High Efficiency Microwave Amplifiers and SiC Varactors Optimized for Dynamic Load Modulation

Christer Andersson
High Efficiency Microwave Amplifiers and SiC Varactors Optimized for Dynamic Load Modulation

Christer Andersson

© Christer Andersson, 2013
Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 3511
ISSN 0346–718X
Technical report MC2–243
ISSN 1652–0769
Microwave Electronics Laboratory
Department of Microtechnology and Nanoscience - MC2
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
Phone: +46 (0) 31 772 1000

Printed by Chalmers Reproservice
Göteborg, Sweden 2013
Abstract

The increasing use of mobile networks as the main source of internet connectivity is creating challenges in the infrastructure. Customer demand is a moving target and continuous hardware developments are necessary to supply higher data rates in an environmentally sustainable and cost effective way. This thesis reviews and advances the status of realizing wideband and high efficiency power amplifiers, which will facilitate improvements in network capacity and energy efficiency.

Several demonstrator PAs are proposed, analyzed, designed, and characterized: First, resistive loading at higher harmonics in wideband power amplifier design suitable for envelope tracking (ET) is proposed. A 40 dBm decade bandwidth 0.4–4.1 GHz PA is designed, with 10–15 dB gain and 40–62% drain efficiency. Its versatility is demonstrated by digital pre-distortion (DPD) linearized measurements resulting in adjacent channel leakage ratios (ACLR) lower than \(-46\) dBc for various downlink signals (WCDMA, LTE, WiMAX).

Second, a theory for class-J microwave frequency dynamic load modulation (DLM) PAs is derived. This connects transistor technology and load network requirements to enable power-scalable and bandwidth conscious designs. A 38 dBm PA is designed at 2.08 GHz, maintaining efficiencies \(>45\)% over 8 dB of output power back-off (OPBO) dynamic range. From this pre-study a fully packaged 86-W peak power version at 2.14 GHz is designed. ACLR after DPD is \(-46\) dBc at a drain efficiency of 34%.

For DLM PAs there is a need for varactors with large effective tuning range and high breakdown voltage. For this purpose, SiC Schottky diode varactors are developed with an effective tuning range of 6:1 and supporting a 3:1 tuning ratio at 36 V of RF swing. Nonlinear characterization to enable Q-factor extraction in the presence of distortion is proposed and demonstrated by multi-harmonic active source- and load-pull, offering insights to tunable network design.

Third, a method to evaluate and optimize dual-RF input PAs, while catering to higher harmonic conditions and transistor parasitics, is proposed. The method is validated by a PA design having a peak power of 44 ± 0.9 dBm and 6 dB OPBO PAE exceeding 45% over a 1–3 GHz bandwidth.

The results in this thesis contribute with a novel device and analysis of high efficiency and wideband PAs, aiding in the design of key components for future energy efficient and high capacity wireless systems.

**Keywords:** GaN, energy efficiency, HEMT, load modulation, nonlinear measurements, power amplifiers, SiC, varactors, wide bandgap technology
List of publications

Appended publications

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


Other publications

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


Notations and abbreviations

Notations

$\beta$ Normalized transistor output current amplitude
$\epsilon_r$ Relative dielectric constant
$\eta$ Drain efficiency
$\Gamma_L$ Load reflection coefficient
$\Gamma_S$ Source reflection coefficient
$\omega_0$ Reference (angular) frequency
$f_0$ Reference frequency
$I_{DC}$ PA DC supply current
$I_{ds}$ Transistor drain-to-source current
$I_{gs}$ Transistor gate-to-source current
$I_{max}$ Transistor maximum current
$N_D$ Doping concentration
$p$ Output power probability density function
$P_{out}$ PA RF output power
$P_{DC}$ PA DC power consumption
$V_{br}$ Transistor breakdown voltage
$V_{DC}$ PA supply voltage
$V_{ds}$ Transistor drain-to-source voltage
$V_{gs}$ Transistor gate-to-source voltage
$V_{knee}$ Transistor knee voltage
$X_{Cds}$ Reactance of transistor effective output capacitance
$R_{opt}$ Class-B output power optimal load resistance

