals means that the previous efficiency definition in section 2.4.1 is no longer sufficient, so section 3.1.1 will discuss efficiency in supply modulated system in more detail. Another consideration of having additional components in the system is the interaction between them, the most obvious being the modulated supply voltage. This modulated voltage depends on the modulated input signal of the PA, the relation between modulated signal and modulated supply voltage being determined by the shaping function, more on that in section 3.1.2. The shaping function has a significant impact on the baseband bandwidth of the modulated supply voltage, this is going to be discussed in section 3.1.3. Depending on the architecture, the modulated supply voltage and the RF voltage might not be in phase at the PA transistor’s drain, section 3.1.4 will explore the impact of that on the system performance and discuss ways to optimise the delay. Ideal systems would have no further interactions; in a real system however, this is not the case. Depending on its architecture, the supply modulator introduces additional frequency components, the spectrum and frequency of which depends on the modulators architecture. Section 3.1.5 introduces
the effects of those components on the PA. Additionally, the different sub-system will present non-ideal impedances to the other parts of the system and introduce new-linearities that need to be considered. Their behaviour as described in literature will be discussed in section 3.1.6 and section 3.1.7.

3.1.1 Efficiency in ET systems

For the case of a PA measured under constant amplitude excitation, the efficiency is easily calculated, see section 2.4.1. In an ET system, the efficiency of the supply modulator needs to be factored in resulting in the overall efficiency

\[
PAE_{ov} = PAE_{PA} \cdot PAE_{SM} = \frac{P_{RF,out} - P_{RF,in}}{P_{DC,in}} \quad (3.1)
\]

If the comparison between different PA architectures is to be fair, this equation has to also take the energy consumed by the circuitry around the modulator into account, a challenging endeavour since there is no clear point at which to stop including elements. One could argue that most of the digital signal processing is already in place in most systems anyway but the point of view that including the signal generation, e.g. the FPGA in many of the high bandwidth buck-converter implementations presented in section 3.2.3, is also a valid one. While this extra power is not particularly significant in macro base stations [6], it might consume more power than is gained by the efficient PA architecture, possibly making envelope tracking an impractical solution for PAs used in smaller cells, such as femto cells. Characterising the overall system efficiency yields the information necessary to make these decision.

There are multiple possibly interesting measures of PAEs in a modulated system: The peak PAE, i.e. the PAE at the maximum output power, the peak envelope power (PEP), and the average PAE. Equation 3.1 works for all of them. In literature, there is often little information about which PAE is plotted. With PAs optimised for maximum PAE at the back-off to target specific modulations, the peak PAE will be higher than that at PEP since the PA is operated at PEP for only a fraction of the time. Average PAEs may appear to be the easiest way to compare the efficiency of PAs; the fact that PAs are designed for different types of modulation, with different PAPR, bandwidth and dynamic behaviour means that the average PAE is a useful measure in determining the suitability of a PA but not more than that.

The last type of efficiency is the least relevant from a system level but the most interesting for the PA designer, the instantaneous efficiency. A time-domain representation of how the efficiency behaves over time can show whether ET works as expected, and requires the PA’s output power and instantaneous power flow into the transistor to be measured and time aligned.
3.1.2 Shaping Functions

The dynamic supply voltage needs to change with the magnitude of the input power envelope, and the function that translates input to supply voltage is the shaping function. In an ideal ET PA, the supply voltage has no impact on the PA performance [11], the voltage trajectories only differ in resulting efficiency as long as they keep the drain voltage from reaching zero. To optimise for efficiency, the drain voltage would be kept close to zero at all times. In a real transistor, the knee region restricts the shaping function by introducing a minimum voltage the drain voltage is to remain above. This leads to the introduction of a lower limit for the supply voltage, the functions implementing such limit are detroughing functions [253].

The two dominant ways of generating shaping functions are look-up-tables (LUT) based on measurements and equation based shaping functions. Equation based functions have the advantage of being easily implemented, the most common equations based shaping functions are power

\[ V_{\text{mod, power}} = \sum_{k=0}^{\infty} a_k \cdot v_{\text{envelope}}^k \]  

(3.2)

exponential [254]

\[ V_{\text{mod, exp}} = V_{\text{max}} \cdot (v_{\text{envelope}} + D \cdot e^{\frac{v_{\text{envelope}}}{D}}) \]  

(3.3)

and raised cosine function [255]

\[ V_{\text{mod, cos}} = V_{\text{max}} \cdot (1 - (1 - D) \cdot \cos(v_{\text{envelope}} \cdot \frac{\pi}{2})) \]  

(3.4)
where

\[ D = \frac{V_{\text{max}}}{V_{\text{min}}} \]  \hspace{1cm} (3.5)

The LUT approach works by measuring the PA statically for a number of supply voltage and assembling a shaping function that tracks the desirable characteristics. This can be used to design shaping functions that deliver flat gain [256]–[258], flat phase [187] or track the maximum PAE [256], [257]. The approach can also be used to design shaping functions that limit the intermodulation distortion [259], however, the trade-off between efficiency, gain and phase is not discussed here. In ET systems with a modulator that produces a significant amount of switching noise or ripple, optimising the shaping function is possible by optimising the duty cycle of the switching modulator for a minimal ripple-resultant distortion [260]. The two approaches can be combined by generating a LUT to deliver the desired characteristic and adapting the coefficients of an equation based shaping function to track the LUT values [256]. The shaping function has a direct impact on AM/AM and AM/PM distortion and efficiency [256], shaping functions optimised for flat gain have been shown to provide high linearity even without additional DPD [258].

### 3.1.3 Baseband Bandwidth

Another result of shaping functions mapping the input envelope to the supply voltage is that the baseband bandwidth occupied by the modulated supply voltage is not necessarily equal to the modulation bandwidth of the transmitted signal [256]. As the magnitude of the input signal is obtained by taking the square root of the, band limited, input signal, its bandwidth is theoretically infinite [261]. Depending on the shaping function, this infinite bandwidth can be reduced back to a band-limited baseband bandwidth, however, see Fig. 3.1.2. Power envelope tracking (PET) reduces the required baseband bandwidth by using exclusively even terms, with the baseband bandwidth reduced to \( n \) times the mod-

**Figure 3.1.2:** Simulated spectra of a 10 MHz 64 QAM signal with different shaping functions
ulation bandwidth with \( n \) being the number of even terms in the power shaping function (3.2). The magnitudes of the frequency components depends on the coefficients, see Fig. 3.1.2. The raised-cosine function has a baseband bandwidth of 3 times the modulation bandwidth, the highest third of this bandwidth is very low in amplitude though so neglecting it is unlikely to have a significant impact. The linear and the exponential shaping functions however do not reduce the baseband bandwidth, resulting in a theoretically infinite baseband bandwidth. Realistically, frequency components can be neglected if they are very small, and a magnitude of 50 dB below the main lobe has been used in [262], [263], albeit without any information provided regarding how that value was chosen. In the light of this, the baseband bandwidth of the exponential function can be considered lower than that of the linear function as it drops more rapidly.

By applying a low pass filter to the envelope signal generated by the shaping function, the baseband bandwidth can be further reduced [262]. This can, however, lead to the supply voltage dropping too far, resulting in the RF voltage reaching the knee region. This can be prevented by iteratively adapting the filtered envelope signal until it stays above the knee region at all times while keeping the reduced bandwidth [262]. The memory effects introduced by this approach can be compensated for by applying specially adapted DPD algorithms. While LUT algorithms are able improve the linearity considerably, only deterministic memory mitigation algorithms can fully pre-distort the signal. Instead of looking at the problem in the frequency domain and using low pass filters, the bandwidth can also be reduced by approaching the problem from the time domain. By limiting the slew rate of the signal, the dynamic of the signal can be reduced [264], [265]. This approach requires the system to have a delay as a reduced slew rate does not allow fast transitions from low to high; the magnitude of the delay increases with the amount of slew rate reduction. The lower the required slew rate, the easier it is for the modulator, as the low slew rate does not allow the supply voltage to drop as much a reduced slew rate reduces the achievable average efficiency of the PA [266].

If the ET PA is required to amplify multi-band signals, i.e. two frequency bands at the same time, calculating the shaping function becomes more complex since the input signal now consists of two carriers with two envelopes. One way to approach this is using a two-dimensional LUT with the two input powers as inputs [267]. Another approach regards the two input signals as one wideband signal, the bandwidth can thus be approximated as the difference between the upper edge of the highest frequency and the lower edge of the lowest frequency [263]. Tracking the sum of the envelopes instead of the sum of the modulated signals reduces the bandwidth of the modulated supply voltage the modulator has to output significantly, in the case of [263] by a factor of around 50, to a bandwidth of around twice the sum of the individual signals. By tracking the average of the two baseband signals instead of the sum, the efficiency can be increased [268], [269] at comparable linearity after DPD at the cost of DPD complexity. The cost in terms of baseband bandwidth is not reported.
Developments in Supply Modulation - Supply Modulation at System Level

If the modulator is used with a signal with a bandwidth too high for the modulator to generate, the real supply voltage will differ from the intended supply voltage, this will lead to electric memory effects [270], [271]. Using a combination of buck-converter and a linear amplifier with an adjustable frequency limitation, this effect can be evaluated systematically [272]. While the difference between desired and measured supply voltage is easily measurable and can be expressed as normalised root mean square error [270], this inability to track the supply voltage properly is rarely discussed in literature, the common comparison showing a figure with both desired and generated waveform makes it hard to tell how well a modulator is able to track the envelope and how much distortion it will cause.

3.1.4 Time Alignment

Up to this point, it has been assumed that the RF signal and the modulated supply voltage are perfectly synchronised at the intrinsic drain of the transistor. As the signals travel different paths, this is not necessarily the case. A misalignment in time will lead to the drain voltage being shifted in time, being lower than necessary for one part of the signal and higher than the other, see Fig. 3.1.3. The red line shows how a misaligned supply voltage will not provide the supply voltage necessary to keep the drain voltage from interacting with the knee region. This can be considered an electric memory effect and manifests itself in the linearity of the signal, increasing the asymmetry between higher and lower ACPR [273], [274] and degrading EVM [275]. The accuracy requirements of the time alignment increases with the modulation bandwidth; for a modulation bandwidth of 1.25 MHz, an accuracy below 10 ns is acceptable [274] while the requirements increases to 2 ns for a signal with a modulation bandwidth of 20 MHz [273]. The accuracy requirements also depend on the architecture; the window of acceptable time misalignment is smaller in EER PAs than it is in ET PAs [276]. Another factor is the shaping function; the further the PA operates from the knee region, the more tolerant it is to time misalignment [277].
Developments in Supply Modulation - Supply Modulation at System Level

The ACPR can be used to optimise the delay between the two paths, minimising it to obtain the optimum time delay [254], the optimum delay can also be established by optimising for minimum EVM [275]. Implementing this delay is trivial in a bench top PXI setup where signals can be generated arbitrarily and well synchronised. In analogue systems, a realisation using cables as delay line is feasible, in digital systems this depends on how baseband and RF signals are generated. In the case of some of the baseband bandwidth reduction techniques mentioned in section 3.1.3, the filtering or slew rate reduction requires significant delays realising which may not be feasible in an analogue approach. Independent of the approach chosen, keeping the RF signal and the modulated supply voltage aligned is crucial to achieving high linearity.

3.1.5 Supply Modulator Ripple

Many modulator topologies utilise auxiliary frequencies in the conversion from the fixed DC supply to the modulated supply voltage, such as the switching frequency in buck-converters or the RF frequency in RF type converters. While most of this auxiliary voltage component is filtered out, filtering results in losses, these increase with the order of the filter [278] making lower order filters desirable. As a result, there is typically a residual voltage ripple superimposed on the modulated supply voltage, as represented in Fig. 3.1.4, which means that the PA will be subjected to a certain degree of supply voltage ripple in many ET systems. Depending on the level of ripple, this can be problematic because it introduces distortion due to knee region interaction. As discussed in section 2.4, a PA will be most efficient if the minimum drain voltage is as small as possible, i.e. the knee region in a real device. For modulated signals, the PA operates in output power back-off (OBO), i.e. at a reduced output power most of the time. In a supply modulated PA, the supply voltage will be reduced for reduced power levels, moving the minimum drain voltage towards the knee region as discussed in section 2.4.7. This means that in supply modulated PAs, the minimum drain voltage is close to the knee region at all
points in time. In the presence of ripple, this can prove problematic. Any ripple will be added to the modulated supply voltage which means that the RF load-line is moved toward the knee region for half of the time, increasing distortion, and away from it for the other half, reducing efficiency, see Fig. 3.1.5. This is reflected in an increase in EVM and ACPR [274].

Additionally, any frequency component present in the supply voltage will be subject to a mixing process as the transistor is a non-linear device [279]. In a memoryless system, the mixing can be described using a series analysis, see (3.6) [279].

\[
V_{out}(V_{in}, V_{ripple}) = \sum_{i,j=0}^{\infty} a_{i,j} V_{in}^i V_{ripple}^j
\]  

(3.6)

where \(a_{i,j}\) are the gain terms as a function of the \(i\)th order of the input voltage and the \(j\)th order of the voltage ripple, \(V_{out}\) is the output voltage, \(V_{in}\) and \(V_{ripple}\) are the input voltage and the output voltage ripple respectively. The first-order ripple-induced side-band will therefore be

\[
V_{out}(t) = \frac{1}{2} a_{11} v_i v_r \cos((\omega_0 \pm \omega_r)t)
\]  

(3.7)

with \(V_{in} = v_i \cos(\omega_0 t)\) being the input voltage and \(V_{ripple} = v_r \cos(\omega_r t)\) being the ripple. The equation shows that the whole signal will be mixed with the ripple voltage, leading to additional side-bands separated from the carrier by the ripple frequency. In wideband PAs, these side-bands might be well within the RF bandwidth, and thus a problem for reliable communication as they interfere with other communication systems.

### 3.1.6 Baseband Impedance

The baseband impedance \(Z_{BB}\) is the impedance that the modulator presents to the power amplifier over the modulation bandwidth, see Fig. 3.1.6. Any current flowing through this impedance, for example as a result of the amplification of a modulated signal, will also generate a voltage drop that will be superimposed on the modulated supply voltage and thus change the voltage at the drain...
of the transistor. As a result, the new, distorted, modulated supply voltage deviates further from
that required to keep the RF load-line from interacting with the knee region [247], [280]. Just as
the ripple is up-converted, so too are the additional voltage components [281]. Unlike the ripple
however, the resulting mixing products are either close to or within the signal bandwidth. If the
mixing products are within the baseband bandwidth, they lead to AM/AM and AM/PM distortion.
Out-of-band, the mixing products add to the side-bands, and, depending on the phase of the of these
distortion products, they can either increase or reduce the magnitude of the side-bands. Therefore, the
baseband impedance of the converter leads to an electrical type memory effect which can be observed in
modulated measurements as spread of the AM/AM and AM/PM measurements and as an asymmetry
in the neighbouring channels [94], [282] . The baseband impedance has therefore a significant impact
on the PA linearity [247] as can be demonstrated using baseband load-pull [15]. In classical PAs,
this issue can be countered by designing a bias network which presents low impedances over very
wide bandwidths using a range of capacitors [247], [283]. This is not an option in supply modulated
PAs however, as the capacitors would provide a low impedance at the modulation frequencies. As a
result, the baseband impedance presented to the PA is equal to the output impedance of the supply
modulator. The output impedance of a modulator depends on its topology and will be treated in the
respective subsection on pages 90 pp. The modulated supply voltage generated by the modulator can
also be designed to present the PA with negative baseband impedances, allowing linearisation of the
PA [284]. This is discussed more thoroughly in section 3.4.

Another example where the baseband impedance can not be controlled using a bias network is
during load-pull measurements. Here, the transistor is measured on its own and, depending on the
setup, it might be impossible to achieve the desired baseband impedance as the measurement system
might require a certain distance between transistor and the first possibility to inject DC and provide
the baseband impedance. In this case, a DC coupled amplifier can be used to provide the desired
baseband impedance to the device [285]. This setup also offers the possibility to present arbitrary
Developments in Supply Modulation - Supply Modulation at System Level

baseband impedances to the device, e.g. mimicking a specific baseband impedance termination, offering
the possibility to analyse different bias networks and their impact on the performance.

In addition to the baseband impedance at the output of the transistor, the baseband impedance
at the input of the transistor plays a role in the linearity of the PA [286], [287]; optimising it can
therefore significantly improve the linearity and linearisability. Simulations or measurements of the
baseband impedance of modulators are very rarely reported, possibly due to the fact that the only
reported baseband impedance measurement technique requires a specially crafted Schroeder 2K-tone
complex modulation [288].

While the effects of a non-zero baseband impedance are usually undesirable, the baseband impedance
can also be exploited. By designing the bias network carefully, the voltage generated by the baseband
impedance can be used to reduce the voltage for low power levels and increase it for high power levels,
effectively supply modulating itself [289]. This increases the efficiency significantly, albeit at the cost
of linearity.

### 3.1.7 PA Impedance

The other side of the PA-modulator interface is the PA impedance $Z_{\text{PA}}$, i.e. the impedance the PA
presents to the modulator, see Fig. 3.1.7. With

$$ P_{\text{DC}} = V_{\text{DC}} \cdot I_{\text{DC}} $$

(3.8)

$$ P_{\text{RF}} = \eta \cdot P_{\text{DC}} $$

(3.9)

and

$$ Z_{\text{DC}} = \frac{V_{\text{DC}}^2}{P_{\text{DC}}} $$

(3.10)

the static PA impedance can be calculated as

$$ Z_{\text{DC}} = \eta \cdot \frac{V_{\text{DC}}^2}{P_{\text{RF}}} $$

(3.11)

The shaping function maps the supply voltage to the input power, it has therefore a direct effect
on the PA impedance. The efficiency $\eta$ depends on supply voltage and input power, $P_{\text{RF}}$ on input
power and supply voltage. This demonstrates the impact of the supply voltage on the PA impedance.

Fig. 3.1.7 shows the PA impedance that results from the different shaping functions presented in (3.2)-
Developments in Supply Modulation - Supply Modulation at System Level

Figure 3.1.7: Simulated real part of the PA impedance presented by the PA for different shaping functions (3.4). This is consistent with the PA resistance reported in [260]. As long as the ratio of voltage and current is constant, so is the impedance [290], once that ratio changes, e.g. as the minimum supply voltage is reached and only the current changes, the impedance increases. Depending on the quiescent current, the current will become constant for small input powers, significantly reducing the efficiency at low power levels.

So far, the impedances were assumed to be DC impedances depending on RF power, supply voltage and efficiency. In a real supply modulated PA, the PA presents load impedances to the supply modulator not only at DC, but over the whole baseband bandwidth, see section 3.1.3. These power and frequency dependent dynamic impedances can be characterised by measuring baseband voltages and currents between PA and modulator [288], [291]. The dynamic behaviour of the PA or whether the impedance depends on the modulation scheme is currently unexplored, however. While it has been demonstrated that the impedance of the PA has an impact in the modulator filter design, this has only been shown with a static impedance so far [291], not taking into account the impedance changing with RF power level and applied voltages.

3.1.8 Supply Modulator Linearity

The shaping function maps the input to the output voltage, this output voltage then needs to be supplied to the PA by the supply modulator. As any transistor-based circuit is inherently non-linear, any type of supply modulator is going to introduce non-linearities and thus change the shape of the output voltage. Additionally, the modulator’s amplitude and phase response will change with frequency. By taking this behaviour into account, the linearity of the ET PA can be improved considerably [276], [292]. As the ET PA and the modulator interact, the linearity of the modulator can be characterised using static loads to be able to isolate the different error sources. It is important to know the PA impedance however as the error signal depends significantly on the load impedance [293].
3.1.9 Interaction between ET PA and Modulator

Despite the interaction between the two main elements in an ET system being very interesting, there is little research in the area. It has, however, been reported that the internal resistance of the supply modulator degrades the PAE significantly if the PA is operated in P-mode or C-mode [249].

3.2 Supply Modulators

The supply modulator has to turn the DC voltage it is supplied with to the voltage determined by the shaping function, see Fig. 3.2.1. In principle, this can be achieved by utilising the techniques used by DC/DC converters; DC/DC converters are available with efficiencies of more than 97% [294]. Unlike those DC/DC converters however, the output voltage required for supply modulated PAs has to change with the power level of the PA; not a trivial undertaking with modulation frequencies exceeding 100 MHz and baseband frequencies exceeding even that as discussed in section 3.1.3. The research into supply modulators has intensified in the last few years, working to accommodate the increase in required modulation frequencies. Not surprisingly, a lot of the research is following the methods typically employed in commercial DC/DC converters and adapting them for the high bandwidths [294], such as linear amplifiers (LA) and switch mode modulators (SMM). Others approach the problem from the RF side using the principle of RF generation and rectification [295], [296]. Another approach to generate the modulated supply voltage is following the principle of digital to analogue converters (DACs), synthesising modulated waveforms using a number of discrete voltages [297]. As all of these topologies come with advantages and disadvantages, there is a considerable amount of research into combining two or more of those techniques to compensate for their individual weaknesses at the cost of higher complexity.
3.2.1 Linear Amplifier

One of the most straightforward ways to generate a modulated supply voltage is generating the desired signal with a reduced amplitude and amplifying it with a linear amplifier (LA), e.g. an operational amplifier [174]. The efficiency of linear amplifiers is inherently limited by the same mechanism that limits the efficiency of PAs, however, as discussed in section 2.4, making this approach unsuitable to achieve high efficiency. Due to the ability of linear amplifiers to work from DC to frequencies in the double digit MHz region [298], this type of modulator can be useful to characterise ET PAs and the interaction between PA and modulator. By limiting the linear tracking to the high power levels that happen with a low probability, especially for high PAPR signals, the relatively low efficiency of a linear amplifier can be mitigated. This has been demonstrated using a charge pump that is used to boost the supply voltage [299]. Another approach is using a diplexer to add a time-variant AC signal to a constant DC signal in an approach called auxiliary envelope tracking (AET) [300]. There have been attempts to increase the efficiency of a linear amplifier by bringing the supply voltage of the linear amplifier closer to the desired output voltage, achieving efficiencies up to 50% - 60% [265], [301]. Linear amplifiers have the advantage of not introducing additional frequency components as long as they are operated outside compression and thus do not introduce supply modulator ripple. Additionally, they are available with very low baseband impedances [302].

3.2.2 Multi-level Modulator

Another way to generate the supply voltage uses a DAC, as mentioned before, piecing together the required supply voltage from a discrete set of voltages. In the context of supply modulated PAs, it is usually referred to as multi-level converter (MLC) or multi-level modulator (MLM). As with any DAC, the number of discrete voltages limits the shape the supply voltage can attain. The most simple version of the multi-level modulator, switching between two discrete voltages depending on the output power, dates back to the 1980s and is also known as class G PA [303]. As the drain voltage must not fall below the knee voltage, such a two-level modulator needs to switch between the two levels based on the drain voltage and thus output power. The thresholds at which to switch is determined by the lower supply voltage level [304]. The optimum ratio of high and low supply voltage level depends on the PA and the type and modulation frequency of the signal [304]. This ratio also determines how often the modulator switches between levels and how long the modulator stays at the voltage levels. With a 40 MHz OFDM signal and the two levels at 40 V and 25 V for example, the reported PA operates at the higher voltage for 13.5% of the time [304]. In the case of a class G PA, the switching speed has to be only slightly higher than the modulation frequency, a working modulator with a 50 MHz switching frequency has been shown to be able to track a 40 MHz OFDM signal.
Developments in Supply Modulation - Supply Modulators

Figure 3.2.2: (a) Schematic and (b) waveforms of a multi-level supply modulator (after [305])

[306], while with switching frequencies of 300 MHz, modulation bandwidths of 75 MHz have been demonstrated [307]. By increasing the number of supply voltage levels, the efficiency can be further increased while keeping the switching frequency constant [308], the efficiency can also be improved by increasing the switching frequency [308]. Another way to utilise this combination of supply voltage levels and switching frequencies is reducing the required switching frequency by increasing the number of switches. Three [309] and four [305] level modulators have been demonstrated, see also Fig. 3.2.2. If stricter switching timings can be guaranteed, the series diodes preventing direct connection of the different supply voltages [306], [309] can be removed as the switches fulfil the same function [305]. It is not only the number of switches that determines the efficiency but also how well they operate. For increasing switching frequencies, the rise and fall times start to be significant. This slew rate of the switch is shown to have a direct impact on the efficiency and linearity of the supply modulated PA [298].

Instead of switching between a number of fixed voltages, the modulated supply voltage can be generated by taking $N$ discrete fixed voltages and using switches to connect between 1 and $N$ of them in series resulting in $2^N$ possible supply voltages [249], [297], see also Fig. 3.2.3. This allows a significant improvement in terms of accuracy, keeping the modulated supply voltage closer to the ideal and the PA more efficient. However, an increased number of switches that need to be controlled complicates signal generation. Additionally, the switches are used very unevenly as the least significant bit, i.e. the switch switching the smallest fixed voltage, has to switch more often see Fig. 3.2.3. Depending on the realisation, the switching frequency has been reported to be considerably higher than the modulation frequency, a factor of 7 in the case of the example in [297] or lower by a factor of 3 - 10 for the different switches [249]. As switching is the main loss mechanism [297], this increase in switching processes has a direct impact on the efficiency, especially as the switching frequency increases, this shows in the efficiency drop from 92% for a 4 MHz WCDMA signal to 78% for a 20 MHz WiFi signal. Using a step-wave switched capacitor converter, the number of switching processes per switch can be reduced
at the cost of a large number of switches which also means a large number of digital drive signals [310].

As briefly discussed in section 3.3, the characteristics of the transistor change with the supply voltage, an effect which will be treated in more detail in section 3.3.1. As a result of this supply voltage dependent characteristics, larger steps between the voltage levels lead to more non-linearities that are introduced by the modulator. One of these non-linearities is in-band, the change in AM/AM and AM/PM due to the change in supply voltage [311]. Another non-linearity is out of band and introduced by a step in the supply voltage which is up-converted with the RF signal at the drain side of the transistor. This up-conversion can lead to out-of-band components that are not measured in typical ACPR measurements and can thus stay undetected. One way of dealing with these issues is using a low-pass filter between modulator and PA [309] to smooth the modulated supply voltage at the cost of additional components and losses. Snubbers, i.e. RC series networks in parallel to switches, can also be used to reduce the impact of the steep switching edges, they are particularly useful to suppress overshoot and ringing [305]. Instead of using filters, the multi-level modulator can also be used to supply a modulated supply voltage to the linear amplifier which is then able to generate the modulated supply voltage for the ET PA efficiently and over a wide bandwidth [310], [312] at the cost of efficiency. A different combination of multi-level modulator and linear amplifier is through the connection at their outputs [313]. Here, the linear amplifier smooths the output voltage by providing the high frequency components the multi-level converter is not able to provide, increasing the accuracy with which the envelope is tracked.

As each switch needs a dedicated digital input, generating the input signals for these modulators is non-trivial as they need to be tightly synchronised. FPGAs are able to provide a solution to this issue [305], offering a high number of channels and a good resolution of 125 ps [305]. As neither output voltage nor output current of the FPGA are sufficient to drive the switches directly, each switch needs a driver, increasing complexity and reducing efficiency. A less complicated way of controlling the
Developments in Supply Modulation - Supply Modulators

Figure 3.2.4: (a) Schematic and (b) waveforms of a switch mode modulator (after [294])

Switch-mode power supplies (SMPS) are known for their efficiencies exceeding 98% [315] so they were an obvious starting point for high efficiency supply modulators. Adapting them to be used as modulators changed the requirements. Most SMPS are designed to output a single voltage but in the case of the switch-mode modulator (SMM), the output voltage needs to change very fast. On the other side, base station communication equipment has well-defined and stable DC voltages which eliminates the need for a boost stage and power conditioning, simplifying the design. The dominant design is that of a buck converter. It works by using the switches $Q_{HS}$ and $Q_{LS}$ to switch the output between the supply voltage of the modulator $V_{DC}$, and a low voltage, often 0 V, see Fig. 3.2.4. This results in a PWM-type waveform that is low-pass filtered to yield the desired modulated supply voltage $V_{DD}$ [316]. The reason that this very simple technique has resulted in countless publications and patents over many decades is that despite its simplicity, it is non-trivial to implement efficiently, especially as
Developments in Supply Modulation - Supply Modulators

the switching frequency increases. Achieving a high efficiency means managing the losses in the buck converter, these losses can be divided into four major categories [317], [318]:

- Conduction losses, the power that is dissipated in the switches when they are fully turned ON
- Switching losses, the losses resulting from switching between ON and OFF states
- Driving losses, the power lost in drivers, including resistors in the driving circuitry
- Passive losses, the power dissipated in the passive elements of the SMM, e.g. inductors and capacitors, but also in other passives such as the PCB itself.

To manage the losses that result from the transistors characteristics, circuit design, and the excitation of the switches, the underlying issues have to be addressed.

Transistors as Switches

Up to this part in this thesis, transistors have been mostly treated with respect to their behaviour and suitability to operate within power amplifiers. In the past, the distinction between transistors for PAs and transistors working as switches has been a valid assumption as there was little overlap. This has changed however with modern GaN HEMT technology being the first choice for switches in high frequency modulators: GaN HEMTs with a gate length of 150 nm, usually used in RF amplifiers where they offer around 15 dB of gain at 10 GHz [263], are now used as switches [317]. This makes sense since switches for SMMs share many requirements with transistors for PAs: Low capacitances, low on-resistance, little trapping. For this reason a lot of research has gone into the characterisation of the voltage dependence of the intrinsic capacitances [319], [320] and the analysis of their physical origin [86]. While the on-resistance can be reduced almost arbitrarily by increasing the device gate width, a larger device will have more capacitances, calling for a trade-off between on-resistance and capacitances, one of the main design decisions in SMM design. Trapping is a characteristic that is as undesirable in switching as it is in RF PAs. While it is called knee-walkout in the RF domain, see section 2.2, it is usually called referred to as $R_{ON}$-Modulation [321] in the context of switches. While it reduces the achievable efficiency by enlarging the knee region in RF PAs, it increases the on-resistance from the static values measured in static DCIV setups depending on applied voltages [321]. This increase in on-resistance directly increases losses, limiting the achievable efficiency. One thing that is rarely investigated in transistors for PAs is the operation in the third quadrant, i.e. for negative drain voltages. If used in PAs, there is little interest in it resulting in some major transistor manufacturers completely neglecting the modelling in this region, see Fig. 3.2.5. In designing SMMs, however, these operating conditions can be more relevant. Many implementations of SMMs use reverse diodes across the transistors [317], [322]; depending on the gate bias level, the transistors themselves can operate
Circuit design

The basic circuit of a SMM is fairly simple, as shown in Fig. 3.2.4, consisting of the switches, the drivers and the output filter. Driving the low side transistor is easy since its source voltage is fixed so the only challenge is charging the gate quickly enough. The high side however is significantly more challenging as the source is tied to the switching node, see Fig. 3.2.4. This means that applying the gate source voltage used to drive the high side transistor can not be done using a direct connection but requires additional circuitry. One approach to overcome this limitation uses isolated DC/DC converters and digital isolators [326], the band limitation of which limits the slew rate and thus switching frequency. Other ways of driving the high side use configurations such as active pull-up and bootstrap of which the active pull-up shows the best efficiency if modified, the resistors used in them will however form an RC filter with the capacitances of the transistors, limiting the achievable rise time and thus achievable switching frequency due to the introduced losses. A third way of getting around the issue of driving the high side switch is using a reverse buck-converter that switches between the PA’s RF ground and DC ground while the PA’s DC input voltage is fixed [327], this requires a PA which is floating at
Developments in Supply Modulation - Supply Modulators

DC, however. Additionally, the reported reverse buck-converter is not able to closely track the supply voltage [327]. Depending on the architecture, the driver losses can be responsible for between a quarter to about 60% of the overall losses [317], [318], making it an important area of research [270], [322], [328].