Abbreviations

ACLR Adjacent Channel Leakage Ratio
ALM Active Load Modulation
AM/PM Amplitude Modulation / Phase Modulation
CMOS Complementary MetalOxideSemiconductor
$CO_2e$ Carbon-dioxide equivalents
CW Continuous Wave
DC Direct Current
DLM  Dynamic Load Modulation
DPD  Digital Pre-Distortion
DSM  Dynamic Supply Modulation
DUT  Device Under Test
EER  Envelope Elimination and Restoration
EM   Electro-Magnetic
ET   Envelope Tracking
FET  Field-Effect Transistor
GaAs Gallium-Arsenide
GaN  Gallium-Nitride
GSM  Global System for Mobile communications
HBT  Heterojunction Bipolar Transistor
HEMT High Electron Mobility Transistor
HWR  Half-Wave Rectified (sinusoidal)
IM   Intermodulation
IPBO Input Power Back-Off
LDMOS (Silicon) Laterally Diffused Metal Oxide Semiconductor
LINC Linear amplification using Nonlinear Components
LSNA Large Signal Network Analyser
LTE  Long Term Evolution
LTE-A Long Term Evolution Advanced
MEMS Micro-Electro-Mechanical Systems
MIMO Multiple Input Multiple Output
MMIC Monolithic Microwave Integrated Circuit
MN   Matching Network
OPBO Output Power Back-Off
PA   Power Amplifier
PAE  Power Added Efficiency
PAPR Peak to Average Power Ratio
PCB  Printed Circuit Board
PDF  Probability Density Function
PWM  Pulse Width Modulation
PSD  Power Spectral Density
QAM  Quadrature Amplitude Modulation
Q-factor Quality factor
RF   Radio Frequency
SEM  Scanning Electron Microscope
SiC  Silicon-Carbide
TLM  Transmission/Transfer Line/Length Method
VCO  Voltage Controlled Oscillator
VNA  Vector Network Analyser
WCDMA Wideband Code Division Multiple Access
WiMAX Worldwide Interoperability for Microwave Access
# Contents

Abstract iii

List of Publications v

Notations and abbreviations vii

## 1 Introduction

1.1 Motivation .................................. 1
1.2 Thesis contributions ........................ 2
1.3 Thesis outline ............................ 4

## 2 Power amplifier principles and wideband design

2.1 Transistor model .......................... 5
2.2 PA modes of operation ....................... 8
  2.2.1 Class-B theory ........................ 8
  2.2.2 Harmonic resonators and tuning .......... 11
2.3 Transistor technology ....................... 11
  2.3.1 GaN HEMTs ........................ 11
  2.3.2 Comparing technologies ................. 12
2.4 Wideband PA design ........................ 12
  2.4.1 Comparing techniques ................. 13
  2.4.2 Resistive harmonic loading .......... 15

## 3 High efficiency transmitters

3.1 Efficiency enhancement principles ............ 20
3.2 Dynamic supply modulation ................... 21
  3.2.1 Linear amplification using nonlinear components 23
3.3 Dynamic load modulation .................... 24
  3.3.1 Class-J varactor-based DLM .......... 26
3.4 Active load modulation ..................... 31
  3.4.1 Doherty .......................... 33
  3.4.2 Chireix .......................... 34
  3.4.3 Dual-RF input Doherty-outphasing continuum 36

## 4 SiC varactors for dynamic load modulation

4.1 Microwave power varactors .................. 43
4.2 Epitaxial design .......................... 46
4.3 Fabrication ............................. 48
Chapter 1

Introduction

1.1 Motivation

The use of wireless technology is changing the lives of people all around the world. Proven during the Arab spring, mobile phones are a tool of democracy, allowing people to organize themselves, while documenting and distributing their uncensored reality to a global audience. In rural areas, not the least in developing countries, wireless infrastructure can enable communications where terrain or long distances would otherwise make fixed solutions prohibitively expensive. In the industrialized world high speed mobile internet access allows for constant connectivity, providing flexibility both in how we work and how we play.

Since a couple of years the global number of mobile phone subscriptions exceeds the human population [1]. Further, as shown in Fig. 1.1, the data traffic in the mobile networks increased by more than an order of magnitude over the period 2008–2012. This trend is expected to continue at least over the next four year period. Requiring significant bandwidth, more than 50% of the mobile traffic in 2012, was video data [2]. No doubt the rapid consumer adoption of smartphones and mobile broadband is the driving force behind this development.