Another important part of the circuit of a SMM is the output filter, consisting of one or more LC stages. The inductor in the filter of the modulator serves two purposes. The first one is critical to achieve high efficiency: it needs to fulfil the ZVS requirements, i.e. having a resonance at the switching frequency between filter inductor and switch-node capacitance [329], [330]. An example of this is the efficiency of a buck-converter switching at 100 MHz which drops slightly when reducing the inductance from 47 nH to 39 nH, when further reducing it to 27 nH, the efficiency drops considerably [317]. The second purpose is filtering the switching frequency, making sure that the magnitude is low enough to fulfil the system requirements in terms of ripple voltage and thus not lead to too much distortion at the PA as discussed in section 3.1.5. The filters are therefore expected to have a good stop-band attenuation, little pass-band attenuation and little phase shift in the pass-band to not distort the supply voltage too much. Additionally, losses in the filter are important, with the losses mainly depending on the losses in the inductor. These losses have been shown to be causing up to 10% of the overall losses in an ET PA [278]. Additionally, the filter configuration has a significant impact on the ability to accurately reproduce a desired signal, especially in terms of voltage overshoots [331]. To reduce the resonance of the filter, a dampening circuit may be used. While introducing losses in form of a resistor, these losses should not be significant in the frequency range of operation if designed properly [265].

Generating the switching signals is a challenging task, especially as the switching frequency increases. The two main ways are using digital electronics such as field-programmable gate arrays (FPGAs) [317] or high-speed arbitrary waveform generators (AWGs) [318] and analogue using comparators [332]. Both approaches have limits, time resolution in the digital and slew rate limits in the

Figure 3.2.6: Drivers and output filter in a buck-converter type modulator
analogue case. Both limit the achievable duty cycle as the transistors have to be fully switched for the SMM to operate efficiently. As the switching frequency goes up, the rise time and fall time takes up more and more time of the cycle. Additionally, moving the duty cycle away from 50% results in shorter on or off times, which means that rise time and fall time become increasingly relevant for low and high output voltages. As a result, the efficiency drops with increasing switching frequency, due to the switching taking up more than a negligible portion of the time [333], additionally limiting the output voltage range.

Buck-converters usually operate with a fixed switching frequency, resulting in a distinct peak in the frequency domain, leading to distortion and mixing as discussed in section 3.1.5. As the switching frequency is not phase coherent with the RF frequency, resulting distortion is hard to predict and thus pre-distort. If the switching frequency can be generated to be synchronous with the RF signal, the ripple-induced side-bands can be removed using digital pre-distortion [334]. This requires a synchronised clock or a common signal source for the switching and the RF signal. It also requires the DPD bandwidth to cover both ripple induced side-bands, i.e. at least two times the switching frequency plus the modulation frequency. Another way of removing the effects of ripple is by characterising it between supply modulator and PA and using this information to pre-distort the PA [335]. A different approach to reducing the ripple is by using not one switching frequency but a whole spectrum of switching frequencies. By spreading the switching over a spectrum, the power in the individual frequency components is reduced, e.g. spreading it perfectly over one hundred frequency points would reduce the power in the individual frequency components by 20 dB. While real systems will not achieve a perfect spreading, a 19 dB reduction when spreading the spectrum over 290 frequency points has been demonstrated [336].

Multi-Phase Converters

Instead of increasing the switching frequency, the achievable bandwidth can be increased by using multiple buck converters that are working not synchronously but interleaved [301] in multi-phase switch-mode modulators (MP SMM) moving the switching frequency close to the modulation frequency. An 8 phase converter with a switching frequency of 4 MHz allows tracking of the supply voltage with a modulation frequency of 5 MHz [301], for a four phase converter the ratio of switching frequency to modulation frequency needs to larger, with $f_{sw} = 25$ MHz and $f_{mod} = 20$ MHz [330]. With only two phase-shifted converters, the switching frequency $f_{sw} = 50$ MHz needs to be significantly larger than the modulation frequency $f_{mod} = 20$ MHz [329], however this ratio of 2.5 is still significantly less than it would be without multi-phase architecture. In a MP SMM architecture, the multiple buck converters have individual inductors [301] which sets new constraints on the inductors and thus the filter design [301]. The individual inductor is also required to provide the individual
buck-converter with the environment it needs for ZVS [330]. The constraints can be reduced by using a fourth order filter, allowing one inductor per buck-converter to ensure ZVS operation and a second filter stage shared by all buck-converters [330]. Instead of buck-converters, a multiphase converter can be realised using phase shifted switched capacitors, eliminating circulating currents [337]. The downside of multiphase converters is that they require accurate phase alignments between the branches and two drive signals per phase, resulting in a complex drive signal generation. In multi-phase converters where the number of phases $N$ is even, the triangular PWM waveforms are out of phase by $\frac{2\pi}{N}$ which means that there are always two phases $180^\circ$ out-of-phase [330]. This effectively cancels the ripple in multi-phase converters, ideally reducing it to zero which means that the output inductors only have to ensure ZVS and are not required to filter switching ripple.

**Multi-level Buck Converters**

Another way of increasing the efficiency of the buck converter of enabling higher switching frequencies is by combining multi-level modulator and buck converter in one design. As discussed in section 3.2.2, the multi-level modulator switches between different supply voltages, making it unable to follow the supply voltage closely with a realistic number of supply voltages. In a multi-level buck converter, the voltage is switched between these discrete voltages, allowing a significantly higher voltage resolution [309], [338]. Amongst the advantages of this topology are lower ripple voltage due to the lower currents that are being switched, more efficiency since the voltages that the transistor switches between are lower and better thermal behaviour as the losses are distributed across more devices. However, the signal generation is significantly more complicated than in buck converters or multi-level modulators.

**Combining Linear Amplifier and Switching Mode Modulator**

As SMMs struggle to cope with high modulation frequencies efficiently despite all of these developments, they are often combined with linear amplifiers. In this approach, also called hybrid supply modulator, the SMM provides the power at low frequencies at a high efficiency while leaving the high frequency components to a less efficient linear amplifier [271], [339], see Fig. 3.2.7. The figure shows that most of the power is contained in the lower part of the spectrum with 80% being with a few kHz of DC, around 15% between that and the modulation frequency and less than 5% being above that [271]. That means that if the efficient SMM covers the 80% around DC, the relatively low efficiency of the linear PA does not have a large impact on the overall efficiency of the hybrid supply modulator. While its losses are still dominating the overall losses of the modulator [340], the system efficiency is considerably higher than possible with a linear amplifier. To optimise the combined modulator, their losses have to be analysed in conjunction. With the SMM being an established design, especially if no high modulation frequencies are required, the design in hybrid supply modulator focuses on the linear
amplifier [341] and the circuit that drives and connects the two circuits. The driving signals for the SMM can be generated by measuring the output current of the linear amplifier [339], [342], although a separate synthesis of SMM and linear amplifier driving signals has been shown as well [343]. If the linear amplifier is using feedback, it is able to correct for the ripple introduced by the buck-converter, resulting in very low ripple voltages, e.g. 8 mV$_{P-P}$ at 0.6 V output voltage [344].

Ways of optimising the overall efficiency combine many of the different modulator techniques discussed in this section. One example of such a combination works by providing the linear modulator with a modulated supply voltage using a SMM, reducing the bandwidth of that modulator allowing efficient operation [344]–[346] at the cost of all the control overhead that this entails. Another combination uses a multi-level modulator [343], [347] instead of a SMM to increase the efficiency. A third approach is splitting the bandwidth not into a region for predominately SMM and one for predominately the linear amplifier but three regions, two of which are covered predominately by a SMM with different switching speeds [348], [349]. Another completely different way of combining switching mode modulators and linear amplifiers is by not connecting the SMM directly to the load at all but instead just tracking the supply voltage of the linear amplifier [301], which allows moving the filters before the linear amplifier, resulting in a low baseband impedance and a tracking accuracy that is higher than that of a buck-converter. Its efficiency is considerably lower than that of other modulator types though.

### 3.2.4 RF Supply Modulators

Another approach is using an RF amplifier and rectifier to generate varying DC signals. They operate by up-converting the envelope signal to RF frequencies, amplifying it to the required power level and rectifying the resulting RF signal, resulting in a time varying DC signal, see Fig. 3.2.8. This converter
Developments in Supply Modulation - Supply Modulators

Figure 3.2.8: Schematic of a RF DCDC Modulator

topology comes from the world of power electronics and is based on the Čuk converter [350]. It has been mostly used to realise compact DC/DC converters as at high frequencies, components sizes shrink [351], [352] and, for microwave frequencies, the converter can even be completely integrated as demonstrated at 4.6 GHz using GaN-on-SiC technology [296]. The technology can also be used to design ET modulators [325], operating efficiently over a voltage range of more than 6 dB, showing promising results modulating a 1 GHz RF PA [295], [353]. As both PA and rectifier are limited in their efficiency, the overall efficiency of this type of modulator is lower than that of multi-level converters or SMMs with the main losses occurring in the PA [354]. As they operate at RF frequencies, designing them to cover a wide bandwidth is simpler than it is in SMMs. Regulating the modulated supply voltage accurately is critical in ET applications. The two methods that have been proposed in literature are the use of an out-phasing PA to drive the rectifier [353] and using a frequency modulation scheme [295], they both provide high efficiency over a wide output voltage range.

With RF supply modulators being quite a niche, there is not a lot of literature on their optimisation. As they can be split up into PA and rectifier, these two components can be optimised individually. In general, PA and rectifier operate in a similar way in terms as far as matching and harmonic terminations are concerned [355], [356] which means that an amplifier with RF power injected into its output will provide a DC signal at the drain bias port [355]–[357]. This can be used in a communication system to switch the PA between being an amplifier and being a rectifier, corresponding to the sending and receiving element [357], [358]. In the case of a transistor as the rectifying element, it will need the an additional gate bias, the level of which significantly changes the efficiency over power behaviour [356], [359]. By selecting the transistor to only require a small positive gate voltage, a self biased rectifier can be realised [360]. In addition to the gate bias voltage, transistor based rectifiers require a gate signal. Figure 3.2.9 shows three ways to generate this gate signal. It can be directly based on the signal as the PA gate signal, supplied externally independent of the other signals in the circuit or generated from the drain signal of the rectifier by means of a feedback network of some sorts. In the last case, the gate signal of the rectifier is always in sync with the drain signal, therefore a rectifier using generating the gate signal using feedback is called self-synchronous. The feedback network can be realised using a physical components [360] or using the feedback capacitance $C_{GD}$ and a terminating
Developments in Supply Modulation - Supply Modulators

Figure 3.2.9: Schematic of RF DCDC converters with different ways to generate the rectifier’s gate signal, after [351]

As RF DCDC converters can easily operate in the GHz region, the operating frequency is significantly above the bandwidth of communication signals which facilitates the filtering of the operating frequency. This means that for a comparable order of the output filter, the RF DCDC converter produces less ripple. Additionally, with the ripple being at the operating frequency in the GHz range, any conversion products are unlikely to result in interference due to their frequency.

3.2.5 Comparison of Supply Modulators

In Tab. 3.1, the different modulator architectures are compared in terms of different relevant metrics. This is a subjective comparison however, as even within one type of architecture, these metrics can differ significantly. They are evaluated on the perceived limitation of the technology, not only the results presented in the literature. One observation is that efficiency and modulation bandwidth are linked, the higher one the lower the other. Comparing supply modulators in the literature is significantly harder though. The efficiencies are hard to compare as in a supply modulator for ET, the modulator would not be used at the maximum power most of the time, this is the power level of which most efficiencies are reported as it is the highest. Even if the average efficiency for a modulated signal is specified, the modulated signals are shaped by different shaping functions and thus not directly comparable. The modulation bandwidth that is reported in the paper is often too high for the modulator to generate accurately so the generated voltage will differ from the intended supply voltage,
Development in Supply Modulation - Supply Modulated Power Amplifiers

Table 3.1: Comparison of supply modulator topologies

<table>
<thead>
<tr>
<th></th>
<th>LA</th>
<th>MLM</th>
<th>SMM</th>
<th>SMM+LA</th>
<th>MP</th>
<th>SMM+</th>
<th>RFM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency</td>
<td>-</td>
<td>++</td>
<td>++</td>
<td>o</td>
<td>++</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Modulation Bandwidth</td>
<td>++</td>
<td>o</td>
<td>-</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>Input signal generation</td>
<td>++</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>+</td>
<td>o</td>
<td>+</td>
<td>-</td>
<td>o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ripple</td>
<td>++</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Integratability</td>
<td>++</td>
<td>+</td>
<td>o</td>
<td>-</td>
<td>o</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Tracking Accuracy</td>
<td>++</td>
<td>o</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseband Impedance</td>
<td>++</td>
<td>?</td>
<td>o</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
</tbody>
</table>

Sometimes significantly [270], [327], [362], leading to inflated modulation bandwidth on the one side and to electric memory effects [270] on the other. Showing a Bode plot of the supply modulator and the normalised error would allow better comparison. Input signal generation is usually just mentioned very briefly, so is ripple and the tracking accuracy. The baseband impedance is usually not mentioned at all, despite its impact on the PA. As discussed in section 2.5, the inductor is the main limitation when it comes to integrating circuits. This is particularly relevant in supply modulator topologies that require an output filter, in those cases the inductor has to be realised off-chip, preventing full integration. As the frequency that needs to be filtered increases, the required inductor value decreases which increases integratability.

3.3 Supply Modulated Power Amplifiers

In the power amplifier, or more specifically its transistor, the RF signal and the baseband signal are combined, as shown in Fig. 3.0.1. The PA is the last element in the chain before the antenna and, as a result, has the highest output power and has to fulfil the stringent linearity requirements. The transistor, as the centre-piece of the PA will be discussed first in section 3.3.1, both in terms of requirements imposed upon it by the applied varying supply voltage and in terms of characterisation under modulated supply voltage. Section 3.3.2 presents a variety of supply modulated PAs before going into detail on output matching network design, integration, tracking of the gate voltage and multi-stage PAs. Just as multiple modulator topologies can be combined to achieve higher efficiencies, different high efficiency amplifier techniques can be combined, this will be treated in section 3.3.3.
3.3.1 Transistors for Supply Modulated PAs

As discussed in section 2.2.3, there are many requirements a transistor needs to fulfil to be suitable for use in a power amplifier. Under supply modulation, the operation conditions the transistor is subject to become more complex and include a modulated supply voltage and, possibly, an additional modulated gate voltage. The transistor needs to be able to operate efficiently and linearly under these conditions. This section will discuss how the transistor’s behaviour changes with supply voltage and in which ways a transistor needs to be characterised be able to judge whether a transistor is suitable for supply modulation.

As the supply voltage changes, so does the behaviour of the transistor. To be useful in modern communication systems with high bandwidth signals with high PAPR, the supply voltage in supply modulated PAs will need to be highly dynamic and be able to attain values in a significant supply voltage range. This range is determined by the PAPR of the signal, the transistor and the supply modulator. Ideally, the supply voltage would go down to zero, in reality, the transistor’s knee region limits the supply voltage range in which efficient operation is possible. Additionally, the supply modulator’s efficient output voltage range is limited, depending on the architecture. To keep the PA efficient in the range where it is most relevant, i.e. the dissipated power is highest, the supply voltage needs to be tracked over the range of the PAPR of the signal. This means that a target supply voltage range for a LTE signal with a PAPR of 9 dB would require the same dynamic range, i.e. 9 dB which corresponds to a voltage range of 2.8, e.g. 10 V - 28 V. If the supply voltage changes over a range that significant, there are three main parameters that change in the transistor: the non-linear capacitances, the trapping effects and the gain.

As the two capacitances \( C_{DS} \) and \( C_{GD} \) transform the intrinsic load impedance of the transistor to an extrinsic load impedance, they need to be taken into account during the matching network design. As a result, any change in their value moves the impedance that the intrinsic transistor is presented with. This means that any non-linear behaviour of these capacitances moves the load impedance, possibly impacting the PA performance in terms of efficiency, output power, gain and linearity [257]. This shift in optimum impedance will be discussed further in section 4.1.1. Fig. 3.3.1 demonstrates measurements of the non-linear capacitances of a Qorvo GaN-on-SiC HEMT. They show that \( C_{DS} \) increases almost linearly with decreasing supply voltages, from 0.28 pF at 2 V to 0.2 pF at 50 V. The feedback capacitance, \( C_{GD} \), changes significantly more, from 0.01 pF at 50 V to 0.12 pF at 2 V. The change in capacitance is additionally non-uniform, slowly increasing with decreasing supply voltage until it reaches a threshold voltage, here around 15 V, at which it increases significantly, the result of the gate-field plate. The field plate extends over the edge of the gate towards the drain, increasing the width of the depletion region towards the drain [86]. As the drain voltage decreases, this gate field
plate depletes the channel less and less, the carriers that are now in the channel form a parallel plate capacitor with the gate field plate. Once the channel is fully populated, the $C_{GD}$ starts to plateau as shown in Fig. 3.3.1. This means that, depending on the values the supply voltage can attain, the matching network needs to take the non-linearity of the capacitances into account. Additionally, load-pull measurements now need to be conducted at multiple supply voltages [363], showing how both the load resistance and the average output capacitance change significantly with the supply voltage and the RF input power [363]. This also requires models to be able to represent these non-linear capacitances [86], [364].

As discussed in section 2.4.7, the efficiency of supply modulated PAs decreases with the supply voltage as the ratio of supply voltage and knee-region decreases. In PAs using HEMT devices, trapping effects mean that the knee-region can depend on the supply voltage, increasing with supply voltage as a result of the surface traps at the interface between passivation and barrier. This constricts the channel and increase the on-resistance and decrease the achievable current [365], as discussed in section 2.2. This effect is known as knee-walkout or RF/DC dispersion [80], [122], [365], the increased losses limit the supply voltage the transistor can be efficiently operated at. Fan diagram measurements allow quantisation of this effect, they can be conducted by sweeping the load impedance at different supply voltages with enough RF input power to drive the PA into either the knee region or the current limit of the PA [122] at each load impedance, effectively mapping out the outside boundaries of the DCIV plane at RF frequencies, as shown in Fig. 3.3.2a. By taking the knee voltage values at different current levels, the knee-walkout can be evaluated as a function of the targeted current level and thus the targeted power level. Fig. 3.3.2b shows the impact of knee walkout. In a transistor without knee-walkout, the knee-voltage would be constant and the achievable ideal class B efficiency would increase with supply voltage, as discussed in section 2.4.3. For all three different current levels of around 260 mA/mm, 375 mA/mm and 500 mA/mm, the knee-voltage increases with supply voltage,

![Figure 3.3.1: Measurement results of a Qorvo GaN-on-SiC HEMT showing the drain voltage dependent output capacitances $C_{DS}$ and $C_{GD}$](image-url)

Figure 3.3.1: Measurement results of a Qorvo GaN-on-SiC HEMT showing the drain voltage dependent output capacitances $C_{DS}$ and $C_{GD}$
Figure 3.3.2: Measurement results of a GaN-on-Si HEMT (a) fan diagrams showing the knee-walkout and (b) impact of knee walkout on efficiency for different target current levels, solid grey lines indicate maximum class B efficiency considering the knee voltage as shown in Fig. 3.3.2b, resulting in the maximum achievable efficiencies staying almost constant for each current level. It also shows that by reducing the targeted current level, the achievable efficiency increases significantly, albeit at the cost of achievable output power which also increases the capacitance-per-Watt figure of merit introduced in section 2.2.3.

A third effect the of the supply voltage modulation on the transistor is the gain of the transistor. The gain of transistor varies with the supply voltage [366], a small amount of gain variation has been shown to be beneficial, offering the flexibility required for baseband injection to linearise the AM/AM characteristic [366]. If the gain drops too much, the PAE and the linearity of the PA will be affected however. The gain variation, its impact on the PA behaviour, its causes and possible mitigation strategies are treated in section 4.2 on pages 137 pp.

3.3.2 ET PAs in Practice

Over the years, a significant amount of ET PAs have been published, demonstrating working supply modulated PAs in output power ranges from hundreds of mW [94] to several hundred Watts [172], [367]. This section aims to give an overview over the applications they were designed for and technologies employed in the design. The role of the output matching network, multi-stage supply modulated PAs and gate voltage modulated PAs will be discussed in more detail following the general overview.

Supply modulation is mostly discussed as a method to increase the efficiency in mobile communications [293], [298], [368] as mobile communications such as 5G are a technology driver for PA architectures that promise high efficiencies over a wide power range. In these applications the achievable RF bandwidth of the PA is significant advantage, a 1.5 GHz - 2.5 GHz continuous mode class F PA is shown to work efficiently over a wider power range [369], the paper fails to go into detail in
many aspects such as ET operation over bandwidth, modulator and symmetry of ACPR, however. If a base station is to transmit at lower power levels, the supply voltage can be reduced to achieve this efficiently, e.g. in the case of a transmitter for GSM with its constant amplitude that is adjusted to different power levels over the day to cover different cell sizes, a feedback loop can be used to generate the supply voltage based on the output signal [370]. However, mobile communications are not the only application of supply modulation, any application with high PAPR values benefits from it. One such application is digital terrestrial TV broadcasting in the UHF range with a COFDM modulation with a PAPR of 9 dB, making supply modulation an attractive solution [367]. In this application, the substantial power levels of up to 470 W PEP can be produced by two push-pull LDMOS devices, with the modulated supply voltage being supplied by a now discontinued Emerson multiswitcher [367]. Even higher power levels, 650 W, are needed in driving linear accelerators, however, in this instance GaN HEMTs are used instead of LDMOS [172] to achieve higher efficiencies. It is not only RFPAs that benefit from supply modulation, however. By dynamically tracking the supply voltage of a class D audio amplifier, its efficiency can be increased significantly in output power back-off [371]. This is particularly relevant since most audio amplifiers are rarely operated at their maximum specified output power [371]. The integration of the PA part of an supply modulated PAs follows the same strategies as integrating conventional PAs, therefore the interesting part is not the integration of the PA itself but that of combining ET PA and modulator on one MMIC. As discussed in section 3.2.2, a cascaded multi-level modulator can be easily integrated with the PA due to its lack of inductor [249], resulting in a monolithic ET PA. The results show an efficiency below that achievable with a hybrid modulator however, due to the transistor technology that was developed for RFPAs and the lack of p-type devices [249].

**Output matching**

If the output capacitance varies significantly as discussed in section 3.3.1, the impedance providing a good trade-off between output power and efficiency moves [372]. This means that the output matching network design needs to consider not only the behaviour at the maximum supply voltage but also the efficiency trajectories at lower supply voltages. By factoring the modulation standard into this design process, the average efficiency of the PA under modulation with the target signal can be optimised [121], [372]–[375], this optimisation can be done manually by comparing the load-pull data at different supply voltage or using equations [375]. If the PA is not focussing on a specific modulation standard, it can be matched to have the maximum efficiency at a supply voltage lower than the maximum [298], [376]. If the PA is merely a test object, the efficiency characteristics might be irrelevant, in this case the PA might be matched to highest efficiency at the maximum supply voltage as it would be in traditional, non-ET, PA design [256]. In devices with a output capacitance that is significantly
non-linear over supply voltage, the second harmonic generated in it can be used to move the harmonic load impedance and thus change the PA class of operation, e.g., to keep the transistor to from reaching voltage breakdown [377].

**Muli-stage ET PAs**

In communication systems, amplifiers need to achieve certain gain requirements, they therefore comprise more than one stage [378]. In the case of ET PAs, that leads to the question of which PAs to apply the tracking to. One option is tracking exclusively the final stage PA [94], another tracking both the driver PA and the final stage PA [379]. In the case of more than two stages, there are even more cases [380]. In addition to the question of which PAs to track, there is also the questions of how to track the different PAs, all with the same supply voltage [380], [381] or, alternatively, with different, individually shaped supply voltages for the different stages [382]. Significant PAE improvements when tracking both driver PA and final stage PA have been reported [381], albeit without detailing the reason why the PAE would be significantly bigger.

**Gate Envelope Tracking PAs**

While the previous examples have only taken the modulation of the drain supply voltage into account, it is also possible to modulate the gate voltage of the transistor. While this does not have the potential to increase efficiency to the same degree as modulating the drain voltage, modulating the gate requires only very low power levels, making it significantly easier to realise. It was first demonstrated as a way to move the quiescent current between class B and class A while keeping the current waveform from reaching zero, operating in class A over a wide power range [383]. Class A PAs are not used any more in fields where efficiency is of concern, tracking the gate voltage is now all about linearity as compared to an operation at a static gate bias, a tracked gate voltage allows both AM/AM and AM/PM distortion to be reduced while slightly increasing efficiency [384]. Just like with the modulated drain voltage, the applied gate voltage follows a shaping function, which determines linearity and efficiency [384]. In the case of a multi-stage PA, the two gate voltages can be tracked separately, providing a wide range of gate voltage combinations that can be used to optimise linearity further, again with little to no cost in terms of PAE [385]. Tracking both drain and gate voltage allows further improvement in terms of linearity and efficiency although the reported efficiency values do not seem to be consistent with the plotted PAE trajectories [386].

### 3.3.3 Supply Modulating High Efficiency PA Topologies

As supply modulation increases the efficiency of individual PAs, it can be combined with any other high efficiency PA topology. While it means that the resulting structure has the disadvantages of both supply modulation and the topology it is combined with, the improved efficiency might make it
worthwhile.

**Doherty PA**

As discussed in section 2.4.7, the DPA comprises at least two transistors, which means that there are at least 4 voltages that can be supply modulated. In conventional DPAs, the transistor of the peaking amplifier can not be fully utilised as it is biased in class C, tracking the gate voltage of one or more of the peaking amplifiers will therefore increase the output power that is achievable from the DPA [199], it can also be used to improve the linearity [387]. While neither of these papers show the improved output power that should be achievable, this is demonstrated in [388] where the gate voltage of two peaking amplifier stages of a 3-stage DPA are modulated, significantly improving the efficiency and output power. Comparable to section 3.3.2, the gate voltage of the main PA can be tracked to increase the efficiency of the PA [190]. The two gate voltage tracking schemes can be combined, tracking both the main and the peaking gate voltage for increased efficiency and linearity [389], [390]. Tracking the main transistor’s drain voltage allows the output power range of high efficiency to be extended significantly [314], [391]. The benefit of ET heavily depends on how favourably the transistor reacts to the supply modulation of course, showing good results in the GaAs IC in [391] while the Wolfspeed GaN HEMT based DPA only slightly increases in efficiency [392], partially because its gain reduces significantly with supply voltage. Especially in asymmetric DPA with a peaking amplifier that is significantly larger than the main amplifier, the efficiency dips significantly as discussed in section 2.4.7. By supply modulating the peaking amplifier’s drain voltage, this dip in efficiency can be partially recovered [187]. Instead of modulating the individual supply voltages of the two PAs, they can be modulated together, this has shown to improve the efficiency over the whole output power range [393].

**LMBA**

Like the Doherty PA, the LMBA has multiple voltages that could be supply modulated. As the two balanced PAs are operating in the same operational conditions, it make sense to regard them as a single PA as far as supply modulation is concerned. While there has been no reported gate modulation in LMBA so far, some of the different options of tracking the supply voltages of balanced PAs and control PA have been evaluated [394]. The results show that applying modulated to both balanced PA and the control PA gives the best efficiency as expected, the fact that these PAs are not close together on the PCB make it hard to modulate both of them in the presented setup [222].

**Out-phasing PA**

By modulating the supply voltage of both individual PAs in an out-phasing PA in addition to the phase in multilevel LINC architectures, the dynamic range can be improved, even with a limited phase
Developments in Supply Modulation - Linearity and Linearisation

accuracy [395]. The two amplifiers can also be supplied with different supply voltage and phase in an asymmetric multilevel out-phasing [207] scheme. This results in higher efficiencies while reducing the requirements for phase accuracy but imposes new requirements on the ability of the power supplies to react to modulated signals. If the two amplifiers are supplied individually, two different modulator drive signals need to be generated, increasing the system complexity further.

3.4 Linearity and Linearisation

In an ideal ET system, changing the supply voltage has no impact on the linearity of the PA [11]. In the absence of an ideal transistor however, supply modulating a PA has multiple effects on the linearity. Depending on the gain variation and the non-linearity of the intrinsic capacitors of the transistor, the AM/AM and AM/PM characteristic are going to change, usually for the worse. The ability to use the supply voltage to linearise the PA is a redeeming feature, however.

Shaping functions

The dependence of the gain and phase on the supply voltage grants the ability to shape the gain and phase behaviour to a certain degree by adapting the shaping functions of the supply voltage. This is used in iso-gain shaping functions, which improve the PA linearity considerably [366], [386], both in terms of EVM and ACPR. By combining an iso-gain shaping function with a gate voltage tracking to realise an iso-phase tracking, the AM/PM distortion of the PA can be linearised to the point at which it passes the linearity requirements of modern communication standards [93]. Shaping functions can also be optimised by trying to compensate for the non-linearities of the transconductance and the intrinsic capacitances [347].

Baseband injection

As discussed in section 3.1.6, the baseband impedance presented to the transistor has a significant impact on the supply voltage and, with it, on the asymmetry of the power in the neighbouring channels. With full control over the supply voltage, the supply voltage can be changed to present a perfect short circuit to the transistor [396]. By presenting negative impedances to the PA, the intermodulation distortion can be reduced. While the optimum impedances for minimal $IMD_{3L}$, $IMD_{3H}$, $IMD_{5L}$ and $IMD_{5H}$ differ significantly, the overall linearity can be increased [396], [397]. This allows cancelling out the intermodulation distortion introduced in the transistor [398], [399]. The baseband injection does not have to be realised on the PA itself though. In a multi-stage PA, a driver stage can have the baseband injected and achieve improved linearity of the multi-stage PA [379]. The lower power level at which this injection happens can facilitate the implementation. This is due to the
Developments in Supply Modulation - Conclusions

fact that lower power levels also mean that inefficiencies in the modulator have less an effect compared to the a modulator that is capable of modulating the final stage.

As the baseband injection needs to be at least two times the baseband bandwidth, the efficiency of a modulator supporting baseband injection would thus be lower because it needs to accommodate higher frequencies.

**DPD**

As supply modulated PAs have two input signals, DPD can be applied not only to the RF path but also to the baseband path [400], [401]. The pre-distorted RF signal can be used a basis for a pre-distorted baseband signal that is fed into the supply modulator. Alternatively, RF and baseband could be pre-distorted completely separately. In multi-stage supply modulated PA, the DPD can be based on just the main PA or, alternatively, have multiple loops, linearising the driver PA in addition to the main PA [382]. In the case of multi-level converters, the gain and phase change significantly as the supply voltage changes between two values, this is particularly significant if the difference between the voltages levels is large [311]. As this is a deterministic problem, DPD algorithms can be optimised to take these step changes into account by having different terms for different power levels [311].

### 3.5 Conclusions

This chapter dived deep into supply modulated power amplifiers, examining at them from all angles. Supply modulation is a promising technology for high PAPR, high bandwidth application which makes it an interesting area to both industry and academia. This has lead to a wealth of publications around the topic, ranging from research on constituting parts to system level considerations. Researching the building blocks of a system is easier than working on their interfaces and interactions which is reflected in the number of publications on these topics, revealing some notable gaps. One gap is a the result of the lack of research into system level interactions, resulting in significant lack of data on the impedances of the subsystems, i.e. PA and supply modulators, especially for non-switch mode modulator. These non-switch mode modulators are also under-represented in terms of research, with buck-converters making up the majority of the research. As the modulation frequency they can achieve by themselves does not yet reach the frequencies required by 4G and 5G, supply modulation is a field where topologies are routinely combined to cancel out each others weaknesses. This is true for both the supply modulators where linear modulator and switch mode modulators can achieve higher baseband bandwidths, as well as the PA itself where a high efficiency topology such as the Doherty PA can be combined with supply modulation.
References


Developments in Supply Modulation - Conclusions


Developments in Supply Modulation - Conclusions


Developments in Supply Modulation - Conclusions


Developments in Supply Modulation - Conclusions


Developments in Supply Modulation - Conclusions


Developments in Supply Modulation - Conclusions


Developments in Supply Modulation - Conclusions


Developments in Supply Modulation - Conclusions


Chapter 4

GaN-on-Si HEMTs in Supply Modulated PAs

Contents

4.1 Characterisation ............................................................. 123
  4.1.1 Load-Pull ............................................................... 123
  4.1.2 Characterising Intrinsic Capacitances ............................. 131
  4.1.3 De-embedding of Non-linear Elements ............................ 133
  4.1.4 Small-signal Characterisation and Modelling ..................... 135

4.2 Gain Variation in GaN HEMTs ............................................ 137
  4.2.1 Characterising Gain Variation ...................................... 137
  4.2.2 Impact on Supply Modulated PAs .................................. 143
  4.2.3 Origins of Gain Variation ........................................... 144

4.3 Optimising GaN HEMT Geometry for Supply Modulated PAs ........ 151
  4.3.1 Optimising the Length of the Gate Field Plate .................. 152
  4.3.2 Optimising the passivation thickness ............................. 153

4.4 Conclusions ................................................................. 154

References ................................................................. 154

Parts of the analysis, measurements and conclusions in this chapter have been published in [17]–[19], [402], [403].

As discussed in chapter 2, a transistor is the centre piece of the RF PA and therefore plays a critical role in determining its behaviour. Chapter 3 showed that with the additional requirements introduced by the modulated supply voltage, the transistors characteristics become even more important. As a result, transistors for supply modulated architectures need to be characterised in more ways than transistors for conventional PAs, and this will be discussed in section 4.1. One of the main issues that transistors in supply modulated architectures face is supply voltage dependent gain variation. Section 4.2 will discuss this problem, and propose a new method of comparing different transistors in terms of their gain variation and analyse a transistor in detail to, for the first time, find the origin of gain variation. Based on this, ways to improve and optimise transistors for supply modulation applications are presented in section 4.3.
4.1 Characterisation

As the transistor in supply modulated PA is being used over a supply voltage range, it has to be characterised at more than just one supply voltage. This adds another dimension to conventional load-pull measurements which will be discussed in section 4.1.1. The operation over a supply voltage range also requires a more extensive parasitic capacitor characterisation and modelling, this will be shown in section 4.1.2. Using these modelled non-linear capacitances, the non-linear de-embedding to the current generator plane and the role the different capacitances play will be discussed and demonstrated in section 4.1.3. The supply voltage dependence is not limited to large-signal measurements however. As linear analysis is sufficient in some cases, using small-signal models can be advantageous due to their relative simplicity. Section 4.1.4 discusses such a multi-bias small-signal model of a GaN HEMT which includes the effects of both gate and source field plates.

4.1.1 Load-Pull

As transistors in supply modulated PAs are operated at different supply voltages, the load-pull measurements for the transistor characterisation also need to be conducted at multiple supply voltages. If the supply modulator can only provide discrete values, as is the case with the multi-level converters discussed in section 3.2.2, the choice of supply voltages for characterisation is straightforward. If the supply voltage is not discrete, the question of which supply voltages the transistor needs to be characterised at deserves more consideration. By using small steps between the supply voltages at which the transistor is measured, the characterisation can be done comprehensively. This comes at a cost, however. Increasing the number of step increases the number of individual measurements which can take a considerable amount of time and yield a significant amount of data which need to be evaluated, requiring a trade-off.

Designed for the characterisation of transistors at static supply voltages, the active harmonic load-pull setup discussed in section 2.3.3 allows an arbitrary impedance environment to be presented to the transistor. If combined with a probe station, active load-pull systems allow measurements of unpackaged devices at the device plane itself, Fig. 4.1.1 shows such a setup comprising a Keysight PNA-X, four signal generators and 4 amplifiers. The setup is limited to CW signals, presenting exclusively static, single frequency impedance environments.

Active Mesuro load-pull measurements in detail

As mentioned, the active load-pull setup used for most measurements in this thesis is based around a Keysight N5242 PNA-X non-linear vector network analyser (VNA) which is designed for a wide range
Figure 4.1.1: Picture of the active harmonic load-pull measurement system used for the device-plane measurements in this thesis.

of linear and non-linear measurements. As heart of the load-pull setup, it is used to measure the magnitude and phase of the voltages and currents at the transistor at different frequencies. To be able to get to the signals to the PNA-X, two directional couplers are connected between the signal sources on either side and the gate and the drain of the transistor. An accurate measurement requires three calibration steps: absolute power calibration, phase calibration and vector calibration. If the DUT is fitted with a SMA connector, the calibration sequence is not important. As most measurements in this thesis were conducted on-wafer, the power and phase calibration need to be conducted first at SMA level before the probe station is connected, the vector calibration is then conducted at wafer level. The PNA-X is connected to a Keysight U8488A USB power meter to enable the PNA-X, which is very accurate at measuring relative power, to accurately measure absolute power. The phases of the voltages and currents at the different frequencies are critical but can not be measured directly as the PNA-X measures the frequency components at different points in time, an external Keysight U9391C comb generator provides the information necessary to measure the phase accurately. After these steps, the probe station is connected and the system is vector calibrated to be able to calculate the voltages and currents at the transistor from those measured at these directional couplers. As the name vector calibration implies, both magnitude and phase are being accounted for. The vector calibration process depends on the transistor fixture. Most measurements in this thesis were conducted on-wafer in a probe station in the setup shown in Fig. 4.1.1. To calibrate to the pads of the transistor, an on-wafer SOLT calibration kit is used; in this case a calibration kit on an Alumina substrate is used. Once calibration is complete, the DUT is placed on the chuck where it is kept in place by the under pressure generated by a vacuum pump. This facilitates placing the ground-signal-ground
probes on the pads of the transistor as shown in Fig. 4.1.2. Lowering the probes onto the pads on both sides inserts the DUT into the calibrated measurement setup. As mentioned in section 2.3.3, the impedances at the transistor can then be determined by injecting a signal into the device ports. As the power is measured in the directional couplers and deembedded to the calibration plane, the power generated in the signal generators does not have to be calibrated. During the measurement, the power is measured and adjusted to achieve the desired amplitude and phase at pads of the transistor itself. The DC power supplied to the transistor is supplied and measured by a Agilent 6622A precision power supply.

**Accuracy of the active Mesuro load-pull measurements**

The measurement accuracy of the active load-pull measurement depends on the accuracy of the PNA-X itself, that of the DC supply and that of the calibration. As all measurement instruments are temperature dependent to a certain degree, the room in which the equipment is located, is kept at 25 °C. Under these conditions, the DC supply is specified to have a voltage readback accuracy of 0.05% + 50 mV and a current readback accuracy of 0.1% + 8 mA [404]. The power meter used for the calibration is accurate to ±0.52% with an additional power linearity error of ±0.5% [405], as the power meter is not matched perfectly, the mismatch, specified as VSWR of 1.06, introduces an additional error as the mismatch results in a reflected voltage of 3% which is superimposed at an unknown phase. Quantifying the accuracy of the PNA-X is not trivial. A major source of error is the directivity of the couplers which is between 40 dB and 50 dB, assuming no errors during the calibration [406]. As a result, the measurement accuracy decreases as the measured impedance moves away from a matched
Figure 4.1.3: Measured load-pull contours of an E008 GaN-on-Si HEMT presented with a class-F impedance environment at different supply voltages: green at 30 V, blue at 20 V, orange at 10 V, grey points are the measured points. (a) PAE in % and (b) Output power in dBm

condition [407]. A second source of error that is hard to quantify is the repeatability of the SOLT calibration. Not only does it rely on the standard being exactly as it was during characterisation and the accuracy of that characterisation, it also relies on a reproducible connection in the measurement system. While the phase calibration is important to be able to evaluate the waveforms qualitatively, a measurement error of the phase does not reduce the power measurements at the individual frequencies.

Load-pulling transistors for supply modulated PAs

While not entirely representative of the behaviour of the transistor with a modulated input signal and a modulated supply voltage, these static measurements form a valuable basis for the ET design process as they offer a lot of information. Fig. 4.1.3 shows the results of such a measurement, which included an input power sweep for each of the targeted impedances, shown as grey dots in the figure. The figure shows the output power and PAE contours of a bare die 0.25 μm GaN-on-Si HEMT designed by Sheffield university using their E008 process. This measurement took approximately 30 hours, the transistor was presented with a class F environment at each of the impedance and input power steps, which required specific second and third harmonic impedances. Here, the second harmonic was set to a short circuit while the transistor was presented with an open circuit at the intrinsic current generator plane at the third harmonic. In the Mesuro setup used for this measurement, this required a number of iterations for the setup to find the correct combination of amplitude and phase at the harmonics, especially at low power levels where the low harmonic content coming from the transistor makes it challenging to accurately set the impedance. These contour plots contain a lot of information. The PAE is above 74% in a large area at 30 V, while at 10 V, the maximum efficiency is between 70% and
72\% which is consistent with the theory as a reduction in supply voltage at a constant knee region reduces the achievable efficiency. As the contour with the maximum efficiency moves with supply voltage, the PA can not be at its most efficient for all supply voltage, requiring a trade-off. This is also demonstrated in Fig. 4.1.4 which shows that the output power decreases with supply voltage as expected, additionally the region of highest output power moves towards lower impedances as the maximum current swings stays while the maximum voltage swing decreases. Unlike with the PAE, there is no reason to seek a trade-off of the output power as a reduced power at a reduced supply voltage is desired. Assuming a fixed impedance environment, the output power in the ideal case is directly related to the square root of the supply voltage, i.e. for a supply voltage of $V_D = \varsigma \cdot V_{DD,max}$, the output power is $P_{out} = \varsigma^2 \cdot P_{out,VDD,max}$ where $\varsigma$ is a scaling factor, $0 \leq \varsigma \leq 1$. A reduction from 30 V to 20 V should result in an output power reduction of -3.5 dB, a further supply voltage reduction to 10 V in an output power reduction of -9.5 dB relative to the output power at 30 V. Figure 4.1.3b, however, shows that the behaviour of the output power depends significantly on the load impedance. At around 400+j*125 Ω, the output power behaves as expected, dropping by 3.5 dB at 20 V and 9.2 dB at 10 V; for lower impedances, e.g. 265+j*0 Ω, the power drops by 2 dB at 20 V and 7 dB at 10 V. This is important as it means that the transistors needs an even lower voltage to efficiently produce the desired output power which in turn requires the supply modulator needs to cover a wider voltage range. During the design process, the parameters that need to be considered are output power, PAE at the different supply voltages and output power reduction with supply voltage.
In the case of the transistor with the measurements results in Fig. 4.1.3, a suitable impedance might be $250+j*100 \, \Omega$, offering high output power at 30 V, a trade-off of the efficiencies at 10 V, 20 V and 30 V, and an output power reduction of 2.7 dB at 20 V and 7.8 dB at 10 V. The figure shows that at this impedance, no measurement was performed, which offers two alternatives. Generating a model that allows interpolation or resorting to analysing the performance at the closest measured impedance, which in this case would be $265+j*84 \, \Omega$. In this case, the latter option was chosen, the measured gain and PAE at this impedance are shown in Fig. 4.1.5. As these are the raw measurement data, the power and efficiency values look as expected. In the figure, the measurement results at additional supply voltages of 15 V and 25 V are shown, they were not included in the contour plots to keep the contour plots legible.

**Design of calibration and measurement setup**

Transistors are available unpackaged either as single bare die or packaged in either a plastic or ceramic package. A single die, such as the transistor used in the previous measurement, has only the parasitic elements of the transistor itself, allowing relatively easy deembedding to the current generator plane, $I_{DS}$ in Fig. 4.1.6. Any type of transistor packaging will introduce additional elements and Fig. 4.1.6 demonstrates the difference between the two. In a packaged device, deembedding to the current generator is significantly more complex due to the added parasitic components that can not be easily measured. Bare die transistors, such as the one measured in Fig. 4.1.3, are typically measured using ground-signal-ground (GSG) probes with pads that are very close to the transistor, which has the additional advantage that the measurement system can be easily calibrated to the device plane. For packaged transistors, this is less trivial. Not only is it impossible to directly calibrate to the device plane, it is even challenging to calibrate to the package plane as measurement systems have coaxial connectors which necessitates a calibration kit that connects to the coaxial system and is able to shift

![Image](image-url)

**Figure 4.1.5:** Measured gain (dotted lines) and PAE (solid lines) over output power characteristics of a GaN HEMT for different supply voltages at $Z_L = 265+j*84 \, \Omega$
Figure 4.1.6: Comparison of bare die and packaged device, showing device and package parasitics

the reference plane of the measurement system to the package plane. A calibration and measurement kit is developed to allow shifting the reference plane of the measurement system to the package of the transistor. To be able to accommodate different types of transistors and allow using the fixture as a carrier for complete PAs, it is designed in a modular fashion. The connections to the measurement systems are on the heatsink, the connection PCBs on input and output side also offer four SMA connectors that can be used to supply bias voltages or acquire measurement signals. The heatsink conducts the heat away from the transistor and has the additional function of providing structural support, every other component is attached to it, see Fig. 4.1.7(c). The DUT and the calibration standards are inserted on a copper carrier into a cavity in the heatsink, as shown in Fig. 4.1.7(a).

The cavity allows the insertion of different types of carriers which makes it suitable for transistors with different packages. In the case of a transistor packaged in an SMD package, the carrier would be planar, with the PCB and the packaged transistor sitting on top of it. For transistors with a flanged package, a carrier with a small cavity for the transistor’s flange can be inserted with the PCB surrounding the transistor. In the case of a bare-die transistor being used, it can be bonded to the carrier. In this case, the area the die is bonded to can be raised to the same level as the PCB, reducing the parasitic inductance between transistor and PCB. For each of these options, the package plane would be at a different position, necessitating separate calibration standards. Using the characterised open, short, load and through standards in Fig. 4.1.7(b) allows a calibration to the exact position of the package plane in the case of the packaged device used in this measurement. To connect the carrier PCB to the connector PCBs, small PCBs with GSG traces are pressed onto the traces on either side and clamped down, reliably connecting them. The calibration ensures that the impedance transitions at the interface between the connector PCB and the carrier PCB and between
SMA connector and connector PCB are calibrated out which means that their exact impedance is not critical. By fabricating all calibration standards and the DUT PCB on the same substrate and in the same orientation, any anisotropies or variations in the dielectric constant of the substrate can be calibrated out as well. To be able to supply the transistor with the appropriate gate and drain supply voltages, a bias T is added on each side of the transistor. While the DC current into the gate is negligible and a low power bias T is sufficient, the current levels on the drain side are significantly higher, so the bias T is realised using two 3 dB hybrid couplers back-to-back with an additional baseband termination to provide a well-defined low-impedance at low frequencies.
4.1.2 Characterising Intrinsic Capacitances

With their ability to present arbitrary impedances to the transistor under test, active load pull systems can be used to provide short circuits at the device plane to directly characterise the intrinsic non-linear capacitances using Y-parameter measurements. To characterise the capacitances, one port of the transistor is excited with a single frequency, low power RF signal for different gate and drain bias voltages, while the other port is presented with a very low impedance at the same RF frequency. Under these conditions, RF voltages and currents are measured. Presenting highly reflective impedances to either port of the transistors can lead to instabilities, therefore the measurements are limited to gate voltages that pinch off the channel completely. This pinched-off operation is also critical to limit the dissipated power in the transistor. While measuring the s-parameters can yield the same result without the limitation of the complete pinch-off, the used system allowed y-parameter measurements with no additional effort. This set of measurements was conducted at 900 MHz. As the intrinsic capacitances are a result of the device geometry, the calculated values are independent of frequency as long as the materials the device is made up of are non-dispersive and the calculations take all intrinsic components into account. This series of measurements results in the IV traces shown in Fig. 4.1.8, they can be easily converted to the capacitance if the circuit can be assumed to be linear, a valid assumption if the excitation is small enough. If the dynamic drain voltage is negative enough, the transistor starts to rectify the excitation signal, limiting the measurement range in the negative supply voltage range.

As in a reciprocal network, $Y_{DS} = Y_{22} + Y_{12}$ and $Y_{12} = \frac{I_{GS}}{V_{DS}}$, $Y_{22} = \frac{I_{DS}}{V_{DS}}$, the capacitors impedance can be calculated as $\Im(Y_{DS}) = \Im\left(\frac{I_{GS} + I_{DS}}{V_{DS}}\right)$, for simplicity’s sake, the peak-to-peak voltage and current is used for this calculation, a valid assumption as long as the system can be assumed to be linear.

$$C_{DS} = -\frac{\Im(Y_{DS})}{\omega}$$ (4.1)
and

\[ C_{GD} = - \frac{3m(Y_{12})}{\omega} \]  \hspace{1cm} (4.2)

The measurements yield the capacitances in Figs. 4.1.9 and 4.1.10. Based on these measurements, the capacitances are modelled mathematically, following an approach similar to the one presented in [323], using \( \tanh() \) functions. This modelling is necessary to be able to conduct the non-linear deembedding. Fig. 4.1.9 shows a good agreement between measurements and model in the area where the measurements were conducted. \( C_{DS} \) can be seen to increase dramatically for a small supply voltage, a result of the dynamic drain voltage reaching the point where the transistor starts rectifying. The calculation uses peak-to-peak values, Fig. 4.1.8 shows that the rectification increases the peak-to-peak current considerably. If \( C_{GD} \) is plotted as a function of \( V_{DS} \), the resulting curves move with applied gate voltage, as shown in Fig. 4.1.10b. By plotting \( C_{GD} \) over the applied voltage between gate and drain, \( V_{DG} \), it becomes clear that \( C_{GD} \) is not a function of \( V_{DS} \) but exclusively of \( V_{DG} \), as demonstrated in Fig. 4.1.10b.
4.1.3 De-embedding of Non-linear Elements

Waveform engineering is based on the design of voltage and current waveforms at the intrinsic current generator plane [408]. While the Mesuro load-pull system supports de-embedding, its de-embedding is limited to linear networks and can thus not take non-linear capacitances into account. This means that the measured data, an array of multi-harmonic voltages and currents at the device plane, needs to be post-processed to obtain the intrinsic waveforms. As the linear series inductances have been calibrated out, the voltage which is measured by the load-pull system is very close to the voltage at the current generator plane. To obtain the current at the current generator plane, $i_D$, the current through the capacitances has to be subtracted from the measured current as shown in Fig. 4.1.11.

$$i_D = i_2 - i_{Cds} - i_{Cgd} \quad (4.3)$$

Due to the non-linearity of the capacitances, this de-embedding has to be conducted in the time domain. As the time step can be set to be sufficiently small, the differentiated voltage can be approximated by calculating the difference in voltage between two discreet points and multiplying it by the sampling frequency. The capacitances are modelled as described in section 4.1.2, they depend on the applied drain and gate voltage. In the algorithm, shown below, the time varying capacitances $C_{DS}$ and $C_{GD}$ are calculated and multiplied with the time-varying differentiated voltages $\Delta V_D$ and $\Delta V_g$, yielding the currents needed to calculated the intrinsic drain current.

```python
for j in range(0, time_max):
    c_ds[j] = 0.98*((0.15 - 0.0015*v_ds[j]) * math.tanh(v_ds[j] + 5 - (3.5/2.5*v_gs[j] + 6)) - 0.15)
    c_gd[j] = 0.164 - 0.1 * math.tanh(0.3 * (v_gd[j])) - 0.032 * math.tanh(0.2 * (v_gd[j] - 14)) - 0.000374 * v_gd[j]
    i_ds[j] = c_ds[j] * 1e-12 * ((v_ds[j] - v_ds[j-1]) * f)
    i_gd[j] = c_gd[j] * 1e-12 * (((v_ds[j] - v_ds[j-1]) - (v_gs[j] - v_gs[j-1])) * f)
    i_t = i_t - i_ds - i_gd
    i_f = fftp.fft(i_t)
```

This drain current can also be transformed back into the frequency domain, enabling the calculation of the harmonic impedances. To demonstrate the de-embedding, the results from the capacitance...
characterisation in section 4.1.2 and a measurement of the same device operated as a rectifier are used as an example. Using a rectifier to test the de-embedding has the advantage that the current waveform has two well-defined points, the two zero crossing points of the voltage waveform where the current waveform has to be zero as well due to Ohm’s law. To evaluate the impact of the non-linear deembedding, the measured current waveform is compared to the waveform where a linear $C_{DS}$, a non-linear $C_{DS}$ and both non-linear $C_{DS}$ and $C_{GD}$ have been deembedded as shown in Fig. 4.1.12. The figure demonstrates that without de-embedding, little information can be drawn from the current waveform, making it hard to judge whether the current waveform at the current generator plane has the intended shape. The linear de-embedding is able to partially reveal the intrinsic current waveform, but struggles around the zero current transitions. The condition of simultaneous zero voltage and zero current is not met with a linear de-embedding. Using the modelled $C_{DS}$, the waveform only changes slightly, becoming flatter in the transistor’s off-state, the voltage and current still do not simultaneously cross zero however. The small difference between linear and modelled $C_{DS}$ is expected as $C_{DS}$ does not change dramatically with the applied voltages, as shown in Fig. 4.1.9. If the current flowing through $C_{GD}$ is taken into account as well, the zero crossings happen almost simultaneously. This ability to accurately determine the waveforms at the intrinsic current generator plane demonstrates the use of non-linear de-embedding in waveform engineering. It also shows however that in the absence of reliable models of the intrinsic capacitances, a linear de-embedding will be a reasonable first guess. The difference of the current waveform in the time-domain also results in a different current in the frequency domain and, with it, a different impedance. Figure 4.1.13 demonstrates the impact the different de-embedding strategies have on the information that can be derived from the drain impedances. In a rectifier, the transistor serves as the RF load, therefore the fundamental impedances should all be purely resistive. Figure 4.1.13 shows that with a non-linear de-embedding,
the impedances are significantly closer to fulfilling this condition, again demonstrating the superiority of non-linear de-embedding over the linear one.

### 4.1.4 Small-signal Characterisation and Modelling

While large signal measurements are crucial to evaluate transistors, especially if a PA is to be designed around them, some characteristics can be evaluated in a small-signal model which is easier to understand, modify and model. While it is intrinsically unable to exhibit non-linear behaviour as it is build of linear building blocks, a transistor can be measured and modelled for a number of discrete supply voltage points, reproducing the supply voltage dependence of capacitances, resistors and current sources. This modelling requires a number of s-parameter measurements of the transistor, they are all combined to form the model with its fixed and varying elements as depicted in Fig. 4.1.15. The measurements used here are forward biased and pinched off at 0 V to reveal capacitances and series resistances, and measurements at a class AB bias point for a number of supply voltages, allowing the full modelling.

The usual small-signal equivalent circuit models [125] do not consider field plates specifically. The prevalence of gate and source field plates in AlGaN/GaN HEMTs as a tool to mitigate trapping effects, see section 2.2, makes it necessary to extend the usual model to accommodate them. This extends the models proposed by [84], which proposed separate models for gate and source field plates. Additionally, the model needs to be able to work at multiple supply voltages to make it applicable to transistor for supply modulated applications. As many transistors have both gate and source field plates, see Fig. 4.1.14, the model needs to represent this. Combining the commonly used small-signal

![Figure 4.1.13: Drain impedances of a transistor operated as a rectifier at the fundamental frequency for different power levels with (a) no deembedding, (b) linear deembedding and (c) full non-linear deembedding](image-url)
GaN-on-Si HEMTs in Supply Modulated PAs - Characterisation

Figure 4.1.14: Schematic of an AlGaN/GaN HEMT including gate and source field plate. Elements based on [86], [125]

Figure 4.1.15: Small-signal equivalent circuit of a HEMT with gate and source field plate, elements visible in Fig. 4.1.14 are enclosed in the dotted rectangle, non-linear components that change with applied voltages are marked with an arrow

equivalent circuit model [125] with the model in [84] results in the model shown in Fig. 4.1.15.

To examine whether this model is sufficient to describe the behaviour of a GaN HEMT with gate and source field plate, it is used to model a Sheffield E008 GaN-on-Si HEMT with a gate length of 0.25 µm and a gate width of 250 µm, comparable to the transistor shown in Fig. 4.1.16.

The s-parameters of the transistor are measured at \( V_D = 0 \) V, for both pinched-off and conducting bias points. Additionally, the s-parameter measurements are conducted for different supply voltages. The extrinsic parameters can be fitted by focussing on the measurements at \( V_D = 0 \) V first as in these cases, some elements of the model can be neglected. The following step is supplementing them with the other measurements in an approach comparable to the one in [126], [409], [410]. The supply voltage dependent \( g_m \) is directly taken from a static DCIV measurement, the initial values of the equivalent circuit components were based on the device geometry as far as possible allowed reducing the risk of non-physical values: the starting values of \( R_S \) and \( R_D \) could be estimated from the physical separation of the ports and the measured sheet resistance of the wafer. The drain resistance changes with supply voltage however, as the voltage has a direct impact on the width of the depletion region
GaN-on-Si HEMTs in Supply Modulated PAs - Gain Variation in GaN HEMTs

4.2 Gain Variation in GaN HEMTs

To achieve perfect envelope tracking, the output power has to be independent of the supply voltage [11]. This requires the gain to be independent of the supply voltage, as, for a constant input power, a change in gain will result in a change in output power. Real transistor will have a gain that changes with supply voltage, this section will discuss gain variation; how to characterise it, its impact on supply modulated PAs, its physical origin, as well as ways to reduce it.

4.2.1 Characterising Gain Variation

It is easy to observe that gain changes with supply voltage, as shown in Fig. 4.2.1. To be able to evaluate it, this observation needs to be complemented with more rigorous definitions and measurements. This is not a straightforward process as gain variation depends on a number of factors including source and load impedance and gate bias. These will be discussed and a guideline that sets the conditions that allows meaningful comparison will be proposed. The first factor to be examined is the dependence

Figure 4.1.16: Microscope image of a 0.25 µm gate length GaN-on-Si HEMT with a gate width of 250 µm, 2 * 125 µm

[86]. Fig. 4.1.15 provides an overview over the non-linear elements, indicated by red arrows. The fitting results are compared to the measurement results in Fig. 4.1.17. Model and measurement model are similar for all extreme cases: conducting and pinched-off at $V_D = 0$ V, as well as at the lowest and highest supply voltage of $V_D = 10$ V and $V_D = 30$ V, respectively, see Fig. 4.1.17. This demonstrates that the model can be used to represent a real transistor with gate and source field plate in static, linear multi-bias simulations.
of gain variation on load impedance, see Fig. 4.2.1. In these two figures, the transistor can be seen to exhibit very different gain variation at different load impedances. Fig. 4.2.2 illustrates this further by plotting gain variation contours. As the measurements sweep supply voltage, load impedance and input power, the dimension of the data is reduced by averaging the gain over power levels. This introduces a small error due to different compression levels at different load impedances and supply voltages but as most gain values are acquired in the back-off region, this error is small enough to not present a problem, especially since this plot is only illustrative. The contour plot demonstrates that for the measured transistor, the gain changes by between 6 dB and 13 dB in the measured load impedance range and over the supply voltage range of 10 V to 30 V. The gain variation in impedances that provide low output power and low efficiencies is irrelevant as these would not be chosen in PA design anyway. As discussed in section 4.1.1, choosing the load impedance for a supply modulated PA is not trivial, depending on the system requirements and the personal choice of the PA designer.
Figure 4.2.1: Measured gain (dotted lines) and power added efficiency (solid lines) over output power characteristics of a Sheffield E008 GaN HEMT for different supply voltages and with a different fundamental load impedances: (a) 121+j*51 Ω and (b) 265+j*84 Ω

Nonetheless, studying the gain variation at a $Z_{\text{opt}}$ that provides a trade-off between power and efficiency allows a meaningful, albeit not perfect, characterisation and, importantly, allows a meaningful comparison of transistors. This importance to choose a suitable $Z_{\text{opt}}$ becomes obvious when comparing the impedance of lowest gain variation, in this case $Z_1 = 122+j^*51$ Ω, with an impedance that might be chosen for a design due to the trade-off between efficiency and output power, in this case $Z_2 = 265+j^*84$ Ω, the impedance presented in Fig. 4.2.1; without the output power and efficiency information, $Z_1$ might be used as a reference for comparisons. This reliance on $Z_{\text{opt}}$ as a way to describe gain variation also means that comprehensive characterisation of the gain variation requires large signal measurements, that provide the data necessary to choose a suitable $Z_{\text{opt}}$.

The second factor that impacts gain variation is the gate bias. Typically, the gate voltage of a PA remains fixed during operation. The quiescent current and the transconductance $g_{\text{m}}$ however depend on the supply voltage, see Fig. 4.2.3(a). As a result of that, a variation in supply voltage leads to a variation in threshold voltage and, with it, a change in conduction angle which directly impacts the class of operation for the PA. As only class A and class B have a theoretically linear gain [12], the PA’s gain characteristic will change with the supply voltage [12], [243]. The measured gain curves in Fig. 4.2.1 illustrate that. In these, the gain stays close to constant at 10 V, indicating class B, and decreases with power level for supply voltages larger or equal than 15 V, with the decrease in gain
Figure 4.2.2: Contours showing the difference in gain in dB between supply voltages of 30 V and of 10 V of a Sheffield E008 GaN-on-Si HEMT at 1 GHz. Grey dots represent measured impedances, stars show the impedances for maximum efficiency and power at 10 V and 30 V. All measurements conducted under CW excitation, gain averaged over power levels.

Figure 4.2.3: Measured static transconductance and transfer characteristics; (a) $g_m$ vs. $I_D$ and (b) $I_D$ vs $V_{GS}$ of a Sheffield E008 GaN HEMT for different supply voltages.

Compression increasing with the supply voltage, showing different degrees of class AB behaviour.

This also drives home the point that large-signal measurements are required to characterise gain variation, this dependence on the chosen bias point and the output power is only observable in large-signal measurements. It also means that any definition of gain variation needs to consider power level and bias point to be meaningful.

The third challenge is the input impedance of the transistor. There are different definitions of gain and not all of them take the input matching into account. As long as the transistor input impedance stays constant with supply voltage, the gain variation does not depend on the chosen definition of gain. If, however, the transistor input impedances changes with supply voltage, the gain definition needs to take that into account. In LDMOS devices, there is little change due to the nearly constant gate-source capacitance $C_{GS}$ and the low gate-drain capacitance $C_{GD}$ [243], [411]. GaN HEMTs however have a
considerably larger gate-drain capacitance that additionally changes considerably with $V_{DS}$ \cite{27}, \cite{323} and with $V_{GS}$ \cite{323}, leading to a large change in input impedance. Additionally, $C_{GS}$ changes with $V_{GS}$ \cite{323} and $V_{DS}$ \cite{27}, \cite{323}, resulting in an input impedance that changes with power level as well as with supply voltage, see Fig. 4.2.4. The figure demonstrates that the input impedance changes significantly with supply voltage and input power level, with the impedances converging for high input powers.

The last challenge in charactering the gain variation is that it depends on the supply voltage range. Taking Fig. 4.2.1(b) as an example again, the gain drops by 1.5 dB for a supply voltage drop from 30 V to 20 V. When reducing the supply voltage further, the gain drops more dramatically. Obviously, a large gain variation in a region the transistor is not going to be operated in is not an issue. The desired range of supply voltages depends on the targeted communications signal, the PA should be able to provide a high efficiency down to the average of the signal, requiring the transistor to be characterised over a range of the PAPR of the signal.

All of those challenges need to be addressed to allow a meaningful comparison of different transistors. Due to the variety of the challenges, a set of guidelines detailing measurements and operating conditions is the logical approach \cite{17}. The transistor has to be:

1. Matched to a load impedance that provides a good trade-off between power and efficiency
2. Biased for a flat gain at the lowest supply voltage in the range of interest
3. Matched on the input for PAs, in which case transducer gain should be compared, whereas in load-pull measurements the power gain should be compared

Under these conditions, the gain variation is obtained by comparing the gain traces at 1 dB compression for all supply voltages. A demonstration of how these guidelines can be used to characterise gain variation can be seen in Fig. 4.2.5, where a gain variation of 10 dB is observed. The guidelines are used to compare the gain variation of transistors in literature. As different publications use different...
Flat gain at lowest supply voltage

Impedance that offers trade-off between PAE and $P_{out}$

Gain Variation

1dB compression

Figure 4.2.5: Demonstration showing how to evaluate the gain variation of a transistor according to the proposed guidelines

Voltage ranges, the gain variation needs to be normalised to take the voltage range into account. Equation (4.4) demonstrates a simple way to achieve this normalisation:

$$GV_{\text{norm},V} = \frac{\text{Gain Variation (dB)}}{\text{Voltage Range (dB)}}$$ (4.4)

It results in values close to zero for devices exhibiting little gain variation and gain variation becoming more significant with increasing value of $GV_{\text{norm},V}$.