There appear to be at least three key enablers to increased capacity in future mobile networks. Firstly, an increase in the number of base stations. Specifically incorporation of smaller pico-/femto-cell base stations and WiFi access points to locally serve users and offload the macro base stations. Secondly, increased spectrum allocation for mobile communication, either for accommodation of a larger number of users or for carrier aggregation to increase per-user bandwidth. Lastly, increased spectral efficiency by employing MIMO type architectures to increase capacity by utilizing orthogonalities offered by multiple antennas. The product of the above listed capacity improvements potentially implies a thousandfold data rate increase by 2020 [7].

Customer demand is a moving target and deployment of higher capacity networks is ongoing. Today 4G connections only represent 0.9% of all mobile connections, but account for 14% of all data traffic [2]. However, Fig. 1.1 shows that 2012 was the first year that actual data traffic was lower than forecasted in the previous year. This is partly due to operators throttling
connection speeds and eliminating unlimited data plans to maintain an even quality of service [2]. Problematically, as data rates have been increasing, operator revenue per user has slowly been decreasing [1, 7]. Effectively, for a sustainable business model, the cost per bit must decrease at the same rate as the capacity increases. Unlike the operator fragmented spectrum allocation of today, future infrastructure therefore needs to more intelligently pool resources to meet capacity targets in a cost effective way [7].

In a breakdown of mobile network operational costs, base station electricity consumption is reported to be the largest post with an equally large share of total network carbon footprint in CO$_2$e [1]. The transmitter power amplifier (PA) consumes a significant part of the total base station power, with numbers reported in the 40% range [8]. Power consumption of supporting functions, e.g. DC power supplies and cooling systems, scale in relation to the PA inefficiency and can combined claim a similar power share. Improvements in PA energy efficiency are therefore important economically, ecologically, and indirectly in terms of network capacity and coverage.

### 1.2 Thesis contributions

This work has focused on the development of theory and technology for design and realization of energy efficient PAs with wideband performance. Improvements in energy efficiency targets network operational costs and carbon footprint. Resulting reduction in DC supply and cooling requirements reduces the base station form-factor and cost. This enables simpler and less intrusive deployment of base stations to create denser networks. Wideband PAs allows for frequency agility and simplifies the technology platform required to cover the growing mobile spectrum allocations at reduced cost.

Most PAs require specific reactive conditions at the higher harmonics in order to achieve high efficiency operation. This introduces difficulties to implement PAs with more than an octave of bandwidth. In [A] resistive harmonic terminations are employed to realize a decade bandwidth GaN PA. Compared to other modes of operation, this resistive mode of operation offers an ap-
1.2. THESIS CONTRIBUTIONS

pealing trade-off between efficiency and bandwidth. Used in conjunction with envelope tracking (ET) this type of wideband PA could be relevant in modern multi-band base stations.

One of the main energy issues in PAs for base stations is the high peak-to-average power ratio (PAPR) of modern spectrally efficient communication signals. This forces the amplifier to operate in output power back-off (OPBO) most of the time, where the efficiency typically is low. Varactor-based dynamic load modulation (DLM) is an interesting concept for OPBO efficiency enhancement. A complicating factor in this type of solution is the interplay between two different semiconductor devices, i.e. transistor and varactor. DLM PA design has therefore tended to be empirical efforts based on transistor load-pull characterization, limiting fundamental understanding of the architecture potential. In [B] a theory for DLM based on class-J mode of operation is presented. This theory enables concrete analysis of varactor technology requirements and evaluation of transistor technologies for DLM. Further, it is shown how varactor-based DLM can simultaneously offer both OPBO efficiency enhancement and potentially octave bandwidth frequency reconfigurability.

Future wireless networks will require a greater diversity in base station power levels, from macro- to femto-cells. However, not all OPBO efficiency enhancement concepts are easily scaled in output power. In [C] the class-J DLM theory is applied to scale up the 10-W GaN demonstrator in [B] to realize a packaged small form-factor 86-W PA. This power level is unprecedentedly high, surpassing other published varactor-based DLM PAs by nearly an order of magnitude.