Table 4.1: Comparison of gain variation reported in literature

<table>
<thead>
<tr>
<th>Transistor / Process</th>
<th>Technology$^a$</th>
<th>Voltage Range</th>
<th>$GV^b$</th>
<th>$GV_{\text{norm},V}^c$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wolfspeed CGH40010F</td>
<td>GaN-on-SiC</td>
<td>16 - 28 V, 5 dB</td>
<td>4 dB</td>
<td>0.8</td>
<td>[366]</td>
</tr>
<tr>
<td>- ” -</td>
<td>- ” -</td>
<td>10 - 28 V, 9 dB</td>
<td>7 dB</td>
<td>0.78</td>
<td>[256]</td>
</tr>
<tr>
<td>Wolfspeed CGH27015</td>
<td>GaN-on-SiC</td>
<td>5 - 25 V, 14 dB</td>
<td>8 dB</td>
<td>0.57</td>
<td>[268]</td>
</tr>
<tr>
<td>MACOM NPT1012</td>
<td>GaN-on-Si</td>
<td>14 - 28 V, 6 dB</td>
<td>2 dB</td>
<td>0.33</td>
<td>[98]</td>
</tr>
<tr>
<td>MACOM NPTB00004</td>
<td>- ” -</td>
<td>12 - 28 V, 7 dB</td>
<td>2 dB</td>
<td>0.29</td>
<td>[224]</td>
</tr>
<tr>
<td>FBH</td>
<td>GaN-on-SiC</td>
<td>7.5 - 40 V, 15 dB</td>
<td>7.5 dB</td>
<td>0.5</td>
<td>[304]</td>
</tr>
<tr>
<td>RFMD/Qorvo RF3934</td>
<td>GaN</td>
<td>15 - 60 V, 12 dB</td>
<td>4 dB</td>
<td>0.33</td>
<td>[373]</td>
</tr>
<tr>
<td>Qorvo</td>
<td>GaN-on-SiC</td>
<td>10 - 20 V, 6 dB</td>
<td>8 dB</td>
<td>1.3</td>
<td>[298]</td>
</tr>
<tr>
<td>Sheffield E008</td>
<td>GaN-on-Si</td>
<td>10 - 30 V, 10 dB</td>
<td>7-10 dB</td>
<td>0.7-1</td>
<td>Fig. 4.3.2</td>
</tr>
<tr>
<td>Mitsubishi</td>
<td>GaN</td>
<td>5 - 25 V, 14 dB</td>
<td>11 dB</td>
<td>0.8</td>
<td>[362]</td>
</tr>
<tr>
<td>Eudyna FLL120MK</td>
<td>GaAs</td>
<td>3 - 10 V, 10 dB</td>
<td>4 dB</td>
<td>0.4</td>
<td>[374]</td>
</tr>
<tr>
<td>Avago ATF50189</td>
<td>GaAs</td>
<td>1.5 - 6.5 V, 13 dB</td>
<td>2.5 dB</td>
<td>0.2</td>
<td>[224]</td>
</tr>
<tr>
<td>Motorola PFP9045</td>
<td>LDMOS</td>
<td>15 - 30 V, 6 dB</td>
<td>1.5 dB</td>
<td>0.25</td>
<td>[412]</td>
</tr>
<tr>
<td>?</td>
<td>GaAs HVHBT</td>
<td>3 - 28 V, 19 dB</td>
<td>5 dB</td>
<td>0.26</td>
<td>[293]</td>
</tr>
</tbody>
</table>

$^a$ HEMT unless specified otherwise, $^b$ Gain Variation, $^c$ Voltage normalised Gain Variation
Table 4.1 compares the gain variation and the normalised gain variation of transistors reported in literature. Measurements presented in literature have not necessarily been conducted under the guidelines presented here; they should not be far off however. It is safe to assume that the PAs presented in literature have been optimised for a trade-off between output power and efficiency. The table demonstrates that gain variation is present in all devices of all technologies. In GaN HEMTs, however, the range of normalised gain variation is significant, between 0.29 and 1.3. Due to the low number of non-GaN HEMTs, unfortunately, no meaningful statement can be made.

4.2.2 Impact on Supply Modulated PAs

Gain is prominent in two characteristics of PAs, the AM/AM distortion, see section 2.6 and the PAE. As the gain directly effects the PAE as discussed in section 2.4.1, a reduction in gain will reduce the PAE. If the gain drops with supply voltage, the achievable PAE will drop with the supply voltage as well. Depending of the gain variation exhibited by the device, the useful voltage range might have to be limited, reducing the usefulness of the PA for signals with high PAPR [173], [256] as the PA will spend most of the time in the low power area where the supply voltage and thus the PAE are low. To avoid this, the gain at the lowest supply voltage has to be high enough to guarantee a reasonable PAE. This reduction in gain limits the frequency range over which the device can be used efficiently, depending on the minimum supply voltage.

An increase in the AM/AM distortion is another direct result of gain variation. As the supply voltage depends on the power level, the gain at low power levels will be lower than at high power levels. As the PA is usually driven into a certain degree of compression to achieve efficiency targets, a small amount of gain variation is beneficial in reducing AM/AM distortion [366]. In the case of a PA driven 1 dB into compression, a gain variation of 1 dB can be used to linearise the PA and achieve a flat gain [257] by adapting the shaping function. In the case of multi-level converters with few supply voltage levels, the step changes will result in significant distortion as the switch between two output voltage levels, the step changes will result in significant distortion as the switch between two output levels results in a switch between two levels of gain [413].

The shaping function is also impacted by gain variation. As discussed in section 3.1.2, it relates the instantaneous magnitude of the input envelope to the dynamic supply voltage. The discussed shaping functions depend on the gain of transistor not changing too much with supply voltage. A shape similar to the shapes presented in section 3.1.2 might not be realisable with a transistor with large gain variation, as shown in Fig. 4.2.6(a). In this example, the input power necessary to achieve high efficiency is independent of the supply voltage. As a result, the shaping function required to provide reasonable efficiency over the power range, shown as thick, black line in Fig. 4.2.6a, looks fundamentally different from the ones presented in section 3.1.2. The 'zig-zag' curve that is observable is a result of
10 V to 30 V in 5 V steps, the quiescent current is kept at 60 mA or 240 mA/mm. Biased at this supply voltage, the small-signal gain is measured and analysed. During these measurements, the input impedance and the bias current are kept constant, eliminating the established other causes of gain variation. The s-parameters are then measured at supply voltages from 10 V to 30 V in 5 V steps, see Fig. 4.2.1 so s-parameters are able to measure the small-signal gain variation. Using s-parameters the input impedance variation and the bias current are kept constant, eliminating the established other causes of gain variation.

4.2.3 Origins of Gain Variation

Section 4.2.1 described some of the causes of gain variation, namely input impedance variation and moving threshold voltage. To trace down the origin of gain variation, the s-parameters, i.e. the small-signal behaviour of the transistor is measured and analysed. During these measurements, the input impedance and the bias current are kept constant, eliminating the established other causes of gain variation. The prevalence of gain variation under these small-signal conditions can be observed in Fig. 4.2.1 so s-parameters are able to measure the small-signal gain variation. Using s-parameters also means that other large signal effects such as thermal effects or trapping do not impact the measurements, modelling and analysis. The s-parameters are then measured at supply voltages from 10 V to 30 V in 5 V steps, the quiescent current is kept at 60 mA or 240 mA/mm. Biased at this quiescent current, the $g_m$ is close to its maximum at 5 V slightly decreasing with increasing supply voltage.
Figure 4.2.7: Measured extrinsic maximum stable gain/maximum available gain (MSG/MAG) (solid lines) and current gain $|h_{21}|$ (dotted lines) of a GaN HEMT for different supply voltages

Voltage as a result of the self heating of the HEMT which reduces the transconductance at higher supply voltage, see Fig. 4.2.3. The s-parameters measured under these conditions were then analysed to allow comparison of current gain and the two relevant power gain figures, maximum stable gain (MSG) in the region where the transistor is not unconditionally stable and maximum available gain (MAG) where it is. Figure 4.2.7 compares these gain figures for the different supply voltages. The current gain decreases with increasing supply voltage by about 35% from 8.2 GHz to 5.2 GHz, a direct result of the decreasing $g_m$. The power gain figures increase with increasing supply voltage, showing the expected gain variation results also present in the large signal measurements. The MSG for frequencies up to around 7 GHz drops by 6 dB, the $f_{\text{max}}$ from 18.2 GHz to 12.7 GHz, a drop of 30%, when the supply voltage is reduced from 30 V to 5 V. One important observation is that the drop in gain is not linear with the change in supply voltage. The MAG/MSG is almost identical for supply voltages between 5 V and 15 V while there is a significant difference between the curves at 15 V and 20 V. The small-signal gain variation is present over the complete frequency range which indicates that the root cause is independent of the frequency and can be examined at any frequency. In the frequency range up to 7 GHz, the transistor is not unconditionally stable, the gain figure of interest in this range is the maximum stable gain, while above 7 GHz the maximum available gain is the relevant figure. As the definition of the maximum stable gain is simpler, as shown in (4.5) [95], the investigations into the origin of gain variation will be conducted in the sub-7 GHz range where the stability factor $K < 1$.

$$G_{\text{MSG}} = \frac{|S_{21}|}{|S_{12}|}$$

(4.5)

The equation shows that in that frequency range, only gain, $S_{21}$, and isolation, $S_{12}$, are required to calculate the maximum stable gain. This can be broken down further by converting the s-parameters.
to Y-parameters [414].

\[
G_{MSG} = \frac{|Y_{21}|}{|Y_{12}|} \quad (4.6)
\]

In an intrinsic transistor, the two Y-parameters \(|Y_{21}|\) and \(|Y_{12}|\) are associated with \(Z_{GD}\) and \(g_m\) [415]. Real transistors are more complex and have a number of parasitic components in addition to the intrinsic elements in the simple model [415], influencing \(G_{MSG}\). As all of the non-linear elements bar \(R_S\) and \(R_D\) are intrinsic to the transistor, the definition in (4.6) is sufficient in describing the maximum stable gain which means that \(Z_{GD}\) and \(g_m\) are the dominant factors in small-signal gain variation. The following sections will therefore discuss the impact of these two factors to establish the origin of small-signal gain variation.

### 4.2.3.1 Role of the transconductance \(g_m\)

Figure 4.2.3 on page 140 shows the impact of the drain voltage on the transconductance, showing that \(g_m\) changes little for low currents and starts to diverge as soon as the drain current reaches 10 mA. For currents exceeding this value, \(g_m\) differs significantly for different supply voltages. The higher the supply voltage, the lower the current at which \(g_m\) peaks and starts to decrease, a result of the self heating of the transistor [246]. As both the s-parameters showing the gain variation and \(g_m\) have been measured under the same static and thermal conditions, the \(g_m\) of the transistor during the s-parameter measurements can be assumed to be equal to that acquired during the DCIV measurements.

The bias point for the large signal measurement that yielded the results shown in the previous figures was chosen to result in a flat gain at the lowest supply voltage, see dotted vertical line in Fig. 4.2.8. At this bias point, the \(g_m\) values for voltages between 15 V and 30 V are almost identical while the gain measured in the large signal measurement changes by around 7 dB at the trade-off impedance Fig. 4.2.1 over the same voltage range. This is a first indication that \(g_m\) is not the root
cause of the small-signal gain variation. To explore this further, the model presented in section 4.1.4 is fitted to the measured transistor and, since it is fully parametrised, used to vary the transconductance in simulation. This is used to compare the transistor’s behaviour for other bias points. In addition to the constant current bias point employed in the s-parameter measurements, the bias point at which the large-signal measurement was conducted, \( V_{GG} = 5.4 \) V and a bias point at which the \( g_m \) values are close together for supply voltages of more than 10 V, \( V_{GG} = -4.8 \) V. At \( V_{GG} = -4.8 \) V, the \( g_m \) has its maximum at 10 V and drops with increasing supply voltage, see Fig. 4.2.9. In the simulation, the impact of the \( g_m \) on \( f_{\text{max}} \) is used to determine its impact on gain variation. The simulation demonstrates that while varying the bias point and thus the \( g_m \) has an impact on \( f_{\text{max}} \), any change in \( g_m \) by adjusting the bias point merely moves the \( f_{\text{max}} \) curves around as a whole without changing their shape too much. At the bias point of \( V_{GG} = -4.8 \) V and at the large signal bias point, \( g_m \) decreases when increasing the supply voltage from 10 V while \( f_{\text{max}} \) increases. This is significant as it demonstrates that \( g_m \) is not only not the source of the small-signal gain variation but, depending on the chosen bias point, may even partially reduce the small-signal gain variation. As the transconductance determines the current gain, \( g_m \) has a big impact on \( f_T \), see Fig. 4.2.10. By changing the bias point, \( g_m \) changes

![Figure 4.2.9: Comparison of \( f_{\text{max}} \) and \( g_m \) over supply voltage for the measurement with constant drain current and the simulation for two gate bias voltages, -4.8 V and -5.4 V](image1)

![Figure 4.2.10: Comparison of \( f_T \) and \( g_m \) over supply voltage for the measurement with constant drain current and the simulation for two gate bias voltages, -4.8 V and -5.4 V](image2)
and $f_T$ changes with it. As expected there is a close relation between the two in Fig. 4.2.10. This shows that current gain and $f_T$ are not useful indicators when examining devices for gain variation.

4.2.3.2 Role of the feedback capacitance $C_{GD}$

Another possible factor in gain variation is the feedback capacitance $C_{GD}$. It is extracted using the process that was used to populate the small-signal model in Fig. 4.1.15 which results in the supply voltage dependent values shown in Fig. 4.2.11 where their behaviour over supply voltage is shown.

The figure demonstrates that as $C_{GD}$ decreases with supply voltage, $f_{max}$ increases. The region in which the changes are most significant, i.e. between 15 V and 20 V, is the same for both $C_{GD}$ and $f_{max}$, too. This behaviour would be even more pronounced for a flat $g_m$, see Fig. 4.2.9. This correlation is the first indicator pointing to $C_{GD}$ as a major contributor in small-signal gain variation and, with it, gain variation as a whole. Using the small-signal model of the transistor, the assumption that $C_{GD}$ is a major contributor can be explored in simulation by manipulating $C_{GD}$ and evaluating the results of this manipulation on the gain variation. $C_{GD}$ is manipulated while keeping all other values of the model the same. This approach of adjusting $C_{GD}$ is purely theoretical as adjusting $C_{GD}$ would require changing the transistor geometry, resulting in additional changes. Also, the capacitances $C_{GD}$ and $C_{GS}$ are dependent to a certain degree, see Fig. 4.2.12, a behaviour that has been reported in literature [27], [416]. Adjusting one without the other therefore means that the results of the simulation are only going to be valid in a limited range and thus will need to be treated carefully. As the main focus in this experiment is the MSG, the results will still be relevant in the range where the stability factor is smaller than one as here, the gain primarily depends on $g_m$ and $C_{GD}$ as shown in equation (4.5). The results obtained at frequencies where the maximum gain is defined by the MAG will be not meaningful however, as in this region, keeping $C_{GS}$ constant is not a good approximation any more. Using this approach of manipulating $C_{GD}$ allows exploring how the shape of the $C_{GD}$ profile impacts the MSG. Aiming for minimal small-signal gain variation is an obvious goal, see Fig. 4.2.13(a),

![Figure 4.2.11: Extracted gate-drain capacitance $C_{GD}$ and measured transconductance $g_m$ and maximum oscillation frequency $f_{max}$ versus supply voltage](image)

Figure 4.2.11: Extracted gate-drain capacitance $C_{GD}$ and measured transconductance $g_m$ and maximum oscillation frequency $f_{max}$ versus supply voltage
the shape to achieve that can be seen in Fig. 4.2.13(b). Achieving it would mean reducing $C_{GD}$ significantly while keeping the general shape the same. As discussed, $f_{\text{max}}$ is not meaningful due to the limitations of the approach. Another interesting $C_{GD}$ profile is shifting the increase in $C_{GD}$ towards lower supply voltages, see Fig. 4.2.14. This results in the MSG decreasing with supply voltage significantly more gradually than in the original case and is comparable to the transistors with lower gain variation presented in table 4.1. The gain drops slightly as the supply voltage is reduced and decreases significantly once $C_{GD}$ surges. By exclusively changing $C_{GD}$, the small-signal gain variation can be changed significantly in the MSG region, demonstrating that $C_{GD}$ is the major origin of the small-signal gain variation that can be observed in many GaN HEMTs.

To track the origin of small-signal gain variation down to the physical level, the transistor was simulated in TCAD, a semiconductor device modelling tool, at Sheffield University [17] to demonstrate how $C_{GD}$ changes with supply voltage and what factors are responsible for that change. The simulation
Figure 4.2.14: Comparison of extracted $C_{GD}$ and $C_{GD}$ modified to represent a $C_{GD}$ curve shifted to start increasing at lower voltages: (a) Simulated MSG/MAG in dB for extracted (dotted line) and modified (solid line) for different supply voltages and (b) comparison of the two $C_{GD}$ shapes versus supply voltage.

Figure 4.2.15: TCAD simulation results of $C_{GD}$ of a GaN HEMT with a gate field plate length of 1 $\mu$m, 0.5 $\mu$m and with no gate field plate versus supply voltage at $V_{GS} = -5$ V, $C_{GD}$ plateau is labelled in blue.

...results in Fig. 4.2.15 show how the gate field plate is responsible for the increase in $C_{GD}$ with decreasing supply voltage. As discussed in section 3.3.1, the field plate acts as a MISHEMT [84] which depletes the channel under the field plate [86] depending on the drain gate voltage. For low supply voltages, the channel is fully populated, $C_{GD}$ starts to plateau here as shown in Fig. 4.2.15. As the field plate increases in lengths, so does the size of this capacitor, the relationship between gate field plate length and additional $C_{GD}$ due to it is almost linear as long as fringing fields are negligible.

Experimental GaN-on-Si HEMTs with no GFP were fabricated to further verify that gain variation is indeed a result of the GFP. Removing the GFP completely removes the plateau in the $C_{GD}$ trajectory as shown in Fig. 4.2.16a, $C_{GD}$ only increase slightly with decreasing supply voltage which is an effect of the depletion region shifting towards the source contact as the applied drain-source voltage decreases. Figure 4.2.16b shows that, as expected, the small-signal gain now only varies slightly in the absence of gate-field-plates. It additionally shows that the MAG/MSG changes in line with $C_{GD}$, showing little...
4.3 Optimising GaN HEMT Geometry for Supply Modulated PAs

Identifying $C_{GD}$ as the source of gain variation opens pathways to optimising GaN HEMTs for operation in supply modulated PAs. As the trajectory of the $C_{GD}$ has a significant impact on the transistor’s behaviours, the ability to change it will allow the optimisation of the GaN HEMT itself for different applications. As shown in Fig. 4.1.14, the major source of $C_{GD}$ is the gate-coupled field plate (GFP). Field plates reduce the trapping effects at the GaN cap/passivation interface as discussed in section 2.2, and thus reduce the RF/DC dispersion [80], [122] which is important to keep the PA operating efficiently as an expanding knee region will decrease the efficiency, output power and linearity. This decrease in all relevant RF PA parameters illustrates why field plates are significant in GaN HEMT design. There are a few parameters in field plates that can be adjusted to trade-off knee
walkout for a more favourable $C_{GD}$ trajectory. Figure 4.3.1 shows the parameters of FPs that can be adjusted to modify the $C_{GD}$ trajectory: the length of the GFP, as already shown in Fig. 4.2.15, and the thickness between the GFP and the GaN cap. While there are other parameters that determine the knee walkout, e.g. the length of the source-coupled field plate (SFP) and the passivation thickness between SFP and barrier, these have little impact on $C_{GD}$ but will have a significant impact on $C_{GS}$, and with it, the achievable gain.

4.3.1 Optimising the Length of the Gate Field Plate

Reducing the length of the GFP will directly reduce the overlapping area of GFP and channel, thus reducing the area of the parallel plate capacitor they form. This directly reduces the magnitude of the $C_{GD}$ plateau, as shown in Fig. 4.2.15. It will however increase the RF/DC dispersion [80]. This assumption is verified by comparing two GaN-on-Si HEMTs that have GFPs of different lengths but are identical apart from that. The devices are measured under identical conditions, i.e. bias point, drive levels and impedance environment. The measurement results in Fig. 4.3.2 show that the power
levels achievable with the two devices are comparable. The gain however changes when changing the length of the GFP, increasing by around 2 dB for high supply voltages, 4 dB at 15 V and around 3-4 dB at 10 V. This means that the gain variation as defined in section 4.2.1 decreases by around 2 dB. The increased gain at low supply voltages also directly leads to an increase in PAE, e.g. at 10 V, the maximum PAE increases by about 5%. Reducing the length of the GFP increases the RF/DC dispersion however, as expected, due to the increase in trapping effects. Fig. 4.3.3 demonstrates this, showing that the knee voltage of the transistor with the reduced GFP has a higher knee voltage at all supply voltage measurement points. As shown in Fig. 4.3.2, this increase in knee voltage does not significantly impact the efficiency. As the knee region is more pronounced, the linearity might be impacted however. This impact will be subject to further research. As in modern communications systems, the PAPR is quite high, the supply modulated PA will operate at lower supply voltages for most of the time. As a result, a high RF/DC dispersion might be a smaller issue in these applications than a high gain variation. The trade-off between knee-walkout and gain variation is therefore very application specific. An additional limit to reducing the GFP length is that most GaN HEMTs have a T-gate consisting of the gate itself and a wider strip of metal on top of it to reduce the gate resistance and increase reliability. The protruding top bar of the T-gate acts as a field plate, limiting the minimum realisable field-plate length.

4.3.2 Optimising the passivation thickness

Another parameter in Fig. 4.3.1, which influences the $C_{GD}$ is the distance between the barrier layer and the GPF, i.e. the thickness of the passivation. The voltage at which $C_{GD}$ transitions from the plateau to the lower value is determined by the pinch-off voltage of the MISFET, i.e. voltage that depletes the channel under the GFP. Changing the distance between channel and GFP by changing the passivation thickness under the GFP, changes the voltage at which $C_{GD}$ plateaus. By reducing
Figure 4.3.4: TCAD simulation showing the behaviour of $C_{GD}$ for a constant gate field plate configuration and different SiNx passivation thicknesses between field plate and barrier versus supply voltage at $V_{GS} = -5$ V. The thickness, the plateau can be moved towards lower supply voltages, as shown in Fig. 4.3.4. As the GFP and the channel form a parallel plate capacitor, reducing the thickness of the passivation will increase the magnitude of the plateau of $C_{GD}$. This approach could be used to move the plateau towards lower voltages where it has less impact.

4.4 Conclusions

This chapter demonstrated that not all transistors are equal when it comes to using them in supply modulated PAs. The supply voltage dependent non-linearity of the intrinsic capacitances result in the power and efficiency contours shifting with supply voltage and reduces the gain at reduced supply voltages. This gain variation introduces issues in supply modulated PAs and impacts linearity, linearisability, and restricts the shaping functions. Introducing a way of comparing the gain variation of different devices that were measured by different people under different conditions, allows a meaningful comparison. This comparison demonstrates that, while all devices have a small gain variation and many GaN HEMTs have a significant gain variation, some GaN HEMTs show a significantly lower gain variation, demonstrating that it is not an inherent characteristic of GaN HEMTs, but a result of the transistor geometry. For the first time, the origin of the gain variation has tracked down and identified as the gate field plate. The gate field plate inherently leads to a supply voltage dependence of the gate–drain capacitance $C_{GD}$, the geometries of the field plate and the passivation below it determine the degree to which $C_{GD}$ varies. This opens a new pathway, allowing addressing gain variation by changing the geometry of the gate field plate, offering an opportunity to optimise devices for use in supply modulated PAs.
GaN-on-Si HEMTs in Supply Modulated PAs - Conclusions

References


Chapter 5

Input matching Supply Modulated PAs

Contents

- 5.1 Concept ................................................................. 159
- 5.2 Source-pull simulations .............................................. 160
- 5.3 Verification ............................................................. 163
  - 5.3.1 Design of PA ..................................................... 163
  - 5.3.2 Simulation ......................................................... 164
  - 5.3.3 Measurements .................................................. 164
- 5.4 Conclusions ........................................................... 167
References ................................................................. 168

Parts of the analysis, measurements and conclusions in this chapter have been published in [17]–[19].

PA design for supply modulated systems is fundamentally the same as designing it for conventional PAs. Two matching networks match input and output impedance of the transistor to the system impedance, in the case of a multi-stage PA there is an additional interstage matching. The additional considerations that need to go into the output matching network (OMN) design were discussed in section 3.3. This chapter will discuss the input matching network and its impact on linearity in supply modulated PAs. The role of the input matching network (IMN) is to ensure the power available from the system is used in the transistor and not reflected back to the source. This means that it’s design has little impact on the output power and the drain efficiency of the PA, as a result it is not usually treated in detail in PA design. It is directly responsible for the gain of the PA however and, as a result of that, for the PAE. Due to this perceived insignificance of the input matching, little attention is paid to its impact on PA performance apart from gain, despite the significant non-linearities of the gate impedance $Z_G$ of the transistor, see [243]. While there have been single publications in for conventional PAs [244] and the Doherty PA [417], the impact of the input matching network on linearity in GaN HEMT supply modulated PAs is explored here for the first time. Section 5.1 will present a novel concept to optimise the input matching network design to exploit the non-linearities in GaN HEMTs. Source-pull simulations will be used to explore the concept further in section 5.2, the concept is verified in simulation and measurement in section 5.3.
5.1 Concept

The typical PA design flow first establishes a suitable output matching network (OMN) that offers a good compromise between output power, efficiency, linearity, and bandwidth. The IMN is then designed around this combination of transistor and OMN with the intention to provide the maximum possible gain and thus PAE \[95\]. Using conventional, passive components with constant values, this results in a fixed impedance trajectory presenting a fixed impedance at each frequency. The input impedance of a transistor however depends on three parameters in a PA; power level, supply voltage and load impedance. While the load impedance plays a significant role in load modulated architectures \[417\], it is not significant in supply modulated PAs due to the fact that it is a fixed output impedance. This reduces the parameters that can change the transistor’s input impedance to power level and supply voltage. This input impedance variation is a direct result of the intrinsic elements of the transistor changing with the applied voltages as discussed in section 4.1.2, which leads to a change in \(\Gamma_{\text{in}}\), see (5.1) \[95\].

\[
\Gamma_{\text{in}} = S_{11} + \frac{S_{12} S_{21} \Gamma_L}{1 - S_{22} \Gamma_L} \tag{5.1}
\]

This demonstrates that devices with a large variation in \(S_{12}\), i.e. a large variation in \(C_{\text{GD}}\), will have a dynamic \(\Gamma_{\text{in}}\) if the device can not be considered to be an unilateral device. This is the case if the impedance of \(C_{\text{GD}}\) is small enough to allow a significant amount of feedback to happen. From \[18\], \[323\], the variation of \(C_{\text{GD}}\) with the applied gate-drain voltage is known to be significant in GaN HEMTs which results in varying real and imaginary parts of the input impedance, \(Z_{\text{in}}\), as shown in Fig. 5.1.1. Obviously, a passive, static matching network will not be able match all of these impedances, necessitating a trade-off. Additionally, the phase of the input impedance changes, which will manifest itself as AM/PM distortion of the PA. By choosing the impedance \(Z_S\) carefully, it should be possible to reduce the AM/PM distortion in supply modulated PAs, extending the work done on that topic in other PA topologies \[244\], \[417\].
5.2 Systematic investigation using source-pull simulations

To explore the impact of $Z_S$, a source-pull simulation was set up to systematically vary the $Z_S$ and analyse its impact on the performance of the PA. The PA in this simulation is based on the Wolfspeed CGH40010F 10 W GaN HEMT, and has an OMN that is designed for high efficiency operation at 2.9 GHz, following the approach in [257] to sacrifice high efficiency at full supply voltage to increase efficiency at lower supply voltages. In this case, a fundamental load impedance of $12 + j7 \ \Omega$ is chosen to achieve this efficiency trajectory. To ensure stability, a stabilisation network between device and source impedance is included in the simulation, see Fig. 5.0.1 and Fig. 5.2.1. After the optimum impedance for maximum gain at 28 V is established to be $3 + j16 \ \Omega$, $Z_S$ is swept around this optimum, the real part from $1 \ \Omega$ to $19 \ \Omega$ and the imaginary part from $-j5 \ \Omega$ to $j35 \ \Omega$. For this simulation, the output matching network and thus $Z_L$ is kept constant and the supply voltage is swept from 8 V to 28 V in steps of 4 V and input power levels from -15 dBm to 27 dBm. With four sweep parameters, these multi-dimensional results need to be analysed and simplified to make them accessible.

The small-signal gain is measured at the highest supply voltage of 28 V as this is the most obvious starting point when designing the PA. As the maximum gain occurs at a conjugate match [95], the maximum gain occurs at the impedance where the resistive part of the IMN is equal to that of the stabilised transistor and the imaginary parts are conjugates of each other, resonating each other out, see Fig. 5.2.3. Therefore, moving from the optimum impedance in any direction leads to a drop in gain. In the case in Fig. 5.2.3, the gain drops by 4 dB when increasing the reactive parts by $j10 \ \Omega$ in any direction, the same drop in gain can be observed when changing the resistance from $3 \ \Omega$ to $18 \ \Omega$. The higher the Q-factor of the impedance $Z_{in}$, the more sensitive the gain is to variations in the
reactive part of $Z_S$. By increasing the resistive part of the impedance, and thus reducing the Q-factor, the sensitivity of the gain to imaginary part mismatch can be reduced.

The small-signal gain variation is obtained by calculating the difference between the small-signal gain at the highest supply voltage and at the lowest supply voltage at the minimum input power. As a result of this definition as the difference, the small-signal gain variation decreases as the impedance moves towards the conjugate matched impedance at the minimum supply voltage, $Z_{S,\text{Opt,Min}} = 3+j*14 \, \Omega$, as the gain at the highest voltage decreases while that at the lowest increases. Any move of the imaginary part towards $Z_{S,\text{Opt,Min}}$ will reduce the small-signal gain variation while a move in the other direction will always increase it.
Just like the gain variation, the phase variation is obtained by calculating the difference between the highest and the lowest phase value at every impedance at the minimum input power. The simulation results show that the higher the Q-factor, i.e. the lower the resistive part of the impedance, and the closer the reactive part is to the resonance, the higher the phase variation. The phase variation drops in almost all directions from the impedance of highest gain at maximum supply voltage, its dependence on the supply reactance decreases with resistance.

All of the simulation results demonstrate that small-signal gain, small-signal gain variation and phase variation depend significantly on the source impedance $Z_s$ presented to the device. The simulations also show that the impedance of maximum gain at 28 V, an obvious target impedance for PA designers indicated by a red triangle in the figures 5.2.3-5.2.5, is not the optimum impedance for linearity, especially the phase variation is significant. The simulations also show that gain at 28 V can be traded-off for reduced small-signal gain variation and phase variation. One impedance that offers such a trade-off is the one marked by a blue star in 5.2.3-5.2.5. As this trades-off gain for linearity, it will also have a negative effect on PAE.
5.3 Verification

The previous simulations demonstrated the impact of the IMN at an abstract level. Input and output impedances were defined by equations and the simulation accuracy depends on the accuracy of the model. To explore how different input impedances translate into PA performance, particularly for signals exceeding the small-signal conditions, a PA is tested with different IMNs, presenting different impedances to the transistor.

5.3.1 Design of PA

The PA is designed based on the load impedance used in the previous simulation, the input impedances will be based on the data obtained in the source-pull simulation. One IMN of particular interest is the one designed for maximum gain at the maximum voltage of 28 V, $IMN_{\text{max}}$ with $Z_{S,\text{max}} = 3 + j16 \, \Omega$. A second IMN of interest, $IMN_{\text{opt}}$, trades-off gain for linearity, the impedance of this optimised IMN has a significantly higher resistive part ($Z_{S,\text{opt}} = 10 + j15 \, \Omega$). The matching networks are designed using transmission lines and series capacitors, avoiding the use of shunt elements and ground vias for everything but the DC decoupling. The DC decoupling capacitors are integrated into the matching network, the RF grounding are chosen as 10 pF and 100 pF in parallel, the baseband termination is further improved by an electrolyte capacitor. Physical gaps allow the configuration of the different IMNs, shortening the meander and choosing the series capacitor allow for a modification of the optimised IMN while adjusting the length of the open stub provides the same possibility for the IMN optimised for gain.

Figure 5.2.5: Simulation results showing the phase variation versus input reactance and resistance. Red triangle: Source impedance providing the highest gain at 28 V; Blue star: An impedance that offers a good trade-off between gain and linearity.
5.3.2 Simulation

This PA is simulated at 2.9 GHz for static supply voltages, shown as thin lines in Fig. 5.3.2, and with envelope tracking, shown as bold lines. The shaping function used for the ET simulation is a first order power envelope tracking shaping function as proposed in [256] with the boundary voltages of 8 V and 28 V. When changing from the maximum gain IMN to the linearity improved IMN, the small-signal gain drops by approximately 2 dB at 28 V and by 1 dB at 8 V, leading to a small-signal gain variation that is reduced by 1 dB. For the shaping function used in this case, the large signal gain variation increases slightly by around 1 dB.

The small-signal phase variation decreases from 29° to 18°. For large signal excitation, as shown in Fig. 5.3.2b, the phase variation reduces from 57° to 32°. The PAE decreases by about 3% as a result of the reduced gain for static supply voltages. For varying supply voltages, the PAE difference between the different IMNs reduces to less than 1% over the whole output power range, due to the shaping function used. The simulations demonstrate that the IMN has indeed a significant impact on the linearity of a PA and can be optimised for linearity at the cost of efficiency. It also shows that while the PAE at static supply voltages decreases, the impact in an ET PA might be less significant.

5.3.3 Measurements

While the Wolfspeed device model predicts PA performance well if biased at a supply voltage of 28 V [418], the simulation results in Fig. 5.3.2 diverge significantly from the measured ones in [256] while following the same general trend. This demonstrates that measurements are necessary to verify the impact of the IMN on the performance of the ET PA. The PA from Fig. 5.3.1 is therefore realised, see Fig. 5.3.3, and measured for a number of combinations of IMN and supply voltage; for all points, the PA was driven 5 dB into compression. The measurement is conducted using the coupler based measurement part of the active harmonic load-pull setup shown in Fig. 4.1.1 but using passive loads.
Figure 5.3.2: Simulation results comparing the (a) Gain, (b) Phase and (c) PAE of a GaN HEMT PA versus input power for static supply voltages between 8 and 28 V in 4 V steps for the two IMNs, $IMN_{\text{max}}$ (thin blue, solid lines) and $IMN_{\text{opt}}$ (thin orange, dotted lines). Bold line shows PAE trajectory of PA with envelope tracked supply voltage, blue for the maximum gain at 28 V, red for the optimised IMN.

Instead of amplifiers and calibration on an SMA level instead of the on-wafer calibration that was used for the measurements in section 4.

In this setup, four different IMNs were realised, IMN1 as a network and IMN2-IMN4 using the approach of modifying the transmission line length and the series capacitance: IMN1 is close to the $IMN_{\text{max}}$ which is reflected in the input reflection coefficient of -13 dB at 28 V, see Fig. 5.3.4a, it changes significantly with supply voltage until it reaches -6 dB at 10 V. All other IMNs stay within 1 dB over the whole supply voltage range, with $IMN_{4}$ changing the least. While the reflection coefficient of all other IMNs moves with output power, it stays almost constant for $IMN_{4}$, albeit at -5 dB showing a considerable mismatch. Fig. 5.3.4b also show that the phase of all input reflection...
Figure 5.3.3: Picture of the fabricated PA with different input matching structures, designed to be easily reconfigurable by soldering connections and replacing lumped capacitors.

Figure 5.3.4: Input reflection coefficient $S_{11}$ versus output power for different input matching networks and different supply voltages: (a) magnitude in dB, (b) plotted in a Smith chart where 10 V is displayed using the darkest colour and 30 V using the lightest colour.

coefficients change significantly with power. Fig. 5.3.5 shows that while the gain at 28 V is very close to that in the simulation, the small-signal gain variation is significantly more distinct than it is in the simulation. The level of gain variation is consistent with literature [256]. In the case of IMN1, the network optimised for high gain, the gain drops by 9 dB, from 17 dB to 8 dB, when the supply voltage is reduced from 28 V to 10 V. This has a catastrophic effect on the PAE, see Fig. 5.3.6b, as at 2 dB compression, the gain of at 10 V is 6 dB which leads to a PAE of merely 35% despite a drain efficiency of 50%. IMN2 slightly reduces the gain variation by improving the gain at 10 V slightly at the cost of gain at higher supply voltages while IMN3 and IMN4 maintain the trajectories of IMN1, albeit at a reduced level.

As a result of the gain being much lower in the measurements than in the simulations, the PAE looks significantly worse in the measurement, especially for low supply voltages, see Fig. 5.3.6b. The
5.4 Conclusions

This chapter demonstrated how much gain and linearity of a supply modulated GaN PA depend on the supply voltage, the input power and the source impedance the transistor is presented with. This dependence is usually undesired as it results in an input impedance mismatch as soon as the operating conditions deviates from that the one the PA was designed at. The fact that the gain and linearity
change directly points towards a way to optimise the input impedance to improve the PA’s behaviour under supply modulation, resulting in a novel concept to improve linearity. This trade-off is first shown theoretically, demonstrating that the gain and thus PAE at the maximum supply voltage can be traded-off to increase amplitude and phase linearity. By simulating a PA with matching networks that present the desired impedances to the transistor, this approach is further tested before it is realised as a prototype. The measurement on that prototype prove that the concept works and use several matching networks to explore their impact on PAE and linearity. With little to no impact in terms of PAE, the linearity can be increased considerably. This is particularly relevant in transistors that have a high gain at their operating frequency, here even more gain could be traded-off for linearity, albeit at the cost of input mismatch. While this technique was demonstrated on a GaN HEMT, it would also work with other transistor technologies as long as their input impedance changes with the supply voltage making it potentially interesting in GaAs HEMT supply modulated PAs.

References


Chapter 6

RF Supply Modulators

Contents

6.1 Characterising Transistors as Rectifiers ............................................. 171
6.2 Gate Voltage in Transistor-based RF Rectifiers ................................. 173
6.3 Class F RF Rectifiers ........................................................................ 180
6.4 Class B* RF DC/DC Converters .......................................................... 185
  6.4.1 Concept ...................................................................................... 185
  6.4.2 Simulations ................................................................................. 188
  6.4.3 Rectifier ..................................................................................... 191
  6.4.4 Realisation .................................................................................. 195
6.5 Conclusions ....................................................................................... 197
References ............................................................................................... 198

Parts of the analysis, measurements and conclusions in this chapter have been published in [20], [21].

As discussed in section 3.2, RF DC/DC converters have a few advantages that make them interesting from the perspective of integrated ET PAs. One of them is their high operating frequency which allows filtering of the RF frequency with relatively small inductors or transmission line filters, facilitating integration with the ET PA on one substrate. Another advantage is the potentially achievable modulation frequency range, although this has not been demonstrated so far. While some of the most current modulator topologies struggles with modulation bandwidths of 100 MHz, this bandwidth would be considered narrow-band in if it was centred around a carrier of 2 GHz. Even modulation bandwidths of 1 GHz, beyond anything proposed at the time of writing, corresponds only to a fractional bandwidth of 20% at 5 GHz. These theoretical bandwidths assume that both PA and rectifier can be operated efficiently over the whole bandwidth, which is not an unrealistic assumption considering the high efficiencies that have been demonstrated over multi-octave bandwidths as discussed in section 2.4.6. As an RF modulator consists of two parts, PA and rectifier, the DC/DC efficiency is the product of the two individual efficiencies, $\eta_{\text{DC/DC}} = \eta_{\text{PA}} \cdot \eta_{\text{Rectifier}}$. While it has been demonstrated that PA and rectifier can be designed using the same general approach of utilising harmonic terminations [356], the main losses in a RF DC/DC converter are generated in the RF PA [354]. Unlike in PAs for communication applications, the linearity of the PA is not relevant, removing some constraints in PA design.
RF Supply Modulators - Characterising Transistors as Rectifiers

This chapter will start by evaluating the relevant considerations when selecting the RF frequency. The characterisation of RF rectifiers will be discussed in section 6.1, and used, resulting in measurements that allow establishing the role of the gate voltage in RF rectifiers in section 6.2. After that, the impact of harmonic terminations in RF DCDC converters is discussed using a class F rectifier in section 6.3. A new approach to achieving high DC/DC efficiencies based on second harmonic generation and injection is going to be presented in section 6.4.

Selecting the RF frequency

As with the buck-converters in section 3.2, the modulation frequency is not directly related to the RF frequency of the RF DC/DC converter, however, selecting the RF frequency is a major design decision in the modulator design. The considerations are, that it needs to be filtered out at the output, it needs to be high enough to easily accommodate the modulation bandwidth and that it has to allow high efficiency operation. The first two of these requirements call for high frequencies. If modulation frequency and RF frequency are a decade apart, a single LC filter will reduce the RF frequency components at the modulated output by around 40 dB, if the difference between the two frequencies increase further, the RF frequency level can be reduced further with simple filter structures. Another reason to choose a high RF frequency is that it makes it easier to cover the bandwidth; a fractional bandwidth of 5% is easier to realise than one of 20%. The last requirement, high efficiency, is easier realised at low frequencies however, there is a resulting need to trade-off efficiency, bandwidth and ripple. As discussed in section 3.1.5, the ripple needs to be filtered out to avoid mixing products in the passband of the PA. By choosing the ripple frequency carefully, the frequencies of the mixing products of the ripple frequency and the carrier can be moved to a frequency band that can be terminated at the output of the PA, e.g. using a band-pass filter. This partially reduces the impact of the ripple on the linearity of the PA, which means that if the operating frequency is chosen with a PA with a certain bandwidth in mind, the filter can be simplified or the efficiency can be improved. One example for this would be using an RF frequency of 2.9 GHz for a PA for 5G applications which corresponds to a gap in the spectrum that is not currently utilised for 5G as shown in Fig. 2.1.2 on page 13. New spectrum allocations and regional differences need to be kept in mind during the selection process, however.

6.1 Characterising Transistors as Rectifiers

While load-pull is an established technique in PA design for finding optimum impedances and to allow modelling of transistors, there has been little reported work on systematically subjecting transistors
RF Supply Modulators - Characterising Transistors as Rectifiers

Figure 6.1.1: Comparison of simulated class-F PA (red, dotted) and class-F rectifier (blue, solid) load-lines at the current generator plane, grey lines indicate simulated DCIV curves, green boxes indicate quadrants of IV plane

to the impedance environment present in rectifiers. An established way of designing efficient RF rectifiers is designing a PA and feeding its output with the signal that is to be rectified [355]–[357]. However, PAs only operate in the first quadrant while rectifiers operate in both the first and the third quadrant, as shown in Fig. 6.1.1. As the IV curves differ between the first quadrant, where PAs operate and the third quadrant, where rectifiers operate, the optimum impedances are unlikely to be the same. The current in the third quadrant is significantly higher than the one in the first, increasing the achievable current swing, and the voltage swing is increased due to the voltage being limited by the knee region in the third quadrant. With rectifiers being one of the two contributors to conversion losses in RF DC/DC converters, their performance is a key element in efficient supply modulators. As in the load-pull measurements in section 4.1, rectifiers can be characterised in a virtual impedance environment that mimics that of a real circuit. Unlike PAs, however, there are no dedicated measurement systems for rectifiers, the measurement systems used for PAs however can be adapted to be able to characterise rectifiers. In this case, the system introduced in section 4.1.1 is used and modified as shown in Fig. 6.1.2a.

There are distinct changes compared to the characterisation system for transistors for PAs. The DC supply on the drain side is replaced with a DC load with the resistance $R_{DC}$, the voltage drop over this DC load is measured, yielding the load current and load power. The reference input signal in PA characterisation is the fundamental drive signal which means there is no possibility to easily implement fundamental source-pull. In rectifier characterisation, the reference signal is the fundamental power injected into the rectifier on the drain side, allowing the synthesis of arbitrary fundamental impedance on the gate side. As the characterisation system was not designed to be used in that way, automated impedance sweeps do not converge, making measurements very time-consuming as for each measurement point, magnitude and phase of the fundamental on the gate side and the second and third harmonic on the drain side need to be adjusted manually. As the signal generators are not coher-
ent, the phase between them slowly drifts with time, making reproducing impedances time consuming. Additionally, the voltage drop over the DC load can not be measured using the characterisation system as it was not intended to be used in this way, it is measured using a bench-top multimeter.

6.2 Gate Voltage in Transistor-based RF Rectifiers

While both diodes and transistors can be used as rectifying elements at RF frequencies [356], they differ significantly. Rectifiers using diodes only have one input, rectifiers using transistors have both a gate and a drain port. As a result, the dynamic impedance of a diode is only determined by the RF voltage to be rectified, while the dynamic impedance of a transistor depends not only on the RF input voltage, i.e. the drain-source voltage, but also the gate-source voltage. Figure 6.2.1 demonstrates this, showing how the dynamic impedance of the simulated HEMT depend on the applied gate voltage while
that of the simulated diode is fixed. As a result, the efficiency of a transistor based rectifier depends significantly on the gate voltage. To achieve high efficiencies, the gate voltage needs to be below -2.5 V for the transistor to pinch-off in the first quadrant, while it needs to be as close to 0 V as possible in the third quadrant, deviating from this will reduce the efficiency. Transistors have advantages as well, however. While diodes only start conducting once a threshold voltage is exceeded, transistors are able to conduct in the third quadrant from 0 V, giving transistor-based rectifiers the potential to be more efficient than diode-based rectifiers. An additional advantage of transistor-based rectifiers is their power handling capability. Transistors capable of generating hundreds of Watts at microwave frequencies are commercially available and can be used in rectifying circuits while diodes in the RF world are limited to either low power detector diodes or PIN-diodes intended to be used as switches.

To establish the impact of the gate voltage, the case of a static gate voltage is considered, i.e. the gate voltage is equal to the gate bias voltage at all times, e.g. due to a short circuit presented to the transistor at the gate for all frequencies apart from DC. Here, the load-lines follow the IV curves, as shown in Fig. 6.2.2. In this case the transistor can be considered to be a one-port with two bias voltages. For a gate voltage of 0 V, there is no rectification and the transistor itself operates as a the load impedance. The voltage swing is significantly reduced despite the constant input power, meaning that a significant input power mismatch between the power source and the rectifier. At a gate voltage of -2 V, the rectifier does not pinch-off in the first quadrant which would significantly reduce the efficiency. Again, the voltage swing is lower than it is at -4 V, due to mismatch, albeit not as much, showing the impedances are better matched. At -4 V and -6 V, the voltage and current swing are comparable. The main difference is that at -6 V, the maximum current magnitude coincides with a significantly higher voltage magnitude.

The figure demonstrates why it is desirable to have different gate voltages at different parts of
the signal cycle, providing a high impedance if the voltage is positive and a low impedance if the voltage is negative. Ideally, the gate voltage would be at the voltage resulting in the minimum on-resistance for negative drain voltages, here 0 V and at a voltage resulting in a minimum leakage current for positive drain voltages, here any value below -4 V, resulting in a rectangular waveform that switches between these two values. This ideal gate voltage is independent of the input power and allows efficient rectification down to very low input powers. Rectangular signals are hard to realise at microwave frequencies, however, a sinusoidal signal can fulfil the same role at a reduced efficiency.

Self-synchronicity

As discussed in section 3.2.4, there are three common ways of generating the gate signal of the rectifier: driving PA and rectifier separately with signals that can be individually optimised for PA and rectifier, using a power splitter and directing a part of the PA input power to the rectifiers’ gate with the appropriate phase and amplitude adjustment, and lastly using a signal that is generated from the RF signal to be rectified on the rectifier side using either an external feedback network or the intrinsic capacitances of the transistor as a feedback network. If the gate voltage is generated using a feedback loop, a synchronicity between gate voltage and drain voltage can be ensured, hence the term self-synchronicity. Using feedback to generate the gate signal allows the spatial separation to PA and rectifier, and means that PA and rectifier can be simulated, realised, and tested individually, eliminating timing and amplitude alignment issues as long as the feedback network is designed properly. One disadvantage, however, is that the gate voltage depends on the input power while the ideal gate voltage is independent of the input power as discussed previously, this deviation from the ideal case will result in a reduced efficiency in parts of the input power range.

While it is possible to build an external feedback network [360], the intrinsic capacitances of the transistor already form the feedback network shown in Fig. 6.2.3 which can be used to generate the gate signal [351]. By solving the network for the gate impedance $Z_G$, equation (6.1) can be derived.

---

**Figure 6.2.2**: Simulated DCIV curves (dashed lines) and load-lines (solid lines) of a transistor operating as a rectifier with a constant input power for a number of constant gate voltages.
which allows calculating the gate impedance necessary for $v_{Cgs}$ to be a certain value for a value of $v_{DS}$.

$$Z_G = \frac{v_{Cgs}Z_{DG}Z_{GS}}{v_{DS}Z_{Cgs} - v_{Cgs}(Z_{DG} + Z_{GS})} \quad (6.1)$$

Assuming the impedances $Z_{DG}$, $Z_{GS}$ and $Z_{Cgs}$ are dominated by their imaginary part, an assumption, the validity of which depends on the device and the frequency of operation, the equation can be simplified significantly, see equation (6.2).

$$Z_G = -j \frac{1}{\omega_0 C_{DG} \cdot \left( \frac{v_{DS}}{v_{Cgs}} - 1 \right) - C_{GS}} \quad (6.2)$$

The equation shows that as long as the above assumption of negligible resistances holds true, $Z_G$ will be purely imaginary and can be calculated as a ratio of the two voltages, $v_{DS}$ and $v_{Cgs}$. This drastically reduces the source impedances that need to be evaluated to the very border of the Smith chart. A purely reactive impedance agrees with the measurements in [351] where source-pull measurement showed the impedance with the highest efficiency to be at $3.7 + j44.3$, the very border of the measured impedance range. The same measurement also demonstrated that the efficiency increases as it approaches the border of the Smith chart, further confirming that $Z_G$ is overwhelmingly reactive as long as the frequency is sufficiently high. A limitation of using the intrinsic capacitors for feedback is that the intrinsic capacitances that form the voltage divider, $C_{GD}$ and $C_{GS}$, are non-linear and change with the applied voltages in many transistor technologies, including GaN HEMTs, as discussed in section 4. A realisable design, however, requires a single value of $Z_G$ which means that an assumption for an average value for the capacitances needs to be made. The choice of this average value determines at which power level the rectifier operates most efficiently, necessitating a conscious decision of which power level to optimise for, e.g. for medium power levels, where modern communication signals are most probable to be or for high power levels in legacy communication systems or other applications in which the signal has a low PAPR. Another limitation is that the optimum impedance will change with frequency, a possible limitation in broadband self-synchronous rectifiers. Lastly, the analysis
so far only considered the fundamental frequency. The harmonic components generated during the rectification will also be subjected to the voltage divider and thus be present at the gate. This could be either dealt with by presenting the harmonics with a low impedance to keep the gate voltage sinusoidal or by using the harmonics to shape the gate voltage, similarly to the approach in [151] that used second harmonic injection in a PA. Just like the required fundamental impedance, the required second harmonic impedance could be established using (6.2). Any possible increase in efficiency would be at the cost of the additional complexity of the gate impedance network, which would need to operate at multiple harmonics.

**Measurements**

Using the measurement setup discussed in section 6.1, a transistor is measured as a rectifier for different gate impedances and power levels. The measurements are conducted on the Qorvo 6x80 µm 0.25 µm GaN HEMT die at 900 MHz, the same device that was characterised in section 4.1.2. The measurement was conducted in the active load-pull setup shown in Fig. 4.1.1. At the drain port, the transistor is injected with the RF input power that is to be rectified and the second and third harmonic components to achieve a class F operation, the gate is injected with the signal necessary to achieve the targeted gate impedance. Harmonics higher than that are not terminated specifically but measured to the 18th harmonic, the limit of the measurement system. With a maximum voltage swing of around 85 V and a maximum current swing of 0.3 A, the optimum DC load impedance is 

\[ R_{\text{opt}} = \frac{85}{0.3/\pi} = 445 \, \Omega \]

The DC load resistor used in the measurement is chosen slightly smaller to stay close of the breakdown voltage. A resistor with a nominal resistance of 330 Ω is chosen, and is measured to be 328.55 Ω. The gate impedance is kept between \(0.98 \leq |\Gamma_G| \leq 1\) and varied in phase between 0° and 360°, in steps of 30° with additional measurement points between 0° and 60° where the rectifier has its points of maximum efficiency and additionally shows interesting behaviour. The rectifier is measured at two input power levels, one 4 dB lower than the maximum input power and one 25 dB lower than that at -29 dB input power back-off (IBO).

The measurement results in Fig. 6.2.4 demonstrate the impact of phase of the gate reflection coefficient on output voltage and conversion efficiency at two different input power back-off values. As expected from the analysis in this section, the optimum gate impedance changes with the drive power as the intrinsic capacitances change, from \(\angle \Gamma_G = 45°\) at -4 dB IBO to 34° at -29 dB IBO. In addition to moving the phase of high efficiency, an operation at the reduced input power also makes it harder to keep the rectifier efficient as at low input powers, the efficiency drops significantly more when deviating from the optimum phase. At high powers, the efficiency stays above 80% over a gate reflection coefficient phase range of 220° while at low powers, the efficiency drops below 80% if the phase deviates more than a few degrees to either side. To evaluate the effect of the phase more
Figure 6.2.4: Measured DC output voltage (dotted line) and efficiency (solid line) over phase of the input reflection coefficient $\Gamma_G$ at a DC load resistance of $328.55 \, \Omega$ at -4 dB and -29 dB IBO, $|\Gamma_G|$ is kept between 0.98 and 1.

closely, the gate voltages for all the measurement points are plotted in Fig. 6.2.5. Ideally, the gate voltage would be out-of-phase with the drain current, being high where the drain current is flowing and low where it is not. The figures shows them in red, demonstrating that this ideal gate voltage is the exception, with many phases leading to gate voltages that are $180^\circ$ out-of-phase with the ideal gate voltage or very low in amplitude. Interestingly, the number of phases where the gate voltage is the right phase and has a high enough amplitude is the same in the high power and the low power measurements, while the number of measurement points with a high efficiency differ significantly. This means that the losses due to the non-ideal gate voltage are less significant for high input powers; the proper termination of the gate of the transistor can push the efficiency above 90%.

As the RF input power is constant, the DC output voltage in Fig. 6.2.4 behaves similarly to the efficiency curves but shows an interesting behaviour around $\angle \Gamma_G = 25^\circ$ in the low power measurements and around $\angle \Gamma_G = 15^\circ$ in the case of the high power measurements. Here, the DC output voltage is negative showing the significant impact the gate voltage has on the efficiency and DC output voltage. To examine this further, two cases at low power, the phases of $23^\circ$ and $29^\circ$, are compared in Fig. 6.2.6.

Figure 6.2.5: Gate voltages at different phases of the gate reflection coefficient $\Gamma_G$ at (a) -29 dB IBO and (b) -4 dB IBO, $|\Gamma_G|$ is kept between 0.98 and 1, blue line shows single drain current waveforms for phase reference.
as they result in a DC voltages of the same magnitude but with opposing sign. Fig. 6.2.6a shows that at these gate impedances, the load-lines are almost exactly point symmetric, which means that the transistor rectifies the RF signal equally in both cases. The gate voltage is around 180° out-of-phase, despite there being only 6° difference in the phase of the gate reflection coefficient which means that the resonance between the input capacitance of the transistor and the gate impedance, which is purely inductive, has to be somewhere between these two values.

Another interesting point is that of 180° where the voltage is presented with a short circuit and is thus ideally equal to the gate bias voltage of -3.4 V at all times, the case discussed and simulated in Fig. 6.2.2. The measurement agrees well with the simulation, as shown in Fig. 6.2.7. As expected, the load-line follows the -3.4 V trace for both low and high power levels. The figure also shows the magnitude of the higher harmonics in the gate signal, reaching a peak-to-peak voltage of 1.2 V.
6.3 Class F RF Rectifiers

As discussed in section 2.4.5, class F amplifiers are able to achieve high efficiencies by reducing the voltage/current waveform overlap using targeted harmonic terminations. Class F rectifiers do the same with the difference that the conduction occurs in the third quadrant instead of the first. The common design approach for high efficiency rectifiers was discussed in section 3.2.4, it is based on time reversal [355] or the fact that in GaN HEMTs, the DCIV curves in the first and third quadrant behave in a very similar way. In this approach, a PA is designed using the harmonic terminations as desired, and operated inversely, i.e. using the PA output as the rectifier input and the PA DC feed as the rectifier DC output. This approach has been shown to work well [356].

In this section, the class F rectifier will be designed using a different approach, directly synthesising the desired impedance environment using the active load-pull system presented in section 6.1. While time reversal works reasonably well, the measurement of a complete PA/rectifier prevents observation of the waveforms at the intrinsic current generator plane. Additionally, the assumptions of symmetric DCIV curves do not hold true as previously shown in the DCIV measurements, see Fig. 6.2.1, where the current at a gate bias voltage of 0 V in the third quadrant is significantly higher than that in the first quadrant. Another reason is the fact that the voltage in a rectifier reaches negative voltages increases the maximum voltage swing of the device, lastly, the non-linearity of the output capacitances, especially around 0 V [323], mean that just reversing a PA might not yield the optimum performance.

Measurements

The measurements are conducted using the system presented in section 6.1 using the same device that was used in the previous measurements, the Qorvo 6x80 µm 0.25 µm GaN HEMT die at 900 MHz. Just as in the previous measurements, the transistor is injected with an RF input signal at the fundamental on the drain and the signals necessary to present the desired intrinsic impedances at the fundamental at the gate and at the second and third harmonic on the drain side, leaving all higher harmonics unterminated. In section 6.2, the optimum gate impedance for high efficiency operation was established to be \( \Gamma_G = 1\angle 47^\circ \) or \( Z_G = j115 \Omega \), the following measurements will use this impedance. The gate bias voltage is kept at the same as in the previous measurement, -3.4 V. Unlike in the previous measurements, the impact of the DC load on the rectifier behaviour will be discussed. In addition to the previously mentioned 328.55 Ω resistor, the rectifier will be measured with DC load resistances of \( R_{DC} = 56.4 \Omega , 182.9 \Omega \), and 458.5 Ω which is close to the optimum impedance calculated in the previous section. While the measurement system is able to synthesize arbitrary RF impedances, only passive, constant impedances can be realised in circuits, therefore, the impedances are kept as
constant as possible within the constraints of the system. As the signal generators are not coherent, the phase between them slowly drifts with time, making exact impedances hard to control, especially if the individual magnitudes and phases are controlled manually.

The measured a and b waves are converted into impedances, as shown Fig. 6.3.1. The figures show that all impedances are passive and close to constant at all controlled frequencies on both input and output. An interesting effect can be observed in the measurement of the input impedance $Z_{in,f0}$, the impedance the drain presents to the measurement system. The impedances start close to an open circuit at very high impedances, a result of the signal being too small to be rectified. As the signal level increases, the impedance decreases as the gate voltage increases and the input signal starts to be rectified, Fig. 6.3.2 shows the load-lines for the operation at these low power levels which show very low currents at low input power levels, and increase with the input power. The figure also shows how the whole load-lines move towards higher voltages with increasing power, a result of the increasing gate voltage when the transistor is in the third quadrant.

Although the values of the RF input impedance is related to the DC load resistances, there seems to be no direct connection, as shown in Tab. 6.1. The RF input impedance is higher than the DC load resistance for $R_{DC} = 56.4 \, \Omega$ while it is lower for all other DC resistances. It is not only the RF input impedance that changes with the DC resistances, as $R_{DC}$ determines the current drawn at a fixed voltage, all parameters of the rectifier depend on the DC load resistance. To evaluate this dependency, the results of the RF input power sweep are plotted in Fig. 6.3.3. As the gate impedance is on the edge of the Smith chart, no power is injected at the gate side which means that the RF-to-DC efficiency
in this case is simply $\eta = \frac{DC_{out}}{RF_{In}}$. The measurement results show that, as expected, the harmonic terminations lead to a very high efficiencies for the optimum DC load impedance of 328.55 $\Omega$. For the higher load resistance, the efficiency increases slightly for lower powers but drops earlier, the efficiency stays above 90% for a power range of more than 10 dB and above 80% over a power range of over 20 dB for both 328.55 $\Omega$ and 458.5 $\Omega$ though. As these impedances are far from 50 $\Omega$, the measurement accuracy is reduced due to the limited directivity of the measurement system. Roberg et al discuss in [356] how the efficiency depends on the ratio of $r_{DS,on}$ to RF input impedance and DC load resistance, which is reflected in the measurements. For decreased DC load resistances, the efficiency drops as the voltage swing is small compared to the knee region, something particularly noticeable at $R_{DC}=56.4$ $\Omega$ where the efficiency fails to reach 70%, more than 20 percentage points below the efficiencies at higher DC load resistances. The reason for this, the relatively high on-resistance, is clearly visible in Fig. 6.3.4.

For most measurement points, the current closely follows the IV curves showing a near ideal behaviour of fully conducting for one half cycle and pinching off completely for the other half. However, the limited on-resistance means that a the ratio of knee voltage to RF voltage is around 20%, significantly limiting the achievable efficiency. One interesting measurement point is that with the highest input power. In Fig. 6.3.3, the efficiency and the DC output voltage collapse for high input powers. The load-line of that measurement point is shown as a thin purple dotted line in Fig. 6.3.4.
the reason behind this collapse, the drain and gate waveforms are plotted in Fig. 6.3.5. These two figures show that while the phase of the gate voltage is correct for the low input powers, as discussed in section 6.2, it significantly shifts in phase for the highest input power leading to current/voltage overlap in both parts of the cycle. There are two variables that may be responsible for this phase shift at a high input power, the input impedance control and the intrinsic capacitances. The realised input impedance is measured to be $8.9 - j*109\,\Omega$, or $0.94\angle49^\circ$, and thus in the targeted area, which means that the phase shift has to be an effect of the capacitances, the average value of which changes now that the applied voltages are deep in the third quadrant. One distinct result of the phase shift of the gate voltage is the current overshoot when turning the transistor off. This overshoot is also present at higher impedances, as shown in Fig. 6.0.1 in the appendix on page 264. This is not the only reason for the dropping efficiency at high input power levels, though. In Fig. 6.3.6, two more causes that lead to a drop in efficiency at high input power levels are discernible. The first is that the current increases to the point where it leaves the ohmic region, resulting in an increase of the on-resistance. This effect increases with the current drawn and thus appears significantly earlier at low load impedances, e.g. it is clearly visible in Fig. 6.3.4. A second efficiency-limiting effect only
starts to appear at high voltages and thus predominantly at high DC load impedances and appears to lead to a current flow once a certain voltage, far below the breakdown voltage, is exceeded. Unlike the overshoot due to the phase shift of the gate voltage, this behaviour can not be attributed to the gate voltage, as shown in Fig. 6.3.7, where both effects are displayed: the gate voltage shift has an overshoot at around 0.8 ns and the secondary, pinch-off, effect at around 1 ns. The figures show that while the gate voltage phase does not change, it is slightly skewed towards the right due to higher harmonics, leading to a overshoot at around 0.8 ns, where the gate voltage should be below -3 V but is above 2 V instead. The gate voltage does not explain the current peak at 1 ns however, it is well below the threshold voltage at that time. It is not an effect of drain voltage alone as the current peak only appears at the first voltage peak and there is no trace of it at the second voltage peak. While a measurement issue can not be excluded completely, the asymmetry of the effect indicates that it is not a problem of the deembedding. Assuming a correct measurement, this points to a trapping effect that is either typically not captured or usually presents itself in a different way. As GaN HEMTs in rectifiers are rarely analysed in detail, the source of this current peak remains unexplained for now and is an interesting question requiring further research, maybe in conjecture with the optimisation of GaN HEMTs for rectification applications.
6.4 Class B* RF DC/DC Converters

Usually, the PA and the rectifier in RF DC/DC converters are separate entities, with the only power flow and interaction happening at the fundamental frequency. The interaction in this case is limited to the rectifier’s input impedance changing and presenting a different impedance to the output of the amplifier as shown in the rectifier measurements in the previous section. In this section, a new RF DC/DC converter topology is presented which, unlike other topologies, treats PA and rectifier as one unit and achieves its high efficiency as a result of their interaction.