Another compelling OPBO efficiency enhancement solution is offered by dual RF-input PA architectures. The availability of two independent inputs allows for significant improvements in efficiency compared to single-input implementations, while providing additional degrees of freedom for post-production tuning and linearization. A general approach to the analysis of dual-input PAs is presented in [D]. It is further shown that suitable choice of combiner topology together with optimum input control can enable OPBO efficiency enhancement over 100% fractional bandwidth. Proven by a corresponding GaN PA demonstrator design, this also holds true after discarding common theoretical assumptions of short-circuited higher harmonics and including microwave transistor parasitics.

All DLM PA designs in this work use SiC varactors specifically developed for the purpose. A non-abrupt varactor C(V) is realized by tailored epitaxial design to support significant voltage swing at high tuning factors. Material design, device layout, fabrication and characterization are covered in [E]. Factors constraining varactor performance, relating to parasitic effects, are explored by pushing the limits of material, lithography and layout in [F].

There are interesting prospects to use varactors capable of handling high power levels in frequency agile circuits. In [G] advanced measurements in the form of active multi-harmonic source- and load-pull are used to study varactor nonlinear behavior. A method of extracting varactor Q-factor in the presence of nonlinear distortion is presented. It is also shown how the loading at higher harmonics affect varactor performance and that nonlinear considerations should be made in varactor-based circuit design.

In summary, this thesis has addressed, proposed and demonstrated solu-
tions relevant to the improvement of capacity and energy efficiency in wireless networks at required lower costs.

1.3 Thesis outline

This thesis is divided into four main parts. In Chapter 2 basic transistor and PA operation is covered. Benchmarking of transistor technologies and an overview of wideband PA design follow, with special focus on the resistive mode of operation in [A]. In Chapter 3 an overview of architectures for transmitter efficiency enhancement is presented. This includes varactor-based DLM for both frequency reconfigurability and high power PA realizations [B,C], together with active load modulation by wideband dual-RF input PAs [D]. In Chapter 4 epitaxial design, device layout, fabrication, and characterization of the SiC varactors intended for DLM applications are covered [E-G]. The thesis is concluded in Chapter 5 with final words on directions for future work.
Chapter 2

Power amplifier principles
and wideband design

This chapter presents an overview of PA operational and design principles. The simplified transistor model used for theoretical analysis throughout the thesis work is introduced. The GaN HEMT, being the transistor technology used in [A-D], is presented and compared to other technologies. Particular emphasis is put on the \( X_{C_{ds}}/R_{opt} \) ratio as a key transistor parameter. A review of wideband PA design follows, with continuous modes and specifically the resistive harmonic terminations in [A] covered.

2.1 Transistor model

Successful PA design is enabled by appropriate use of transistor models. A complex model that accurately reproduces the transistor behavior is useful to predict detailed PA performance, e.g. matching, stability, linearity, gain, efficiency and output power. However, a complex model makes quantitative analysis difficult, for instance when simultaneously optimizing a PA for efficiency and wideband performance. For this reason a simplified transistor model, largely based on the treatment in [9], is adopted in [A-D] and described in this section.

DC output characteristics for a GaN HEMT are shown in Fig. 2.1 (a) and a simplified current model in Fig. 2.1 (b). The transistor output current in the saturated regime is controlled by the gate voltage \( V_{gs} \). In the simplified model the transistor is reduced to a perfectly linear voltage controlled current source, with constant transconductance and no output conductance. The model is assumed valid as long as \( V_{ds} \) is greater than the transistor knee-voltage \( V_{knee} \), corresponding to the voltage where the maximum current \( I_{max} \) is reached. As an example, for the GaN transistor in Fig. 2.1 (a) a setting of \( V_{knee} = 5 \) V and \( I_{max} = 2.5 \) A is reasonable. Maximum valid \( V_{ds} \) in the model is limited by the breakdown voltage \( V_{br} \), typically in the order of 100–200 V for a pinched GaN HEMT. Self-heating effects, noted by negative slopes at higher currents in the real device, is ignored in the simplified model. This is justified as high efficiency operation implies avoiding these regions of high DC power dissipation.
Chapter 2: Power Amplifier Principles and Wideband Design

Figure 2.1: (a) DC output characteristics for a 3-mm GaN HEMT, $V_{gs}$ in steps of 0.25 V from $-4$ V to 1 V, (b) simplified transistor current model.