6.4.1 Concept

As discussed by Dani et al in [147], a high PA efficiency can be achieved by injecting a signal at the even harmonics into the drain of the PA, resulting in both drain voltage and drain current of the PA looking like a half-wave rectified sine wave. This power at the second harmonic usually needs to be generated externally and needs to be in the right phase at the intrinsic current generator plane of the PA. As the rectifier generates power at all even harmonics, a circuit which directly connects PA and rectifier at the fundamental and all even harmonics should result in all drain voltages and currents resemble half wave rectified sine waves, with both voltages being positive, a positive PA current and a negative rectifier current, as shown in Fig. 6.4.1. With all maximum amplitudes normalised to 1 and a negligible $r_{on}$, the waveforms of the PA in this case can be described as

$$ v(\theta) = \begin{cases} \sin(\theta) & \text{if } 0 \leq \theta < \pi \\ 0 & \text{if } \pi \leq \theta < 2\pi \end{cases} $$  

(6.3)
Figure 6.3.7: Measured rectifier waveforms for different input power levels at a DC load resistance of 458.5 Ω (a) drain voltage and current waveforms and (b) of gate waveforms

Figure 6.4.1: Proposed circuit with intrinsic waveforms of a class B* DC/DC converter, MN is a matching network

and

\[
  i(\theta) = \begin{cases} 
  0 & \text{if } 0 \leq \theta < \pi \\
  -\sin(\theta) & \text{if } \pi \leq \theta < 2\pi 
  \end{cases} 
\]  

(6.4)

where \( \theta = 2\pi f_0 t \) and \( f_0 \) is the fundamental frequency. A Fourier series extension of the voltage yields

\[
  v(\theta) = \frac{1}{\pi} + \frac{1}{2} \sin(\theta) - \frac{2}{\pi} \sum_{n=1}^{\infty} \cos(2n\theta) \frac{\sin(n\pi)}{4n^2 - 1} 
\]

(6.5)

The Fourier coefficients of voltage and current are identical for all higher harmonics, at the fundamental frequency, the coefficients are equal in magnitude but multiplied with a factor of -1. Knowledge of
these coefficients allows the calculation of power and efficiency. With voltage and current in phase, $P_{f0}$, the power at the fundamental, can be calculated as $P_{f0} = 0.5 \cdot i_{f0} \cdot v_{f0} = 0.125$, the DC power as $P_{DC} = \frac{1}{\pi^2} = 0.1013$ resulting in a PA efficiency of $\eta_{PA} = \frac{P_{f0}}{P_{DC}} = 123\%$. The fact that the efficiency is above 1 means that there has to be power flow at higher harmonics as expected, this power can be calculated as

$$P_{nf0} = \frac{1}{2} \cdot i_{nf0} \cdot v_{nf0} = -\frac{2}{\pi^2 (n^2 - 1)^2}$$

yielding the even harmonic powers $P_{2f0} = -0.0225$ and $P_{4f0} = -0.0009$, higher even harmonics are very low in power. The negative sign shows that this power needs to supplied to the PA to achieve that shape of the drain voltage, the PA is a load at these frequencies. As, apart from the inverted current, the waveforms of the rectifier are the same, the Fourier coefficients of the voltage are identical and that of the current are all multiplied by -1, which means that all power terms are equal in magnitude but multiplied by -1. The rectifier’s efficiency is the ratio of DC output power to RF input power, the rectifier efficiency is thus $\eta_{-1} = 81\%$, the rest of the power is up-converted to higher harmonics. As the power magnitudes in PA and rectifiers are the same, the rectifier can load the PA, while the rectifier loads the PA at all higher harmonics. Assuming a correct time alignment, both the PA and the rectifier will end up with the targeted half-wave rectified voltage and current waveforms, and, with it, the high efficiency. An advantage of these waveforms over class F or class E is that the slope around the voltage and current transitions is less steep, as demonstrated in Fig. 6.4.2 where the voltage and current waveforms are shown for different numbers of harmonics present in the waveforms. As a result of the smoother transition, a circuit with half-wave rectified voltage and current waveforms degrades less with phase misalignment.

The modulator has to provide a wide range of output voltages to the ET PA. The PA, however, is only efficient at its full voltage swing as discussed in section 2.4. In a modulator, this results in a reduced efficiency for DC powers lower than the maximum which is problematic as the main goal
of ET is the operation of the ET PA at reduced supply voltages and thus powers. Replacing the simple PA with a PA structure providing high efficiencies at OBO is one option, it would however increase complexity significantly. The efficiency can also be recovered if the load impedance of the PA is increased for low power levels, similarly to the main PA in a Doherty PA. Assuming the DC load and thus the input impedance of the rectifier stay constant, a reduced DC output power of $\psi \cdot P_{\text{DC}}$ means that both voltage and current are scaled by $\sqrt{\psi}$, where the $\psi$ can attain values between $\psi_{\text{min}}$ and 1. $\psi_{\text{min}}$ is determined by the desired dynamic range of the modulator, e.g. 0.25 for desired voltage range of 12 dB. To keep the efficiency of the modulator high over a wide range of DC voltages, the drain voltage swing of the PA has to be conserved which means that the current has to decrease with $\psi$. Assuming the drain voltage stays constant, the current can be expressed as

$$i(\theta, \psi) = \psi \left( \frac{1}{\pi} + \frac{1}{2} \sin(\theta) - \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{\cos(2n\theta)}{4n^2 - 1} \right)$$

(6.7)

which directly leads to a new set of harmonic load impedances which are all scaled by a factor of $1/\psi$. This means that the matching network between PA and rectifier needs to transforms $Z_{\text{BB}}$, the impedance the rectifier presents to the PA, to $Z_n(\psi)$, the impedance at the transistor. To keep the matching network realisable, this change in transformation needs to happen over frequency which means that the output voltage of the modulator is varied by varying not the input power of the PA but the operating frequency of the modulator, similar to the approach in [295]. Designing this matching network is going to be a challenge, as the transformation has to be correct not only at the fundamental and at least the second harmonic, the phase has to be correct as well. Additionally, the matching network needs to incorporate the output capacitance of the PA and, if possible, the bias network of the PA and the output filter of the rectifier.

### 6.4.2 Simulations

The concept is evaluated in simulations in Keysight ADS using a Qorvo 250 nm GaN-on-SiC HEMT model which has been designed to be accurate for operation in both the first and the third quadrant as shown in Fig. 3.2.5. The transistor is submitted to a load-pull simulation to establish an impedance that offers a good trade-off between efficiency and output power. Both PA and rectifier are biased in class AB, close to class B to allow it to operate at low input powers. To be able to easily verify the simulations, they are conducted at 900 MHz where the measurement equipment for multi-harmonic load-pull is available. As with the class F rectifier, the optimum DC load resistance can be calculated using the maximum voltage and current swing, yielding an impedance of $R_{\text{DC,opt}} = \frac{V_{\text{PP}}/\pi}{I_{\text{PP}}/\pi} = 340 \, \Omega$. As the waveforms have a different shape compared to the class F rectifier, the optimum impedance changes significantly.
To keep the simulation setup as simple as possible, both PA and rectifier have their own, independent RF input power source, as shown in Fig. 6.4.3 which means that unlike a self-synchronous rectifier, the PAE calculation now needs to take the input power into consideration. The circuit comprises two transistor, one for the PA and one for the rectifier, and a transmission line with an electric length of $\lambda/4$ at the fundamental frequency. The transmission line is crucial as it ensure that at the second harmonic, voltage and current are $180^\circ$ out-of-phase at the rectifier and in-phase at the PA, the conditions necessary for the two parts of the circuit to load each other. The load impedance that needs to be presented to the PA was established in a load-pull simulation, the transmission line can, in addition to providing the appropriate phase relations, act as an impedance transformer, transforming the input impedance of the rectifier, $Z_{in,RF}$, to the optimum impedance of the PA. As the intrinsic capacitances of the transistor are not negligible at the operating frequency of 900 MHz, the phase-shift introduced by them has to be compensated. Here, the compensation is achieved by shortening the transmission line from $90^\circ$ to $76^\circ$. While this leads to a well aligned phase at the fundamental frequency, the alignment starts to deteriorate at higher frequencies as the capacitances have an increasing impact at higher harmonics, resulting in the current waveform becoming asymmetric, as shown in Fig. 6.4.4. To be able to observe the waveforms at the intrinsic current generator plane, the output capacitance $C_{DS}$ is deembedded using a linear capacitance model. The figure also shows
that, as expected, all waveforms are the shape of a half-wave rectified sine wave with little overlap of the voltage and current waveforms. This is reflected in the high DC/DC efficiency of 83.8% at the maximum output power, as shown in Fig. 6.4.5. The PA efficiency exceeds 100% as predicted while the rectifier efficiency exceeds the theoretical limit of 81% indicating that while the general waveforms look as predicted, the power flow is slightly different. Figure 6.4.5 also shows how the PAE of the modulator and its parts change with drive power. The PA efficiency drops in OBO as expected, while the rectifier is efficient over a wide power range which is consistent with the measurements in section 6.3. Interestingly, the rectifier efficiency does not deteriorate for high power levels, indicating that the effect that limited the achievable efficiency in section 6.3 is not present in the model. The waveforms in Fig. 6.4.4 clearly demonstrate the main source of loss. While the voltage magnitude of the rectifier is around 1 - 2 V when the transistor is conducting, the voltage magnitude of the PA during conduction is significantly higher at around 3 - 4 V, which means that the PA is the main source of losses in the modulator. While the individual efficiencies are not immediately clear from the graph in Fig. 6.4.5, they becomes more clear once the power flow between PA and rectifier is analysed in detail, as shown in Fig. 6.4.6. In the figure, positive values show power flowing into the rectifier, negative values for power flowing into the PA, confirming the theoretical analysis in terms
of behaviour at the fundamental and the even harmonics. Interestingly, a $3f_0$ power starts to appear once the PA is over-driven. This explains that the efficiencies starts to deviate from the theoretical values for high power levels.

The PA is only efficient for high input power levels for the reasons discussed in section 2.4. This means that a simple DCDC converter as presented here would be unable to achieve a high efficiency over a wide power range if the output power is to be adjusted using an amplitude modulation. To evaluate the approach that varies $Z_{n}(\psi)$ as proposed in the last section, the simulation is now repeated for a number of characteristic impedances of the transmission line between PA and rectifier. This variation in characteristic impedance means that the transmission line operates as a transformer, resulting in different load impedances being presented to the PA. In the case of this simulation, the load impedance presented to the PA changes from 234 $\Omega$ to 778 $\Omega$, corresponding to $\psi = 0.3$ or an output voltage range of 0.55. As discussed, the reduced length of the transmission line leads to the phases being not perfectly aligned, this impacts performance as the load impedance increases, as shown in Fig. 6.4.7. Despite this, the efficiency stays high over a wide output voltage and thus output power range as demonstrated in Fig. 6.4.7. While this way of varying the impedance is easily done in simulation, it can not be realised and would need to be replaced by a more complex matching network in a realisation. The simulation still shows, however, that the approach of using an matching network to provide the PA with a changing load impedance can lead to the expected high efficiencies, even if the load phases at the harmonics are less than ideal.

6.4.3 Rectifier

While PAs with second harmonic injection have been demonstrated to work as expected [147], the harmonic generation of rectifiers has not been verified so far. This section will rectify that, looking at the harmonic generation of a rectifier under different load impedances. The measurements are conducted using the active harmonic source-pull/load-pull setup presented in section 6.1 and the same DC load resistances used in the previous measurements. As in the previous measurements, the drain
of the transistor is excited with a signal containing components at the fundamental, second and third harmonic allowing synthesis of arbitrary impedances at these frequencies. The gate is injected only at the fundamental, aiming to present a highly reflective load optimised to achieve highly efficient self-synchronous operation. Again, the fundamental is chosen as 0.9 GHz, unlike with the class F rectifier, the harmonics are not terminated highly reflective though. The second and third harmonic are terminated with a resistance equal to that of the DC load impedance, the higher harmonics are terminated with $50\,\Omega$. Ideally, the fourth harmonic would be presented with the same impedance but the measurement setup did not allow that. Following the previously discussed deembedding approach, all capacitances are removed to be able to evaluate the voltage and current waveforms at the current generator plane. Fig. 6.4.8 shows a comparison of simulated and measured drain waveforms for a DC load resistance of $328.5\,\Omega$, $\Gamma_{in} = 1\angle50^\circ$ and input powers between -14 dBm and 35.6 dBm.

Figure 6.4.8: Simulated (dotted) and measured (solid) drain waveforms of a 6x80 µm GaN HEMT with a DC load of 328.5 Ω, $\Gamma_{in} = 1\angle50^\circ$ and input powers between -14 dBm and 35.6 dBm.
load is increased, the efficiency follows a very similar trajectory but starts to drop at lower power as at the higher impedance, the voltage reaches higher values resulting in the voltage peaks appearing at lower power levels. If the DC load is reduced, the ratio of on-resistance and maximum voltage drops, limiting the efficiency as discussed in more detail in [356]. Additionally, the voltage does not reach the values required to cause the voltage peak until the power level at which the drain current leaves the ohmic region and the rectifier would become very efficient anyway. The error bars show the errors in the measurement system, they are largely due to the limited directivity of the couplers. As before, the phase of the gate signal shifts at high power levels at the low impedances, further reducing the efficiencies, as shown in Fig. 6.4.10. Fig. 6.4.10a also shows that the unexplained current peak at 1.1 ns can not be a typical breakdown effect as the peak starts occurring between 40 V and 50 V at 182.9 Ω, while at 328.5 Ω, the current peak is negligible when the drain voltage reaches 50 V is only starts to appear at around 60 V, as shown in Fig. 6.4.8.

Even harmonic generation

So far, the rectifier was evaluated purely at the fundamental and at DC, which, while useful to analyse possible limitations of the rectifier, was not the focus of the measurement. In class B* operation, the rectifier has to be able to supply to the PA with power at the even harmonics to shape the drain voltage of the PA. Fig. 6.4.11 shows the generated harmonic powers normalised to the power at the fundamental. The generated harmonic powers are very small for low input power levels as at those levels, little rectification happens as shown in Fig. 6.3.2. As the input power increases, the generated harmonic powers increase until they reach a plateau on which they stay for a power range of about 25 dB. Compared to the theoretical values of 0.22 for the second harmonic and 0.0088 for the fourth harmonic, the measured ratio of generated powers is lower with around 0.14 for the second harmonic and 0.004 for the fourth harmonic, about half of the theoretical value. They are still above the levels required to achieve significant improvement of the PA performance as reported in [147]. The simple
Figure 6.4.10: Measured drain voltage and current waveforms of a class B* rectifier for different input power levels at a DC load resistance of (a) 182.9 $\Omega$ and (b) 56.4 $\Omega$

Efficiency calculation $\eta_{\text{Rectifier}} = \frac{P_{\text{DC}}}{P_{\text{RF},f_0}}$ only takes the power at the fundamental into account as usually, power at other frequencies is undesired. With the power at the even harmonics being used in the PA, a more comprehensive efficiency definition is required, the definition under these circumstances can be found in (6.8).

$$\eta_{\text{Rectifier, class B*}} = \frac{\sum_{i=0}^{\infty} P_{2i}f_0}{\sum_{i=0}^{\infty} P_{(1+2i)f_0}} \approx \frac{P_{\text{DC}} + P_{\text{RF},2f_0} + P_{\text{RF},4f_0}}{P_{\text{RF},f_0}} \quad (6.8)$$

It allows the calculation of the overall efficiency of the class B* rectifier. As higher even harmonics are significantly lower in power, it can be simplified to only include harmonics up to the fourth. The simplified equation also does not take any power flow at odd harmonics higher than the first into account as they are expected to be very low. Fig. 6.4.12 shows that using this extended efficiency definition which takes the harmonic powers into account, the efficiency of the rectifier is comparable to that of the class F rectifier, reaching up to 95% and staying above 90% for a power range of more than 20 dB, again dropping for high power levels due to the issues discussed earlier.
6.4.4 Realisation

To prove the concept, a demonstrator is designed using two Wolfspeed CGH40010F 10 W GaN HEMT devices. For simplicity’s sake, both PA and rectifier are driven independently. This requires two coherent signal generators, the only one available at the time of the experiment was a Keysight 81150A, which is able to generate coherent signals up to 120 MHz, so this was chosen as the design frequency. An advantage of the relatively low frequency is that parasitic series and shunt elements between the current generator plane and the package have little effect. This means that the matching network in this case can be realised using a single $\lambda/4$ transmission line, shifting the phase of the second harmonic by $180^\circ$. The optimum impedance of the transistor is around $33 \Omega$, therefore this is the characteristic impedance chosen for the transmission line as discussed earlier in this section. Fig. 6.4.13b shows the circuit diagram nd a picture of the output of the PA-rectifier PCB that had been used in the measurements. It shows the transmission line used to match PA and rectifier, the the lumped element bias networks and provisions for integrated measurements that were never implemented due to a lack of time.
The PA was biased in class AB at -2.8 V, the rectifier in class B at -3.2 V. The input power of the converter was swept for multiple supply voltages and DC load impedances, the input powers of PA and rectifier were kept the same with a constant phase shift of 290° taking into account the cable length and the desired phase shift. The two input signals are generated in the Keysight 81150A signal generator, amplified and fed into the inputs, the output voltage is measured using a digital multimeter at the adjustable power load. While the conversion efficiency reaches up to 78% even without any optimisation, as shown in Fig. 6.4.14(a), the measurements show that, as the output voltage increases, the achievable efficiency reduces significantly. Theory and previous measurements predict an increased output voltage at higher load resistances, this is confirmed in these measurements as well. Interestingly, output voltage and efficiency are almost linear. As \( \eta = \frac{P_{out}}{P_{in}} = \frac{V_{out} \cdot I_{out}}{V_{in} \cdot I_{in}} \) and \( \frac{\eta}{V_{out}} = const, \frac{I_{out}}{V_{in} \cdot I_{in}} = const \). With \( V_{in} \) being constant, that means that the ratio of the input and the output current is constant, as long as the PA is not compressed significantly. As the converter consists of two interacting parts and the only available data are DC measurements, further conclusions such as identifying the limitations and the sources of inefficiencies are therefore impossible. The PA did reach saturation in almost at all combinations of input power and load resistor, as shown in Fig. 6.4.14(b). Extrapolating the non-compressed trajectory would place the compressed efficiency...
Figure 6.4.14: Measured DC to DC efficiency of the class B* DC/DC converter at 120 MHz for different input voltages and different load resistances (a) over output voltage and (b) over input power at 28 V and 66 Ω around 60 % with an output voltage of about 32 V, still significantly below the maximum of 78% that was achieved at lower voltages. While the circuit only showed high efficiency at few points, these measurements fulfilled their role as a proof of concept, showing that efficiencies of almost 80 % are achievable in general. To achieve more competitive performance, the circuit would need to be revised and, ideally, designed around a transistor with a model that works in the third quadrant. Additionally, the efficiency worsened as the frequency was reduced. As the signal generator was operating at its maximum frequency, there was no way to establish the optimum frequency. Measuring internal voltages and currents would allow debugging the circuit to track down the origin of any unexpected effects.

6.5 Conclusions

This chapter dealt with RF DC/DC converters, focussing mainly on the rectifier side of them. One major element of the analysis was the role of the gate voltage and ways to generate it. By using the intrinsic capacitances as a feedback network, the voltage at the gate of the rectifier can be generated
from the drain voltage with the gate termination determining the amplitude and phase of the gate signal. This was shown using measurements on a 0.25 µm Qorvo GaN HEMT, varying the gate impedance systematically established for the first time the dependence of the efficiency on the gate impedance. It also showed that the optimum phase depends significantly on the power level, and that the phase dependence drops with increasing input power. Measurements of harmonically terminated rectifiers have shown that this input power dependence has a significant impact in real rectifiers as a phase shift in the gate voltage can significantly reduce the efficiency by turning the transistor on at the wrong time. Based on this analysis, an active harmonic load-pull/source-pull is used to supply the transistor with a class F harmonic impedance environment. For the first time, waveforms and powers are measured at the device plane to establish the factors limiting the efficiency. The measurements showed that an efficiency of more than 90% could be maintained over a wide power and load resistor range. It also revealed the major loss mechanisms in self-synchronous rectifiers: the gate signal, the waveforms leading the ohmic region, and a previously unknown effect that lead to current flow in the off-state. This effect was also present in the second rectifier that was measured. This rectifier is part of a new type of RF DC/DC converter that uses the even harmonics generated in the rectifier to conduct harmonic injection at the drain port of the PA. Simulations demonstrate it working efficiently, the efficiency limiting effects present in the measurement are not all present in the simulation however. Load-pull/source-pull measurement of the rectifier demonstrate that the harmonic generation works as expected, complementing the reports in literature about the feasibility of second harmonic injection into PAs. A proof-of-concept of the combined harmonic generation/injection modulator was designed around two Wolfspeed CGH40010F 10 W GaN HEMT devices, built and measured and while it showed some promising results with efficiencies up to 78%, it also showed a lot of room for improvement.

References


RF Supply Modulators - Conclusions


Chapter 7

Integrating PA and Supply Modulator

Contents

7.1 Baseband Impedance ................................................. 200
  7.1.1 Impact of Baseband Impedance on the PA ................. 201
  7.1.2 Characterising the Baseband Impedance ..................... 202
7.2 Power Amplifier Impedance ......................................... 210
7.3 Supply Modulator Ripple ........................................... 214
  7.3.1 Impact of Ripple on PA Linearity ............................. 215
  7.3.2 Mitigating the Effect of Ripple ............................... 216
7.4 Conclusions ......................................................... 219
References ............................................................... 220

Parts of the analysis, measurements and conclusions in this chapter have been published in [22], [419], [420].

So far, the power amplifier and the supply modulator have been discussed in isolation. As shown in section 3.1, the two parts of the system interact when they are being integrated. This chapter is about characterising the interfaces between PA and modulator as a step towards being able to describe their interactions using the PA designed for the measurements in chapter 5 and a fully functioning supply modulator which has been provided by Manchester [338]. It starts by looking at the output impedance of a modulator, the baseband impedance, in section 7.1, and of the PA in section 7.2. Following that, the impact of the supply modulator ripple is discussed in section 7.3.

7.1 Baseband Impedance

As the baseband impedance plays an important factor in determining the PA performance, as discussed in section 3.1.6, the impact the baseband impedance has on the PA performance will be discussed in more detail in section 7.1.1. For a factor that crucial in determining how well a modulator is going to work as part of an ET system, characterisation is rarely conducted or, at least reported, possibly because of the perceived difficulty of the measurement. A new, simplified approach to characterising the baseband impedance will be discussed in section 7.1.2.
7.1.1 Impact of Baseband Impedance on the PA

As discussed in section 3.1.6, the baseband impedance of the modulator directly results in memory effects which can be observed in modulated measurements as an asymmetry [94] in the intermodulation side-bands and a spreading in AM/AM and AM/PM measurements. The in-band distortion is demonstrated first by comparing measurements of a PA that is excited with a 20 MHz LTE signal for two types of baseband impedances. One is a conventional bias circuit, the other is a supply modulator with an output filter, presenting an impedance significantly above zero, the exact value of the impedance will be characterised in section 7.1.2. The measurement is conducted using the 2.9 GHz PA from chapter 5, leading to the AM/AM and AM/PM measurement results in Fig. 7.1.1. While the PA was driven into different levels of compression for the two measurements, they allow a qualitative comparison of the added distortion. The figures clearly show how the memory effect increases the AM/AM and the AM/PM distortion by spreading the gain and the phase over a significantly larger range. While the spreading of the amplitude and phase distortion points at low powers can be attributed to misalignment the output and the reference waveforms [421], the gain and phase in this case spread at all power levels, indicating a real memory effect. The gain varies by around 2 dB over the whole power range, the phase by 5° to 10°, requiring a significant amount of pre-distortion.

The out-of-band distortion can be shown more clearly using discrete tones that allow direct comparison of tones, the PA from the previous measurement is excited with a five-tone signal consisting of 5 tones of equal amplitude and phase, spaced 1 MHz apart, and connected to the buck-converter from the previous measurement, resulting in the output spectrum in Fig. 7.1.2. The figure demonstrates that the spectrum shows significant asymmetries, not all IMD products are asymmetric however. To evaluate this further, this is modelled using an ideal 5-tone signal and a baseband voltage based on the one measured in the measurement that resulted in Fig. 7.1.2, as shown in Fig. 7.1.3a. The ideal 5-tone
signal is distorted using the power series in section 2.6.1, adding third and fifth order distortion. The spectrum of this distorted signal is absolutely symmetrical, as shown by the black trace in Fig. 7.1.3b. When this signal is multiplied with the baseband voltage in the time domain to represent the mixing occurring in the transistor, the resulting output spectrum shown in light blue in Fig. 7.1.3b becomes asymmetrical. As in the 5-tone measurement results in Fig. 7.1.2, not all frequency components are asymmetrical, however. Due to the simplicity of this model, the increased in-band distortion due to the baseband impedance is not present in the output spectrum. An interesting observation is that the mixing process results in significantly increased fifth-order IMD products, which again is consistent with Fig. 7.1.2. This spreading of the spectrum needs to be pre-distorted as requirements for spectral purity increase with distance from the carrier. Both the in-band and the out-of-band distortions demonstrate that the baseband impedance plays an important role in determining the achievable linearity in supply modulated systems.

7.1.2 Characterising the Baseband Impedance

Seeing the significance of the baseband impedance in supply modulated systems, characterising it is relevant at a system level as it informs the designers of the distortion the modulator will introduce into the system. This knowledge can then inform the judgement of how well the system is going to work and whether corrective action such as redesign of the modulator or additional pre-distortion needs to be taken. In section 3.1.6, the use of complex, bespoke modulated signals as a way of characterise the baseband impedance of modulators was discussed. In this section, a simpler way using multi-tone signals and modulated signals preconfigured in many measurement systems will be presented. This approach uses Ohm’s law and exploits the fact that any current drawn will induce a voltage over the impedance to be probed, in this case the output impedance of the modulator. In a first step, the DUT is set to supply a constant voltage to a load, see 1 in Fig. 7.1.4. This load is designed to draw a
current that changes with time which will result in a time-varying current flow into the load, see 2 in Fig. 7.1.4. This time-varying current will flow out of the DUT into the load, see 3 in Fig. 7.1.4. As all the current components apart from the DC one are generated by the load, measuring the current flow and the voltage between DUT and load will allow calculation of the baseband impedance

\[ Z(f) = \frac{V(f)}{-I(f)} \]  \hspace{1cm} (7.1)

For an ideal voltage source, the baseband impedance would be a perfect short circuit resulting in no voltage drop. Bias networks in PAs with no supply modulation typically try to mimic this and try to lower the baseband impedance of the bias network [247], supply modulated PAs need to rely on the modulator to provide this baseband short circuit.
Implementation

A simple way to realise this time varying load is by using a power amplifier. As the current that is being drawn by the PA depends on the drive signal, it is very easy to control. What needs to be considered though is that the PA can not have its own baseband termination in the frequency range that is to be covered by the measurement, neither as part of the bias network nor as part of the matching network. Therefore, the operating frequency of the PA needs to be significantly higher than the measurement range, there is no additional frequency limitation however. The schematic of the measurement setup can be seen in Fig. 7.1.5, the modulator and the PA are modelled as RC circuits, the output filter of the modulator as an LC circuit. As all the current apart from the DC current is generated in the transistor of the PA, the elements to the right of the voltage probe do not have any impact on the ratio between measured voltage and current. The resistances and capacitances of the PA do however conduct some of the generated current; the higher this portion is, the less current reaches the modulator which reduces the generated voltage and, with it, the measurement sensitivity.

Another limitation of the measurement is the frequency behaviour of the voltage and current probes and the resolution of the oscilloscope they are connected to.

The way the load changes with time determines the frequency range over which the baseband impedance can be characterised. A conceptually simple way to generate the time varying current is using a two-tone signal. If the bias point of the PA is chosen so that some of the input signal is below the threshold voltage, the two-tone input signal will lead to a current with a significant amount of harmonics, see the example using an ideal class B PA in Fig. 7.1.6. All of these harmonic components in the current waveform would result in a voltage drop allowing the calculation of the impedance at that frequency. To cover a wider frequency range or achieve a higher resolution, the frequency spacing between the two tones can be changed. One example would be a chirp, starting with a low frequency spacing and increasing it until it covers the whole frequency range of interest. Another way of increasing the frequency range is increasing the number of tones in the modulated signals, e.g. to a five-tone signal. Replacing the multi-tone modulation with a communication signal with a continuous
signal allows characterisation of the whole frequency range at one go, it does however require a system capable of generating such a signal. Spreading the signal power over more frequencies will, at a constant sum of output power, reduce the power level of the individual frequency components, making the resulting measurement more noisy, an observation consistent with [288] where the signal is bespoke complex multi-tone modulation. Instead of changing the modulation to cover a wider frequency range, the PA can also be driven into compression. In this case, the transistor introduces a whole range of intermodulation distortion products that are higher in frequency than the original signal. The current drawn at these frequencies can then be used to extend the frequency range.

**Realisation**

The method realised and tested on a supply modulator for ET applications. The modulator serving as DUT is a buck-converter developed at Manchester University with a switching frequency of $f_{sw} = 75$ MHz and an output filter which is designed to accommodate modulation frequencies of up to 2 MHz [338]. The NI PXI chassis houses the FPGA card used to control the buck-converter, the NI PXIe-5164 14 bit oscilloscope used to measure the voltage and current, and the VST used to generate the input signal of the PA, see Fig. 7.1.7. The probes used in this measurement are a Agilent 10074C voltage probe with a bandwidth of 150 MHz and a Tektronix TCP312A current probe with a bandwidth of 100 MHz. As the two probes have different delays, the delay is characterised to be 9.83 ns using a square wave generator that is terminated with a 50 Ω broadband load. The PA used to generate the time-varying current is the 2.9 GHz PA used for the measurements in section 5.3, biased in class AB with quiescent currents of 90 mA and 200 mA, depending on the measurement. The modulations used for this measurements were a 1 MHz two-tone signal, a five-tone signal consisting of 5 evenly spaced tones over a bandwidth of 4 MHz, a 20 MHz LTE signal, and a four carrier WCDMA signal with an overall bandwidth of 40 MHz. An additional variable was the output voltage of the modulator which was varied between 10 V and 30 V for different measurements. Two-tone signals are one of the
Integrating PA and Supply Modulator - Baseband Impedance

Figure 7.1.7: Measurement setup: (a) Schematic and (b) Picture

Figure 7.1.8: Measured voltage and current between PA and modulator with a supply voltage of 16 V and a two tone excitation (a) in the time domain and (b) in the frequency domain

The simplest type of modulation which is useful in understanding how things work, for that reason they are the first signal being used to characterise the baseband impedance. The measurement results in Fig. 7.1.8a show the measured voltage and current between PA and modulator, for this figure the ripple was filtered using a digital low pass filter. As expected, the constant 16 V output voltage of the modulator is not constant any more but has a significant amount of additional frequency components.

An interesting observation is that voltage and current waveforms are quite different which means that the baseband impedance has to change significantly with frequency. Due to the being driven into saturation, the current waveform does not look like the ideal waveform in 7.1.6a. By transforming the waveforms to the frequency domain, Fig. 7.1.8b, the ratio of voltage to current can be seen to depend on the frequency. As expected, even with a modulation frequency of only 1 MHz, a significant amount of harmonic frequencies are generated. The measured current drops into the noise-floor at 11 MHz, the voltage stays above it up to around 17 MHz, resulting in a measurement bandwidth of 10 MHz, 10 times the modulation frequency. The measured voltage and current allow calculation
Integrating PA and Supply Modulator - Baseband Impedance

Figure 7.1.9: Measured magnitude and phase of the baseband impedance at 16 V using different types of modulation of the baseband impedance in the frequency domain, see Fig. 7.1.9, where the measurement results using two-tone modulation are shown as small, blue dots. Fig. 7.1.8b demonstrates that there is no impedance information contained between the tones, therefore the impedance can only be calculated at the modulation frequency and its harmonics. Any other frequency will be digitally nulled to keep the results clear as the ratio of two noise signals is going to be random and obfuscate the impedance. The same measurements and calculations are repeated with the five-tone signal, resulting in the large, orange dots, with the 20 MHz LTE signal, resulting in the thin green lines, and with the 40 MHz WCDMA signal, displayed as thin, red line. The baseband impedances obtained using the three different measurements show a good agreement up to 9 MHz, above which the two-tone results start to diverge and become erratic. This is expected as the voltage and current levels of the two-tone modulation start to get close to the noise floor. The five-tone results start to diverge from those obtained using the LTE modulation at around 25 MHz. The exact frequency is hard to pin down as even as the phase results start to spread while the magnitude results are quite consistent apart from a small number of outlying frequency points. While the results for the multi-tone signals are only at discrete steps determined by the modulation, the LTE modulation occupies a continuous spectrum which yields the baseband impedance at all frequency up to around 35 MHz where the phase starts to spread. The WCDMA results and the LTE results are consistent to the point where they are indistinguishable. While the magnitude shows little spreading, the phase spreads considerably more, a the result of the low impedance at these frequencies, the small voltage and current magnitudes making it hard to accurately measure the phase between them. The baseband impedance itself increases
with frequency, becoming increasingly inductive until it reaches the resonance of the output filter at 12.5 MHz after which it becomes capacitive and decreases as the impedance presented by the shunt capacitor of the filter reduces. The baseband impedance at DC is negative as at DC, the current is flowing from the modulator to the PA, there is no current at DC generated in the PA. To verify the measurement, the results obtained using the measurement are compared with an equivalent circuit model of the same modulator developed at Manchester University. The model uses the manufacturer model of the output filter components and a bespoke model that models the transistor as resistor with a value depending on supply voltage and output current. Figure 7.1.10 compares the baseband impedance of measurements and model, showing a very good agreement all the way down to 1 MHz. As the model is based on manufacturer supplied s-parameters with a lowest frequency point of 1 MHz, the values below it have to be extrapolated resulting in discontinuities, shown as dotted line in the figure. A buck-converter at a static output voltage and current can be modelled relatively easily as shown above, which makes it a good DUT for a case study such as this as this means the measurement can be verified. Apart from the discussed frequency limitations, all types of modulation were suitable to characterise the baseband impedance, demonstrating that simple modulated signals can be used to characterise the baseband impedance of a supply modulator using in a one-shot measurement.

Baseband impedances in different modulator topologies

With the measurement verified, the buck-converter can be investigated further. Additionally, it will be compared to a multi-level buck converter. If the modulator is used in an ET system, it will be required to deliver a wide range of output powers. It is therefore of interest to evaluate how the...
Integrating PA and Supply Modulator - Baseband Impedance

Figure 7.1.11: Measured magnitude and phase of the baseband impedance of a buck-converter using a WCDMA signal (a) for different power levels at a supply voltage of 20 V and (b) different supply voltages at 200 mA and in saturation

Baseband impedance changes with the current drawn by it and with the supply voltage. To measure the dependence of the baseband impedance on the current drawn, three measurements are conducted using the WCDMA modulation, one with a quiescent current of $I_{DQ}=200$ mA and an input power that drives the PA into compression, one with the same bias point and the RF input power reduced by 7 dB and one with the reduced input power and the quiescent current reduced to 90 mA. The measurement results in Fig. 7.1.11a show that the baseband impedance does not change significantly with the current drawn by the PA. The main difference is that with reduced input power and reduced drive signal, the frequency at which the measurement becomes very noisy is reduced. In addition to the current drawn, the other interesting parameter is the output voltage. As shown in Fig. 7.1.11b, the baseband impedance changes significantly with the supply voltage. As the filter characteristic stays the same, this change in baseband impedance can be attributed to the internal resistance of the modulator which loads the LC circuit of the filter and changes the Q of the circuit. Therefore, the most interesting part to compare is the real part of the impedance, especially at low frequencies, before the filter dominates, i.e. in the pass band of the filter. Figure Fig. 7.1.12a shows that the measurement is quite noisy due to the low power level at a part of the individual frequencies in the modulated signal. By removing the frequency components where voltage, current or both are below a certain threshold, in this case 2 mV and 300 µA, the noise in the measurement can be reduced, as shown in Fig. 7.1.12b. The figure now clearly shows that the modulator’s internal resistance depends on its output supply voltage. While the internal resistance is relatively constant from 10 V to 20 V, it increases significantly at 25 V before dropping to close to zero at 30 V. The same measurement is conducted on a different modulator, the multi-level buck-converter from [338]. It is characterised when
Figure 7.1.12: Measured real part of the baseband impedance using a WCDMA signal and different supply voltages (a) as measured of the buck-converter, (b) of the buck-converter with filtered noise, (c) of the multi-level buck-converter from [338] operated at a switching speed of 75 MHz with filtered noise and (d) of the multi-level buck-converter from [338] operated at a switching speed of 100 MHz with filtered noise operating at two different switching frequencies, 75 MHz and 100 MHz and the noise level is reduced as discussed above. This supply modulator shows significantly less baseband impedance variation over supply voltage, apart from the measurements at 30 V, the impedance stays close to constant and does not show any dependence on the switching frequency.

7.2 Power Amplifier Impedance

As discussed in section 3.1.7, the DC part of the PA impedance depends on its efficiency, and the ratio of supply voltage squared to output power and thus the shaping function. As the modulator supplies the PA with a modulated voltage, there will also be current flowing at frequencies higher than DC. During the design of the output filter, the impedance the PA is usually assumed to present to the modulator is a fixed resistance [291], [331] that does not change with frequency. Replacing it with a parallel RC network consisting of a fixed resistance and a fixed capacitance to model this impedance during the design of the process will increase the tracking accuracy as it now more closely resembles the real behaviour of the PA [291]. To be able to optimise the output filter, the impedance of the PA over frequency has to be measured. Once it is characterised and modelled, modulator and output
In [288], the PA impedance is measured but the authors only report what is presumably the magnitude of the impedance which is insufficient for any modelling. In [291], comprehensive measurements are conducted on a 4 W class B PA based on a 150 µm GaN-on-SiC HEMT. By using a measurement approach similar to the one in section 7.1, the impedance of the PA is characterised. The measurement is conceptually the same, a time varying signals is generated, the resulting voltages and currents are measured to calculate the impedance. To be able to attribute the voltages and currents to the parts of the system, the PA is operated with a static gate bias voltage and a static drive power. The voltage is now modulated with a signal based on the shaped function intended for a five-tone modulation with additional modifications to provide a wide range of frequencies. As the transistor has a significant amount of gain variation, see Tab. 4.1, the magnitude of the voltage waveform needs to be kept small as a change in supply voltage will result in a change in gain. A change in gain will also change the current which will distort the measurement results. As the measurement depends on all of the current bar the DC current being the result of the voltage drop over the PA, it is impossible to distinguish between the modulated current drawn by the PA and the current generated due to the voltage drop over the PA.
Figure 7.2.3: Measured real and imaginary part of the PA impedance at 20 V and different drive power levels, (a) at $I_{DQ}=90$ mA, (b) at $I_{DQ}=200$ mA

The measurement results in Fig. 7.2.2 shows that the modulator generates voltage components up to 100 MHz with decreasing magnitude. The switching frequency at 75 MHz is clearly visible; as the modulated voltages have a low magnitude to keep the measurement linear, the switching frequency is now in the same order of magnitude as the modulated voltages, despite the output filter. Interestingly, there are a few frequencies at which voltage and current magnitudes are significantly lower, e.g. 5 MHz and 35-40 MHz. The frequencies above 100 MHz are outside the specified frequency range of the current probe. This signal is applied to the PA in CW at multiple drive levels. As in the baseband impedance measurements with multi-tone modulations, this probing voltage only carries information at multiples of 1 MHz. Only those will be plotted, the rest will be digitally nulled for clarity. The measurement results in Fig. 7.2.3 show that using this technique, the impedance of the PA can be easily characterised. As expected, the measurement values become erratic at 100 MHz as the bandwidth of the current probe is exceeded. Figure 7.2.3a shows that at 90 mA quiescent current, the measured resistance values only vary with drive power for low frequencies, from 6 MHz onwards, the resistance is almost independent of drive power. From 70 MHz onwards, the resistance is close to zero. The reactance varies more with drive power, the reactances at the different power levels converge at around 40 MHz though. The impact of the drive power increases when the quiescent current is raised to 200 mA, as shown in Fig. 7.2.3b. The resistance increases as the drive power is increased until it reaches a certain level at which it starts to decrease. Interestingly, the general shape of the resistance curve depends on the drive power, e.g. the resistance at a drive power of 30 dBm is below that at 20 dBm and 10 dBm for low frequencies but decreases less slowly with frequency than the resistances at the lower drive power. The reactances behave in a similar fashion, i.e. they depend more on the drive power and shape in general shape. Importantly, for both quiescent currents, there is no obvious
connection between DC load resistance and PA impedance.

So far, the impedance was analysed on its own. Fig. 7.2.4 shows the equivalent circuit model of the measured impedance, the measured part of Fig. 7.2.1. As the equivalent circuit consists mainly of components in parallel, transforming the measured impedance into the admittance facilitates determining the component values. The series RC in Fig. 7.2.4 represents the RF load impedance of 50 Ω which is in series with the DC blocking capacitor of 10 pF. For simplicities’ sake, these two elements are neglected in the first step, the 10 pF series capacitor will result in a high impedances, especially at the low frequencies in this measurement, presenting a reactance of $X_C = 1591$ Ω at 10 MHz, at 60 MHz where the measurement results start to become noisy it is still 265 Ω. Interestingly, the measured resistances at $I_{DQ} = 90$ mA vary significantly even though the impedance traces didn’t move significantly, the rest of the measured data varies less. Fig. 7.2.5 allows the direct extraction of approximate values for the resistances and capacitances. With a quiescent current of 90 mA, R varies between 130 Ω and 170 Ω, while C varies between 70 pF and 80 pF, increasing the quiescent current to 200 mA decreases the resistance to 60 Ω-90 Ω, the capacitance varies significantly more between

---

**Figure 7.2.4: Equivalent circuit model of the PA at baseband frequencies**

**Figure 7.2.5: Measured real and imaginary part of the PA impedance at 20 V and different drive power levels, (a) at $I_{DQ}=90$ mA, (b) at $I_{DQ}=200$ mA**
80 pF and 120 pF. The measurement results are significantly less noisy for the higher quiescent current. While a decrease in noise level was expected as the current that needs to be measured is larger for smaller impedances, the difference between the resistances is not significant enough to explain the degree in difference present in the measurement. Other interesting effects present in Fig. 7.2.5b include the capacitance which decreases with frequency more than can be explained by the presence of the 50 Ω, 10 pF RC circuit and the rate of the capacitance decreasing which depends significantly on the RF power level. These measurements demonstrate that modelling the PA for the purpose of optimising the output filter of the modulator for it is challenging as it significantly depends on the quiescent current and, to a lesser degree, on the drive power. This additional dependence of the PA impedance on the drain voltage is also present in the measurements on a different transistor in [291]. As the quiescent current is a design choice that can be fixed and the drain voltage is a function of the drive power because of the shaping function, any model either requires knowledge of these factors, assuming them to be constant, or needs to be parametrised. One issue with the measurements and modelling so far is that, due to the limitations of the measurement, they can not capture the dynamic behaviour of the PA’s baseband impedance. Whether a model that considers drain voltage, quiescent current and RF input power is able to predict the instantaneous PA impedance, is an unknown factor and subject to ongoing research.

7.3 Supply Modulator Ripple

As discussed in section 3.1.5 and section 3.2, the ripple introduced by the supply modulator is attenuated by the output filter but there will always be remnants of it in the modulated supply voltage, an example is shown in Fig. 7.3.1. This example has been measured as part of the previous baseband impedance measurements with the modulator from [338] with a switching frequency of 75 MHz set to provide a constant DC voltage and a 1 MHz two-tone excitation, as discussed in section 7.1.2. To evaluate the impact of this ripple on the PA, a series of measurements is conducted that systematically investigate the effect of ripple. The measurement system used for this is based on the National Instrument ET characterisation system, see Fig. 7.3.2, with an additional arbitrary waveform generator that generates the ripple voltage and a supply modulator build of two operational amplifiers which amplify the modulated supply voltage to the desired level. Using a diplexer consisting of a lumped low-pass filter and a lumped high-pass filter, the modulated supply voltage and the ripple voltage are combined. The PA that is supplied with this combined voltage is a 2.09 - 2.19 GHz PA based on the Wolfspeed CGH40010F, a device also used in previous parts of this thesis. The PA is biased in
class AB and driven 1 dB into compression at the peak envelope power using a 10 MHz LTE signal for all data points. The supply voltage is generated from the envelope using a raised cosine function as discussed in section 3.1.2, RF signal and modulated supply voltage are then time-aligned and optimised as discussed in section 3.1.4. The centre frequency of the mixing products is expected to be separated from that of the signal by the ripple frequency so the ripple frequency has to be chosen to be high enough that the mixing products are not going to overlap with the third and fifth order intermodulation distortion of the original signal but low enough to not need an excessive amount of bandwidth in the receiver, resulting in a 41 MHz signal. This is significantly lower than the frequency a buck-converter with a 10 MHz modulation bandwidth would have, see section 3.2, but as the effects of the ripple voltage do not depend on the frequency as long as there is no overlap with either modulation bandwidth or RF bandwidth, the chosen ripple frequency will give representative results. Another simplification is the use of a single frequency. While the ripple voltage of real modulators have more frequency components [334], they are significantly lower in magnitude, lessening their impact, see Fig. .0.2 in the appendix on page 265. To make the ripple voltage magnitude, $\hat{v}_r$, more universally comparable, it is normalised to the supply voltage, in this case 22 V. The normalisation procedure is demonstrated using the two values of 0.06 and 0.1 in Fig. 7.3.3.

### 7.3.1 Impact of Ripple on PA Linearity

Using the established linearity measures EVM and ACPR, the in-band and the out-of-band distortion of the PA is determined. Fig. 7.3.4a demonstrates that the EVM increases from 5.7% with no ripple to 9.6% for a for a normalised ripple voltage of $\hat{v}_r = 0.1$. Following the same trend, the ACPR increases with the ripple magnitude, from -33.1 dBc without ripple to -26.5 dBc for $\hat{v}_r = 0.1$. These established measures do not carry any information outside the carrier and the third-order intermodulation shoulders however, importantly missing any information about mixing products. Investigating the spectra

---

Figure 7.3.1: Measured waveforms between modulator and PA showing the voltage (blue) with ripple and (red) without ripple
at the output of the PA for the different ripple magnitudes in Fig. 7.3.4b rectifies this and clearly shows the ripple-induced side-bands, the results of the mixing between ripple frequency and input signal. As the theory predicts, there are two side-lobes, the centre of which is separated from the centre of the signal by the ripple frequency of 41 MHz. As predicted by (3.7), the magnitude of the side-lobes increases with $\hat{v}_r$.

### 7.3.2 Mitigating the Effect of Ripple

If the ripple magnitude can not be reduced before reaching the PA, other methods of mitigating its effects might be needed. The prevalent ways of linearising ET PAs, DPD and shaping functions [173], [174], will be examined as ways to mitigate ripple in this section, evaluating their capability to linearise the PA in terms of both the classical linearity measures and the ripple-induced side bands. The advantage of using DPD is that it is already a part of most modern transmission systems and can possibly adapted to correct an additional non-linearity. Applying an industry standard generalised memory polynomial (GMP) that comes preconfigured with the National Instruments software, the EVM and the ACPR can be improved significantly, however not to the level of the pre-distorted PA with no ripple, see Tab. 7.1. DPD fares even worse when it comes to linearising the ripple-induced
Figure 7.3.3: Definition of the voltage normalisation

Table 7.1: Comparison of PA linearity with and without DPD

<table>
<thead>
<tr>
<th>$\hat{v}_r$</th>
<th>DPD</th>
<th>EVM</th>
<th>ACPR$_{\text{high}}$</th>
<th>ACPR$_{\text{low}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No</td>
<td>5.7%</td>
<td>-33.1 dBc</td>
<td>-30.8 dBc</td>
</tr>
<tr>
<td>0</td>
<td>GMP</td>
<td>2.3%</td>
<td>-39.0 dBc</td>
<td>-38.5 dBc</td>
</tr>
<tr>
<td>0.1</td>
<td>No</td>
<td>9.6%</td>
<td>-26.5 dBc</td>
<td>-25.8 dBc</td>
</tr>
<tr>
<td>0.1</td>
<td>GMP</td>
<td>3.7%</td>
<td>-34.1 dBc</td>
<td>-35.5 dBc</td>
</tr>
</tbody>
</table>

side-lobes. Their magnitude reduces by less than one dB and the DPD introduces additional distortion trying to correct for this, see Fig. 7.3.5. While this is possibly a result of the limited bandwidth of the DPD, it is something representative of a real system, for a ripple frequency significantly higher than modulation bandwidth, the DPD would need to cover significant bandwidths. In the example of a signal with a modulation frequency of 40 MHz, the switching frequency could be expected to be around 200 MHz. To cover both side-lobes, the DPD would need a bandwidth of 400 MHz. For higher modulation frequencies, switching frequency and DPD bandwidth would increase even further.

An additional issue for the DPD is the lack of correlation between RF signal and ripple signal, again something representative of a real system. While it has been shown that it is possible to pre-distort the ripple-induced side-bands if they are generated to be coherent with the carrier [334], which increases the system complexity significantly. The measurement results in Tab. 7.1 demonstrated that, in principle, the DPD can linearise uncorrelated ripple, however.

A different approach to reducing the impact of ripple on the PA linearity is by adapting the shaping function. As discussed in section 3.1.5, the knee-region interaction is the source of a part of the additional distortion due to the load-line being moved closer to the knee region for half of the ripple voltage period. By increasing the DC offset of the modulated supply voltage, i.e. shifting the whole supply voltage up by the ripple voltage magnitude, the load-line is prevented from being pushed into the knee region by the ripple voltage, never reaching values below those in the ripple-free case with the unmodified supply voltage, see Fig. 7.3.6. As the supply voltage is increased, the efficiency will be reduced however. This is confirmed in the measurements, see Tab. 7.2. The average PAE
Integrating PA and Supply Modulator - Supply Modulator Ripple

Figure 7.3.4: Measured linearity versus normalised ripple magnitude $\hat{\nu}_r$ at a constant PEP and with no DPD: (a) EVM and ACPR (b) Power spectral density (PSD) at the PA output

Table 7.2: Comparison of PA linearity for different levels of voltage shift with no DPD

<table>
<thead>
<tr>
<th>$\hat{\nu}_r$</th>
<th>Voltage Shift</th>
<th>EVM</th>
<th>$\text{ACPR}_{\text{high}}$</th>
<th>$\text{ACPR}_{\text{low}}$</th>
<th>Average PAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0 V</td>
<td>9.6%</td>
<td>-26.4 dBc</td>
<td>-25.8 dBc</td>
<td>35.3%</td>
</tr>
<tr>
<td>0.1</td>
<td>1 V</td>
<td>7.4%</td>
<td>-29.5 dBc</td>
<td>-28.5 dBc</td>
<td>34.1%</td>
</tr>
<tr>
<td>0.1</td>
<td>2 V</td>
<td>6.2%</td>
<td>-32.3 dBc</td>
<td>-30.8 dBc</td>
<td>32.1%</td>
</tr>
<tr>
<td>0.1</td>
<td>3 V</td>
<td>5.5%</td>
<td>-33.9 dBc</td>
<td>-32.6 dBc</td>
<td>31.7%</td>
</tr>
</tbody>
</table>

Fig. 7.3.7. This demonstrates that the $a_{11}$ coefficient in (3.7) that determines the mixing does not depend on supply voltage and the knee region. This demonstrates that through conventional methods, the out-of-band distortion introduced by the ripple voltage can not be easily removed. As a result, it has to be dealt with at the source, i.e. the modulator, by choosing a topology that has very little ripple to start with, by designing the filter to have a very high attenuation at the ripple frequency.
7.4 Conclusions

This chapter discussed the interfaces between PA and modulator. The baseband impedance has been shown to have a significant impact on the linearity of the PA, however, the baseband impedance of supply modulator is often an unknown factor. Using a PA and a modulated signal, the static baseband impedance of arbitrary supply modulators, or any other kind of voltage source, can be measured with a single shot measurement, resulting in the data necessary to develop a static model for the baseband impedance. By conducting this measurement at different conditions, the data necessary for a dynamic model can be collected. The other side of the PA-modulator interface is the impedance the PA presents to the supply modulator. By operating the PA in CW and modulating the supply modulator, the PA impedance can be measured. The measurements showed that the PA impedance depends significantly on the quiescent current and the drive power level, complicating both its modelling and its use during
supply modulator design. The last element of PA-modulator interaction demonstrated the impact of the ripple frequency of the supply modulator on the linearity of the PA, showing that the ripple induced side-bands are hard to get rid of, demonstrating again that integrating PA and modulator in not trivial and results in a number of interactions that need to be considered.

References


Integrating PA and Supply Modulator - Conclusions


Chapter 8

Conclusions and further work

Supply modulation is a highly promising PA architecture to provide the high efficiencies over the wide power ranges and wide bandwidths required by PAs for current and future communication systems. This work investigated the research fields and the developments both of PAs in general and supply modulated systems in particular. Based on this literature review, supply modulated systems are investigated, starting with the transistor used in the final stage of the PA, followed by the input matching network, one possible way to realise supply modulators and ending with the interaction between PA and modulator.

This thesis discussed how the transistor determines how well a PA reacts to supply modulation, presenting some of the traits that make a GaN HEMT suitable for supply modulated PAs. One important factor in that is the supply voltage dependent non-linearity of the intrinsic capacitances. These non-linearities have multiple effects on the PA. As the output capacitance, which comprises $C_{DS}$ and $C_{GD}$, changes with the supply voltage, the power, efficiency, gain and linearity contours shift with supply voltage which necessitates multiple trade-offs during the design process to achieve a high average efficiency and high linearity. Another effect of the non-linear capacitances is the gain variation with supply voltage. Here, the gain changes with the applied supply voltage in addition to changes with drive power. Gain variation is demonstrated as a prevalent phenomenon in GaN HEMTs which introduces several issues in supply modulated PAs and impacts linearity, linearisability, and restricts the shaping function. To be able to compare the gain variation of different transistors, a set of guidelines was established that allows meaningful comparison. Using these guidelines, gain variation has been shown to be present in all types of transistors with GaN HEMTs showing a particularly wide range of gain variation, with many transistors exhibiting a large degree of gain variation. Using a small-signal model, the origin of gain variation in GaN HEMTs was for the first time traced back to the trajectory of the supply voltage dependent $C_{GD}$ that results from the gate field plate. Optimising the geometry of the gate field plate is therefore proposed as a new way of consciously optimising a transistor for supply modulated PAs. Both reducing the gate field plate length and reducing the passivation thickness between gate field plate and barrier have been proposed, any change of the gate field plate geometry will however impact the knee-walkout, requiring a careful trade-off, which is demonstrated using two transistors that differ.
Conclusions and further work -

exclusively in terms of gate field plate length. The gain variation could be reduced at the cost of a slightly increased knee-walkout, demonstrating the feasibility of this approach.

The non-linearity of the intrinsic capacitances has another effect on the linearity of supply modulated PAs. As both the feedback capacitance and the gate-source capacitance are non-linear, the input impedance of GaN HEMTs changes with supply voltage and drive power, which prevents the input from being matched under all conditions. The fact that the input impedance varies significantly implies that there has to be a way of optimising the matching network for the case of a supply modulated PA. Systematically establishing how gain, gain variation and linearity depend on the input impedance presented to the transistor showed new ways of optimising the input matching network for improved linearity, which is demonstrated both in simulation and measurement. Both prove that this novel concept works and that the linearity can be improved with little to no impact in terms of PAE. As this linearity improvement is a result of the non-linear behaviour of the $C_{GD}$, an optimisation to reduce the gain variation is likely to reduce the achievable linearity improvement.

Supply modulation achieves high efficiencies by moving the challenges in achieving high efficiencies over wide power ranges from the PA into the supply modulator, resulting in them having a significant impact on the system efficiency and linearity. This work had a detailed look at RF modulators, a supply modulator topology which has the potential to achieve very high modulation bandwidths by converting the static input DC voltage to a time varying RF signal and rectifying this RF signal resulting in a modulated voltage. With PAs being well understood, this thesis focuses first on understanding the rectifier side of RF modulators, using active load-pull/source-pull measurements of a 0.25 µm Qorvo GaN HEMT to analyse the measured intrinsic gate and drain waveforms of a self-synchronous rectifier for the first time. Using a gate termination, the magnitude and phase of the gate voltage is generated with the intrinsic capacitances of the transistor being used as a feedback network. Again, the non-linearity of $C_{GD}$ has an impact, resulting in the optimum impedance shifting with power level, the behaviour is particularly pronounced for small input power levels. When analysing loss mechanisms in rectifiers, this was demonstrated to be one of the limiting factors as it leads to the transistor conducting at the wrong time. Another loss mechanism that was revealed is a previously unknown effect that results in current flow in the transistor’s off state at high drain voltages, this effect is unexplained at the time of writing. At lower drain voltages, rectifier efficiencies above 90% were demonstrated.
Conclusions and further work -

for wide power ranges and wide load resistor ranges with a class F rectifier. While traditionally, PA and rectifier have been treated separately, this work proposes a novel concept, class B*, which utilises their interaction by using the even harmonic components generated by the rectifier to improve the PA efficiency. Simulations demonstrate this working, and measurements of the rectifier subjected to the class B* impedance environment show that the generated even harmonic power is slightly lower than predicted but sufficient to increase the PA efficiency significantly. A proof of concept using two Wolfspeed CGH40010F 10 W devices is manufactured and tested, showing promising first results with the efficiency reaching up to 78%. At most operating points however, the efficiency was significantly lower, showing a lot of room for improvement and optimisation. Additionally, the efficiency was only high at the maximum input power. A proposed frequency modulation technique might be able to solve this.

With both PA and modulator discussed, their interface is another critical part of a supply modulated system as the impedances these two present to each other have been shown to impact the linearity of the PA and the ability of the modulator to accurately generate the required modulated supply voltage. This thesis describes a new, simple way of characterising these impedances using different kinds of modulated signal, both simple multi-tone signals and realistic wideband communication signals. Using a state-of-the-art modulator, measurement and simulation of the baseband impedance of the modulator were compared to verify the measurement, showing good agreement under static conditions. The measurements also demonstrated that the modulator’s baseband impedance changes significantly with the operating conditions, making them a useful tool in acquiring the data necessary to model this more complex behaviour. Using the same approach, the impedance the PA presents to the modulator can be characterised. The measurements show that this impedances depends significantly on the quiescent current and the drive power, complicating the modelling of this impedance. As it is this impedance the modulator is going to be designed for, a realistic model is relevant to be able to achieve high efficiency and accurate supply voltage tracking. Another interaction of modulator and PA occurs not at the modulation frequency but at the switching frequency of the modulator which is present as an added ripple voltage in the modulated supply voltage at the modulator’s output. The ripple not only reduces the linearity of the PA by pushing the RF load-line into the knee region, it also introduces mixing products, resulting in ripple-induced side-bands which are hard to get rid of.
Conclusions and further work -

Future work

“Don’t adventures ever have an end? I suppose not. Someone else always has to carry on the story.”

J.R.R. Tolkien, *The Fellowship of the Ring*

This work went into detail in different areas, and while it extended the body of knowledge in these areas, there are still threads to follow. With a limited amount of time and resources, there was no way of pursuing all paths. There is a number of ways in which the research presented in this thesis could be continued:

- The gate field plate geometry has been demonstrated to have a significant effect on $C_{GD}$ and, as a result of that, on gain variation and input impedance variation. Systematic variations of the gate field plate could be used to establish an acceptable trade-off between linearity and knee walkout. By using modulated signals and a modulated supply voltage, this optimisation process would yield the impact of the geometry on linearity and efficiency could be measured under realistic conditions. Modulated active load-pull systems such as the one presented in [422] could be used to achieve this.

- While a proof of concept for a class B* modulator has been presented, it only worked well in a limited power range. As the models of the transistors used in the design were not able to operate in the third quadrant, there was little room for optimisation. By choosing different transistors with a model that is capable of representing the transistor’s behaviour under these conditions, a significant improvement of the modulator efficiency should be possible, even when sticking to the simple $\lambda/4$ transformer design.

- To keep the efficiency of a class B* modulator high over a large output power range, an impedance trajectory that moves the load impedance of the PA was proposed in (6.7). By designing a matching network that changes with frequency to follow that trajectory, it should be possible to realise a modulator that is able to output a highly dynamic voltage while keeping a high efficiency.

- The rectifier measurements showed a current peak in the off-state of the transistor when the voltage was rising, capacitive effects and voltage breakdown were excluded as explanations. This behaviour would significantly limit the efficiency of a RF modulator, making it crucial to describe this process in detail. It would be necessary to verify this behaviour and vary the conditions to establish if it is indeed a real effect and not just a measurement artefact and, once
that is achieved, to trace its origin and find out how prevalent this effect is in GaN HEMTs, i.e. whether it is a common effect in GaN HEMTs or just present in a small subset of devices. This would also potentially open a way to optimise GaN HEMTs for rectification applications.

- Measurements of the impedance of the PA showed a strong dependence on quiescent current and drive power. The modulator design requires reasonable assumptions of its load impedance, a model that considers drain voltage, quiescent current and RF input power would be able to supply the modulator designer with the load they need. However, such a model which is able to describe the instantaneous impedances needs to be developed first.

- The other side of the PA-modulator interface, the baseband impedance of the modulator, is hard to model as well. It needs to take load impedance, output voltage and the switching behaviour into account, resulting in a complex model. As the baseband impedance has a significant impact on the system linearity, a realistic model is crucial during system development. Switching models based on transient analysis are not compatible with the envelope simulations used in simulation software, requiring development of a suitable model.

In addition to research directly continuing the research conducted in the course of this PhD, there are also topics that do not directly continue the research conducted in this thesis but branch off instead.

- While typically, the gate field plate is a rectangular feature that extends over the gate edge, other geometries might be able to achieve the desired electric field distribution, e.g. a bar in parallel to gate on a thick layer of passivation, connected to the gate in only a few spots. This should reduce the gate drain capacitance at the cost of a more complex mask design, whether the desired field distributions can still be achieved and whether the surface traps can be suppressed as desired will need to be determined in simulation and measurement.

- Instead of optimising the geometry of the gate field plate, there might be other ways to optimise the transistor for a small and static feedback capacitance. This could include for example the use of different materials between barrier and gate field plate, such as polyimides or polybenzoxazole [423], whether these materials can achieve a beneficial trapping behaviour or whether they would require a thin layer of traditional passivation between them and the barrier is an open question.

- The input impedance of a GaN HEMT has shown to vary significantly with supply voltage and drive level. In the case of a multi-stage PA, all but the final stage see the input impedance of the following stage as their load, any input impedance change directly leads to a load impedance change in the following stage. This could prove problematic and needs to be investigated. However, it might also be possible to use this shift in load impedance to move the load the
driving stages see in a way that the efficiency of the driving stages is kept high over a wide power range.

- The output filters of SMMs are designed to reduce the ripple to an acceptable level, and provide the necessary conditions for zero voltage switching while letting the modulated signal pass with little impact on amplitude and phase of the signal. The impact on the PA plays so far no role in the output filter design. While taking the whole system into account would make the filter design more challenging, it might be able to improve system linearity at little to no cost in terms of efficiency.

- The ripple voltage has been shown to have a detrimental effect on the PA linearity. As discussed in section 3.2.3, spreading the switching of the SMM over a wider bandwidth can reduce the magnitude of individual ripple components. This has however not been shown in combination with supply modulated PAs, however, this combination might prove very beneficial as it should reduce the ripple induced side-bands significantly.

- The ripple-induced side-lobes were shown to be virtually independent of the supply voltage and the knee region but exclusively depend on the $a_{11}$ coefficient. This coefficient should be dependent on the gate voltage, possibly allowing ripple suppression by gate bias modulation.

- While the impedances of both modulated PA and modulator have been characterised individually, their respective counter-part has been kept constant. In a real system, they would be modulated at the same time which might result in interactions between the two systems when, instead of being presented with a fixed impedance with a positive real part, they are presented with an impedance with a negative real part. Measuring this would require knowledge about the power flow between the two sub-systems. Using a measurement system such as the one presented in [396], which incorporates low frequency directional coupler in the supply voltage path, would allow measurements of the impedances of PA and modulator under modulated conditions. In the absence of such a system, using tickle tones of different frequencies for PA and modulator and measuring voltage and current between PA and modulator should lead to comparable results, or, at least, allow to make a statement on whether there is a significant amount of interaction.

- The combiner in sequential PAs is isolating and dissipates power in most parts of the output power range. By replacing the absorber in the combiner with a rectifier, a significant portion of that power could be recovered, improving the efficiency of sequential PAs significantly.
Conclusions and further work -

References


Bibliography


Bibliography


Bibliography


Bibliography


241

Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


List of Figures

1.1.1 PAPR for mobile communication standards after [11] ........................................... 2
1.1.2 Power density function (PDF) for an LTE signal and efficiency of an ideal class B amplifier over normalised output voltage .................................................. 3
1.1.3 Voltages, powers and efficiencies of an ideal class B PA with and without envelope tracking ................................................................. 4
2.0.1 Diagrams showing (a) mobile devices and their connections and (b) a transceiver system (after [14]) .................................................. 11
2.1.1 Development of baseband bandwidth in mobile communication systems .............. 12
2.1.2 Development of available sub-6 GHz frequencies for mobile communication systems worldwide, frequency bands rounded up to full 100 MHz, proposed frequency bands in orange .................................................. 13
2.1.3 Illustration of symbols in digital modulation schemes: (a) 16QAM and signal vector in the IQ plane and (b) 16APSK .................................................. 14
2.1.4 Statistical analysis of two signals, (a) probability density function and (b) complementary cumulative distribution function .................................. 14
2.2.1 Simulated characteristics of a Qorvo GaN-on-SiC HEMT: (a) DCIV and (b) transfer characteristics, drain current vs. gate voltage (solid lines) and transconductance vs. gate voltage (dotted lines) .................................................. 16
2.2.2 Simple model of a FET showing capacitances and resistances .............................. 16
2.2.3 Simplified band structure of a capped AlGaN/GaN HEMT at the gate (after [28]) .. 17
2.2.4 Layer stack of a GaN-on-Si wafer for RF PA applications .................................. 18
2.2.5 Schematic showing substrate coupling ................................................................. 19
2.2.6 Geometry of a two finger GaN HEMT: (a) top view and (b) lateral view .............. 22
2.2.7 Connecting multi-finger transistors using (a) planar gate and drain, (b) planar drain and source, and (c) source via holes .................................................. 24
2.2.8 Intrinsic load-line resistance as a function of supply voltage (solid), for common supply voltages in mobile communications (dotted) ................. 27
2.2.9 Supply voltage over $f_T$ of different GaN foundries, bubble size shows power density in W/mm, data from [100]–[103] and measured ............................................. 28
2.3.1 IV Measurement: (a) Schematic of the measurement setup and (b) measurement results .......................................................... 30
2.3.2 Load-pull measurement results of a Wolfspeed CGH40100F at 2 GHz .................. 32
2.3.3 Fan diagram obtained using load-pull setup, different colours show different supply voltages ................................................................. 33
2.3.4 Active harmonic load-pull setup by Mesuro ................................................................. 33
2.4.1 Comparison of PA classes of operation (a) loadlines at the current generator plane and (b) voltage and current waveforms at the current generator plane ................................................................. 36
2.4.2 Output power, efficiency and gain over conduction angle $\varphi_c$ at maximum drain current 37
2.4.3 Output power and efficiency over input power for an ideal class B PA ...................... 38
2.4.4 Magnitude of harmonic frequency components of the drain current over $\varphi_c$ of ideal conduction angle PAs, normalised to the maximum fundamental current at class A .... 39
2.4.5 Power and efficiency over conduction angle $\varphi_c$ for transistors with no knee region (solid line) compared to transistors with 10% $V_{max}$ (dashed line) and 20% $V_{max}$ (dotted line) 40
2.4.6 Matching networks in PAs and impedance definitions ........................................... 41
2.4.7 Voltage and current waveform and the resulting losses of a class B PA ................. 42
2.4.8 Voltage and current waveform of a class F PA for different numbers of controlled harmonic frequencies ................................................. 42
2.4.9 Theoretical continuous class F mode (a) waveforms at the current generator plane and (b) voltage and current waveforms ........................................... 44
2.4.10 Recovering efficiency at 6 dB OBO: Class B with an efficiency of 19.6% compared to class B with supply modulation and load modulation, both with an efficiency of 78.5% 45
2.4.11 Effect of supply voltage change on class B load-lines for different drive levels showing how reducing the supply voltage results in a minimum voltage close to zero .... 46
2.4.12 Schematic of a supply modulated PA ................................................................. 46
2.4.13 Measured “Booth” chart of a Wolfspeed CGH40010F GaN HEMT, showing output power versus input power for different supply voltages and the approximate regions . 47
2.4.14 Measured contours of (a) PAE in % and (b) Output power in dBm of a 250 nm GaN HEMT at 1 GHz for different supply voltages: 10 V (orange), 20 V (blue) and 30 V (green) ................................................................. 47
2.4.15 Class B load-lines for different normalised load impedances, 0.25, 0.5, 0.75 and 1 . . 48
2.4.16 Schematic of a Doherty PA ................................................................. 49
2.4.17 Efficiency versus output power back-off for different ideal Doherty PAs (green symmetrical, purple asymmetrical) compared to an ideal class B PA (red) .......... 50
2.4.18 Powers, load resistances, voltages, and currents versus normalised output power in a Doherty PA ................................................................. 50
2.4.19 LINC vector diagram (after [207, p.34]) ................................................................. 52
2.4.20 Schematic of Chireix (a) and LINC (b) outphasing transmitter ......................... 53

253
List of Figures

2.4.21 Schematic of a LMBA ................................................. 54
2.4.22 Schematic of a sequential PA ........................................ 55
2.6.1 Vectors in the definition of the error vector magnitude (EVM) after [241, p.20], with \( \vec{s}_m \) being the measured vector, \( \vec{s}_0 \) the original vector and \( \vec{s}_{err} \) the error vector .......... 59
2.6.2 1 dB compression point of an amplifier with \( a_1 = 10, a_3 = -4, a_5 = -1 \) after [242, p.27] 60
2.6.3 Frequency domain response of a non-linear system with a two-tone excitation after [242, p.34] ............................................. 62
2.6.4 Frequency domain response of a non-linear amplifier with a modulated signal .......... 62
2.6.5 Concept of pre-distortion .............................................. 64
3.0.1 Overview of a supply modulated PA system showing the internal power flow .......... 79
3.1.1 Different shaping functions defined as supply voltage dependent on the input envelope magnitude .......................................................... 81
3.1.2 Simulated spectra of the a 10 MHz 64 QAM signal with different shaping functions . 82
3.1.3 Envelope signal (thin, blue line) with correct (solid, green line) and incorrect (dashed, red line) time alignment between RF signal and baseband signal ......................... 84
3.1.4 Frequency components at different points in an ET system .............................. 85
3.1.5 Load line of an ideal class B PA effected by ripple voltage .............................. 86
3.1.6 Definition of impedances at the interface between PA and Modulator ................. 87
3.1.7 Simulated real part of the PA impedance presented by the PA for different shaping functions ................................................................. 89
3.2.1 Block diagram showing signal and power flow around the supply modulator .......... 90
3.2.2 (a) Schematic and (b) waveforms of a multi-level supply modulator (after [305]) .... 92
3.2.3 Comparison of (a) multi-level (after [305]) and (b) cascaded multi-level supply modulator (after [297]) ........................................... 93
3.2.4 (a) Schematic and (b) waveforms of a switch mode modulator (after [294]) ........ 94
3.2.5 DCIV Simulation and measurement results of (a) Wolfspeed CGH40010F GaN-on-SiC HEMT and (b) Qorvo 6x80 \( \mu \)m GaN-on-SiC HEMT .................................................... 96
3.2.6 Drivers and output filter in a buck-converter type modulator ................................ 97
3.2.7 20 MHz LTE envelope signal power spectrum, with (a) linear and (b) logarithmic frequency axis ......................................................... 100
3.2.8 Schematic of a RF DCDC Modulator ....................................... 101
3.2.9 Schematic of RF DCDC converters with different ways to generate the rectifier’s gate signal, after [351] ............................................. 102
3.3.1 Measurement results of a Qorvo GaN-on-SiC HEMT showing the drain voltage dependent output capacitances $C_{DS}$ and $C_{GD}$ ......................................................... 105

3.3.2 Measurement results of a GaN-on-Si HEMT (a) fan diagrams showing the knee-walkout and (b) impact of knee walkout on efficiency for different target current levels, solid grey lines indicate maximum class B efficiency considering the knee voltage .... 106

4.1.1 Picture of the active harmonic load-pull measurement system used for the device-plane measurements in this thesis ................................................................. 124

4.1.2 Picture of a Qorvo 6x80 µm GaN-on-SiC HEMT through the microscope of the probe station ................................................................. 125

4.1.3 Measured load-pull contours of an E008 GaN-on-Si HEMT presented with a class-F impedance environment at different supply voltages: green at 30 V, blue at 20 V, orange at 10 V, grey points are the measured points. (a) PAE in % and (b) Output power in dBm ......................................................... 126

4.1.4 Measured points of maximum output power and maximum PAE for different supply voltages ......................................................... 127

4.1.5 Measured gain (dotted lines) and PAE (solid lines) over output power characteristics of a GaN HEMT for different supply voltages at $Z_L = 265+j*84 \Omega$ ...................... 128

4.1.6 Comparison of bare die and packaged device, showing device and package parasitics . 129

4.1.7 Picture of (a) the heatsink and carrier, (b) the calibration standards, and (c) the complete kit in the measurement setup ......................................................... 130

4.1.8 Measured load-lines used to calculate the intrinsic capacitances ......................................................... 131

4.1.9 Calculated (solid lines) and modelled (dotted lines) non-linear capacitances (a) $C_{GS}$ at $V_{DS} = 0$ V and (b) $C_{DS}$ at different values of $V_{GS}$ ......................................................... 132

4.1.10 Calculated (solid lines) and modelled (dotted lines) non-linear $C_{GD}$ at different values of $V_{GS}$ plotted over (a) $V_{DS}$ and (b) $V_{DG}$ ......................................................... 132

4.1.11 Measured (1,2) and internal voltages and currents ......................................................... 133

4.1.12 Measured current waveform with different levels of de-embedding: No de-embedding, linear de-embedding, non-linear de-embedding of only $C_{DS}$ and non-linear de-embedding of both $C_{DS}$ and $C_{GD}$ ......................................................... 134

4.1.13 Drain impedances of a transistor operated as a rectifier at the fundamental frequency for different power levels with (a) no deembedding, (b) linear deembedding and (c) full non-linear deembedding ......................................................... 135

4.1.14 Schematic of an AlGaN/GaN HEMT including gate and source field plate. Elements based on [86], [125] ......................................................... 136
4.1.15 Small-signal equivalent circuit of a HEMT with gate and source field plate, elements visible in Fig. 4.1.14 are enclosed in the dotted rectangle, non-linear components that change with applied voltages are marked with an arrow. 136

4.1.16 Microscope image of a 0.25 µm gate length GaN-on-Si HEMT with a gate width of 250 µm, 2 * 125 µm. 137

4.1.17 Comparison of modelled (crosses, light colours) and measured (circles, dark colours) extrinsic device S-parameters from 0.1 GHz to 30 GHz for (a) \( V_D = 0 \) V, forward biased, (b) \( V_D = 0 \) V, pinched off, (c) operation at \( V_D = 5 \) V, \( I_D = 60 \) mA and (d) operation at \( V_D = 30 \) V, \( I_D = 60 \) mA. 138

4.2.1 Measured gain (dotted lines) and power added efficiency (solid lines) over output power characteristics of a Sheffield E008 GaN HEMT for different supply voltages and with a different fundamental load impedances: (a) \( 121 + j*51 \) Ω and (b) \( 265 + j*84 \) Ω. 139

4.2.2 Contours showing the difference in gain in dB between supply voltages of 30 V and of 10 V of a Sheffield E008 GaN-on-Si HEMT at 1 GHz. Grey dots represent measured impedances, stars show the impedances for maximum efficiency and power at 10 V and 30 V. All measurements conducted under CW excitation, gain averaged over power levels. 140

4.2.3 Measured static transconductance and transfer characteristics: (a) \( g_m \) vs. \( I_D \) and (b) \( I_D \) vs \( V_{GS} \) of a Sheffield E008 GaN HEMT for different supply voltages. 140

4.2.4 Measured magnitude of \( S_{11} \) of a Sheffield E008 GaN HEMT for different supply voltages and increasing input powers for a load impedance of \( 265 + j*84 \) Ω. 141

4.2.5 Demonstration showing how to evaluate the gain variation of a transistor according to the proposed guidelines. 142

4.2.6 Measured power added efficiencies for different supply voltages and the supply voltage trajectory giving the maximum PAE (solid black, symbol-less lines), (a) over input power and (b) over output power. 144

4.2.7 Measured extrinsic maximum stable gain/maximum available gain (MSG/MAG) (solid lines) and current gain \( |h_{21}| \) (dotted lines) of a GaN HEMT for different supply voltages. 145

4.2.8 Measured static transfer characteristic \( g_m \) versus \( V_{GS} \) (dotted) and drain current \( I_D \) (solid) of a GaN HEMT for different supply voltages and selected bias point for large signal measurements (dotted line, vertical). 146

4.2.9 Comparison of \( f_{max} \) and \( g_m \) over supply voltage for the measurement with constant drain current and the simulation for two gate bias voltages, -4.8 V and -5.4 V. 147

4.2.10 Comparison of \( f_T \) and \( g_m \) over supply voltage for the measurement with constant drain current and the simulation for two gate bias voltages, -4.8 V and -5.4 V. 147
4.2.11 Extracted gate-drain capacitance $C_{GD}$ and measured transconductance $g_m$ and maximum oscillation frequency $f_{max}$ versus supply voltage ........................................ 148
4.2.12 Extracted gate drain capacitance $C_{GD}$ and gate source capacitance $C_{GS}$ versus supply voltage ................................................................. 149
4.2.13 Comparison of extracted $C_{GD}$ and $C_{GD}$ modified for minimum small-signal gain variation: (a) Simulated MSG/MAG in dB for extracted (dotted line) and modified (solid line) for different supply voltages and (b) comparison of the two $C_{GD}$ shapes versus supply voltage ........................................ 149
4.2.14 Comparison of extracted $C_{GD}$ and $C_{GD}$ modified to represent a $C_{GD}$ curve shifted to start increasing at lower voltages: (a) Simulated MSG/MAG in dB for extracted (dotted line) and modified (solid line) for different supply voltages and (b) comparison of the two $C_{GD}$ shapes versus supply voltage ........................................ 150
4.2.15 TCAD simulation results of $C_{GD}$ of a GaN HEMT with a gate field plate length of 1 µm, 0.5 µm and with no gate field plate versus supply voltage at $V_{GS} = -5$ V, $C_{GD}$ plateau is labelled in blue .................................................. 150
4.2.16 Measurement results of a 0.25 µm GaN-on-Si HEMT with no GFP fabricated at Cardiff University showing (a) extracted $C_{GD}$ versus supply voltage and (b) measured MSG/MAG in dB over frequency for different supply voltages ........................................ 151
4.3.1 Parameters in field plate design ................................................................. 151
4.3.2 Measured gain (dotted lines) and power added efficiency (solid lines) versus output power characteristics of two GaN HEMTs for different supply voltages with the same fundamental load impedance of 265+$j*84$ Ω and two field plate lengths: (a) 1 µm, corresponding to the red curve in Fig. 4.2.15 and (b) 0.5 µm, corresponding to the green curve in Fig. 4.2.15. .................................................. 152
4.3.3 Measured RF/DC dispersion versus supply voltage for the original field plate configuration and an device with a reduced gate field plate ........................................ 153
4.3.4 TCAD simulation showing the behaviour of $C_{GD}$ for a constant gate field plate configuration and different SiNx passivation thicknesses between field plate and barrier versus supply voltage at $V_{GS} = -5$ V .................................................. 154
5.0.1 IMN in a PA and impedance definitions .................................................. 159
5.1.1 Example measurements showing measured (a) real and (b) imaginary part of input impedance of a 2x125 µm 250 nm GaN-on-Si HEMT device for different supply voltages and increasing input powers for a fixed load impedance at 1 GHz ........................................ 160
5.2.1 Source-pull simulation setup ................................................................. 161
5.2.2 Smith chart showing the area of the source-pull simulation .......................... 161
5.2.3 Simulation results showing the small-signal gain at 28 V versus input reactance and resistance. Red triangle: Source impedance providing the highest gain at 28 V; Blue star: An impedance that offers a good trade-off between gain and linearity. ............... 162
5.2.4 Simulation results showing the small-signal gain variation versus input reactance and resistance. Red triangle: Source impedance providing the highest gain at 28 V; Blue star: An impedance that offers a good trade-off between gain and linearity. ............... 162
5.2.5 Simulation results showing the phase variation versus input reactance and resistance. Red triangle: Source impedance providing the highest gain at 28 V; Blue star: An impedance that offers a good trade-off between gain and linearity. ............... 163
5.3.1 Schematic showing the individual part of the designed PA .............................. 164
5.3.2 Simulation results comparing the (a) Gain, (b) Phase and (c) PAE of a GaN HEMT PA versus input power for static supply voltages between 8 and 28 V in 4 V steps for the two IMNs, $IMN_{\text{max}}$ (thin blue, solid lines) and $IMN_{\text{opt}}$ (thin orange, dotted lines). Bold line shows PAE trajectory of PA with envelope tracked supply voltage, blue for the maximum gain at 28 V, red for the optimised IMN ................................. 165
5.3.3 Picture of the fabricated PA with different input matching structures, designed to be easily reconfigurable by soldering connections and replacing lumped capacitors ........... 166
5.3.4 Input reflection coefficient $S_{11}$ versus output power for different input matching networks and different supply voltages: (a) magnitude in dB, (b) plotted in a Smith chart where 10 V is displayed using the darkest colour and 30 V using the lightest colour .......................... 166
5.3.5 Measured Gain versus output power of the PA shown in Fig. 5.3.3 for different input matching networks and different supply voltages ........................................ 167
5.3.6 Measured efficiency versus output power of the PA shown in Fig. 5.3.3 for different input matching networks and different supply voltages, (a) drain efficiency and (b) PAE ........................................ 167
5.3.7 Measured normalised transfer phase versus output power of the PA shown in Fig. 5.3.3 for different input matching networks and different supply voltages ........................................ 168
6.1.1 Comparison of simulated class-F PA (red, dotted) and class-F rectifier (blue, solid) load-lines at the current generator plane, grey lines indicate simulated DCIV curves, green boxes indicate quadrants of IV plane ........................................ 172
6.1.2 Measurement setup used for the characterisation of rectifiers (a) block diagram and (b) picture of the setup ........................................ 173
6.2.1 Simulated DCIV characteristics of a Schottky Diode (green line) and a GaN HEMT (grey lines) ........................................ 174
6.2.2 Simulated DCIV curves (dashed lines) and load-lines (solid lines) of a transistor operating as a rectifier with a constant input power for a number of constant gate voltages 175

6.2.3 Definition of impedances and voltages in self-synchronous rectifiers 176

6.2.4 Measured DC output voltage (dotted line) and efficiency (solid line) over phase of the input reflection coefficient $\Gamma_G$ at a DC load resistance of 328.55 $\Omega$ at -4 dB and -29 dB IBO, $|\Gamma_G|$ is kept between 0.98 and 1 178

6.2.5 Gate voltages at different phases of the gate reflection coefficient $\Gamma_G$ at (a) -29 dB IBO and (b) -4 dB IBO, $|\Gamma_G|$ is kept between 0.98 and 1, blue line shows single drain current waveforms for phase reference 178

6.2.6 Measurement results of the rectifier at low power levels with a gate reflection coefficient phase of 23° and 29°, showing (a) load-lines at the drain of the transistor, with the grey, dotted line showing the DCIV curve for $V_{GG} =0$ V as reference and (b) gate voltages at the two phases 179

6.2.7 Measurement results of the rectifier with a gate reflection coefficient phase of 180°, showing (a) load-lines at the drain for high (red) and low (green) power with the grey, dotted line showing the DCIV curve for $V_{GG} =-3.4$ V and (b) gate voltage waveforms with only the fundamental (solid line) and all measured harmonics (dotted lines), blue line shows single drain current waveforms for phase reference 179

6.3.1 Measured drain and gate impedances for different input powers and DC load impedances (a) at the device plane and (b) at the intrinsic current generator plane 181

6.3.2 Load-lines for different power levels at a DC load resistance of 328.5 $\Omega$, grey line shows DCIV curve at $V_{GG}=0$ V 182

6.3.3 Measured efficiency and output voltage of the rectifier over input power for different DC load impedances 183

6.3.4 Measured rectifier load-lines for different input power levels at a DC load resistance of 56.4 $\Omega$ with the dotted line representing the highest input power 183

6.3.5 Measured rectifier waveforms for different input power levels at a DC load resistance of 56.4 $\Omega$ (a) drain voltage and current waveforms and (b) of gate waveforms 184

6.3.6 Measured load-line de-embedded to the intrinsic current generator plane for a load impedance of 328.5 $\Omega$ for different input powers, measured DCIV curve at $V_{GS} = 0$ V (grey dashed line). The green line indicates an ideal ohmic region 185

6.3.7 Measured rectifier waveforms for different input power levels at a DC load resistance of 458.5 $\Omega$ (a) drain voltage and current waveforms and (b) of gate waveforms 186

6.4.1 Proposed circuit with intrinsic waveforms of a class B* DC/DC converter, MN is a matching network 186
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4.2</td>
<td>Calculated intrinsic drain waveforms of the PA in a class B* DC/DC converter for a different number of present harmonic impedances</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Schematic of the amplifier-rectifier simulation</td>
</tr>
<tr>
<td>6.4.4</td>
<td>Simulated intrinsic drain voltage (blue) and current (red) waveforms of power amplifier and rectifier for a range of input powers</td>
</tr>
<tr>
<td>6.4.5</td>
<td>Output voltage of the complete DC/DC converter and power added efficiency of power amplifier, rectifier and the complete DC/DC converter</td>
</tr>
<tr>
<td>6.4.6</td>
<td>Power flow between PA and rectifier for increasing input power at different frequencies normalised to the maximum fundamental power, sign indicates direction of power flow</td>
</tr>
<tr>
<td>6.4.7</td>
<td>Efficiency over output voltage for different PA load impedances</td>
</tr>
<tr>
<td>6.4.8</td>
<td>Simulated (dotted) and measured (solid) drain waveforms of a 6x80 µm GaN HEMT with a DC load of 328.5 Ω, Γ_in = 1∠50° and input powers between -14 dBm and 35.6 dBm</td>
</tr>
<tr>
<td>6.4.9</td>
<td>Measured efficiency of the rectifier for different load resistances</td>
</tr>
<tr>
<td>6.4.10</td>
<td>Measured drain voltage and current waveforms of a class B* rectifier for different input power levels at a DC load resistance of (a) 182.9 Ω and (b) 56.4 Ω</td>
</tr>
<tr>
<td>6.4.11</td>
<td>Generated harmonic powers normalised to the fundamental power over input power for different DC load resistors</td>
</tr>
<tr>
<td>6.4.12</td>
<td>Measured efficiency of the rectifier taking second and fourth harmonic generation into account</td>
</tr>
<tr>
<td>6.4.13</td>
<td>(a) Circuit diagram and (b) Picture of the PCB used for the measurements</td>
</tr>
<tr>
<td>6.4.14</td>
<td>Measured DC to DC efficiency of the class B* DC/DC converter at 120 MHz for different input voltages and different load resistances (a) over output voltage and (b) over input power</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1.1</td>
<td>Comparison of the in-band distortion introduced by a bias network and a buck-converter (a) Gain showing the AM/AM distortion (b) AM/PM distortion</td>
</tr>
<tr>
<td>7.1.2</td>
<td>Measured output spectrum of a PA excited with a 5 tone signal in the presence of a non-ideal baseband termination</td>
</tr>
<tr>
<td>7.1.3</td>
<td>Simulated 5-tone voltages with and without effect of the baseband impedance showing (a) the input waveform (orange) and the baseband voltage (blue) and (b) the resulting output spectra with (light blue) and without (black) baseband voltage mixing</td>
</tr>
<tr>
<td>7.1.4</td>
<td>Schematic showing concept</td>
</tr>
<tr>
<td>7.1.5</td>
<td>Circuit diagram of the relevant elements of the measurement setup</td>
</tr>
</tbody>
</table>
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1.6</td>
<td>Two-tone signal and the corresponding current waveform in a class B PA in the (a) time domain and (b) frequency domain</td>
</tr>
<tr>
<td>7.1.7</td>
<td>Measurement setup: (a) Schematic and (b) Picture</td>
</tr>
<tr>
<td>7.1.8</td>
<td>Measured voltage and current between PA and modulator with a supply voltage of 16 V and a two tone excitation (a) in the time domain and (b) in the frequency domain</td>
</tr>
<tr>
<td>7.1.9</td>
<td>Measured magnitude and phase of the baseband impedance at 16 V using different types of modulation</td>
</tr>
<tr>
<td>7.1.10</td>
<td>Comparison of magnitude and phase of the baseband impedance of the measurement (green circles) and the model (red, solid line, dotted for extrapolated values)</td>
</tr>
<tr>
<td>7.1.11</td>
<td>Measured magnitude and phase of the baseband impedance of a buck-converter using a WCDMA signal (a) for different power levels at a supply voltage of 20 V and (b) different supply voltages at 200 mA and in saturation</td>
</tr>
<tr>
<td>7.1.12</td>
<td>Measured real part of the baseband impedance using a WCDMA signal and different supply voltages (a) as measured of the buck-converter, (b) of the buck-converter with filtered noise, (c) of the multi-level buck-converter from [338] operated at a switching speed of 75 MHz with filtered noise and (d) of the multi-level buck-converter from [338] operated at a switching speed of 100 MHz with filtered noise</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Circuit diagram of the relevant elements of the PA impedance measurement setup</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Measured voltage and current between PA and modulator with a modulated supply voltage and a shaped five-tone excitation in the frequency domain with $I_{DQ}=200$ mA and 10 dBm drive power</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Measured real and imaginary part of the PA impedance at 20 V and different drive power levels, (a) at $I_{DQ}=90$ mA, (b) at $I_{DQ}=200$ mA</td>
</tr>
<tr>
<td>7.2.4</td>
<td>Equivalent circuit model of the PA at baseband frequencies</td>
</tr>
<tr>
<td>7.2.5</td>
<td>Measured real and imaginary part of the PA impedance at 20 V and different drive power levels, (a) at $I_{DQ}=90$ mA, (b) at $I_{DQ}=200$ mA</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Measured waveforms between modulator and PA showing the voltage (blue) with ripple and (red) without ripple</td>
</tr>
<tr>
<td>7.3.2</td>
<td>ET measurement system adapted for ripple measurement, (a) diagram and (b) picture</td>
</tr>
<tr>
<td>7.3.3</td>
<td>Definition of the voltage normalisation</td>
</tr>
<tr>
<td>7.3.4</td>
<td>Measured linearity versus normalised ripple magnitude $\hat{v}_r$ at a constant PEP and with no DPD: (a) EVM and ACPR (b) Power spectral density (PSD) at the PA output</td>
</tr>
<tr>
<td>7.3.5</td>
<td>Measured PSD at the PA output for cases with and without ripple and with and without DPD</td>
</tr>
</tbody>
</table>
7.3.6 Diagram showing the shape of the supply voltage without ripple (thick orange line), with ripple (dashed thin red line) and with ripple but elevated by the ripple voltage (solid thin green line) ........................................... 219
7.3.7 Measured PSD at the PA output for different levels of envelope voltage shift ........ 220

.0.1 Measured drain voltage and current waveforms of a rectifier for different input power levels at DC load resistances of (a) 182.9 Ω, (b) 328.55 Ω and (c) 458.5 Ω, (d) shows the waveforms at 56.4 Ω for the sake of completeness .............................................. 264
.0.2 Measurement results of a real buck-converter with a switching frequency of 75 MHz showing the voltage ripple and the higher harmonics it contains (a) in the time domain and (b) in the frequency domain ................................................................. 265
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Comparison of electrical properties of semi-conductors</td>
<td>17</td>
</tr>
<tr>
<td>2.2</td>
<td>Comparison of substrate materials for GaN HEMTs</td>
<td>19</td>
</tr>
<tr>
<td>2.3</td>
<td>Comparison of intrinsic capacitances for different transistor technologies</td>
<td>26</td>
</tr>
<tr>
<td>2.4</td>
<td>Comparison of 250 nm technologies</td>
<td>28</td>
</tr>
<tr>
<td>2.5</td>
<td>Comparison of high efficiency PA topologies</td>
<td>56</td>
</tr>
<tr>
<td>2.6</td>
<td>Selected two-tone signal distortion products</td>
<td>61</td>
</tr>
<tr>
<td>3.1</td>
<td>Comparison of supply modulator topologies</td>
<td>103</td>
</tr>
<tr>
<td>4.1</td>
<td>Comparison of gain variation reported in literature</td>
<td>142</td>
</tr>
<tr>
<td>6.1</td>
<td>RF input impedance and DC load resistances</td>
<td>182</td>
</tr>
<tr>
<td>7.1</td>
<td>Comparison of PA linearity with and without DPD</td>
<td>217</td>
</tr>
<tr>
<td>7.2</td>
<td>Comparison of PA linearity for different levels of voltage shift with no DPD</td>
<td>218</td>
</tr>
</tbody>
</table>
Appendix

Figure 0.1: Measured drain voltage and current waveforms of a rectifier for different input power levels at DC load resistances of (a) 182.9 Ω, (b) 328.55 Ω and (c) 458.5 Ω, (d) shows the waveforms at 56.4 Ω for the sake of completeness.
Figure .0.2: Measurement results of a real buck-converter with a switching frequency of 75 MHz showing the voltage ripple and the higher harmonics it contains (a) in the time domain and (b) in the frequency domain.