Figure 2.2: Nonlinear equivalent circuit topology for a HEMT at microwave frequencies. Extrinsic parasitics depend on the packaging.

Accurate modeling of transistors at microwave frequencies requires consideration of parasitic, reactive and nonlinear elements. More extensive topologies are found in literature [10], but a fairly representative equivalent circuit topology of a microwave HEMT is shown in Fig. 2.2. The transistor current source is only accessible through a set of parasitic resistances. Specifically, transistor on-resistance and knee-voltage are related to $R_d$ and $R_s$. Capacitances due to charge displacements in the device give rise to reactive currents and voltages at high frequencies. Further, as these capacitances are nonlinear, reactive currents are also induced at higher harmonics. The harmonic content of the intrinsic voltage and current source waveforms affect the transistor output power and efficiency. Appropriate modeling of the transistor nonlinearities is therefore needed for accurate prediction of microwave performance. Models often trade accuracy in different regions depending on application, e.g. targeting the saturated [11] or the linear region [12–14]. The importance of various parasitics surrounding the intrinsic transistor greatly depend on the packaging or embedding structures used, e.g. bare-die compared to ceramic package use, or microstrip compared to coplanar embedding. A complete nonlinear Chalmers/Angelov model is compared with a simplified model in [B].

A schematic of a typical single-ended PA is shown in Fig. 2.3. Transistor gate bias and drain supply are provided through the input and load match-
2.1. TRANSISTOR MODEL

Figure 2.3: Schematic of a typical single-ended PA.

Figure 2.4: A 15-W GaN HEMT from Cree, Inc. soldered to a fixture and bonded to gold pads on adjacent PA matching networks [D].

ing networks, respectively. The design of these matching networks affect the PA output power, efficiency and linearity. All PAs designed in [A-D] employ bare-die transistors. That is, the transistor semiconductor chip is directly soldered to the PA fixture and bondwired to adjacent input and load matching networks, as shown in Fig. 2.4. To facilitate bondwiring the matching network substrates are either gold plated [A-C] or have gold pads soldered to them [D]. Packaged transistors are more practical by comparison, allowing for replacement by screwdriver, and are sometimes conveniently pre-matched for certain frequency bands. However, the pre-matching or parasites of packaged devices are not always well documented. Using bare-die transistors reduces the extrinsic parasites to a minimum of bondwire inductance. Moving the circuit design closer to the intrinsic device reference planes allows for understanding and ultimate exploitation of device technology performance. The versatility of bare-die transistors make them a superior choice in research.

Simplified modeling of transistor behavior is desirable to enable efficient theoretical analysis. The simplified transistor model successfully used in [A-D] to predict PA bandwidth, load modulation and efficiency performance, is shown in Fig. 2.5. Biased at pinch-off (class-B) and excited from the gate side by a sinusoidal signal, the transistor output current is fairly well represented by a half-wave rectified (HWR) sinusoidal. The amplitude of this current waveform is given by a scale factor \( \beta = \{0, 1\} \), which is a function of the
PA input power. To include some rudimentary but essential reactive effect in the model, the nonlinear capacitances are lumped together into a single linear and effective $C_{ds}$ output capacitance. The simplified model does not directly consider the input network design, other than assuming ideal output current shaping. However, the value of the effective output capacitance is typically identified by large-signal load-pull and is somewhat affected by the input network design through transistor feedback.

2.2 PA modes of operation

Mode of operation refers to the intrinsic current and voltage waveforms in and across the transistor current source. The resulting mode of operation is a product of the transistor driving conditions and the imposing PA matching network design. After deciding on a suitable transistor model, actual PA design focuses on the matching networks.

Various modes of operation are presented in literature, e.g. class-A, -AB, -B, -C, -F, -J [9]. These modes all have different merits in terms of efficiency, output power, linearity and gain. Required higher harmonic loading conditions also differ, e.g. short-circuited, open-circuited, capacitive. Common to all however, being transconductance modes with HWR-current waveforms at various conduction angles, is the relative ease of realization with a sinusoidal drive. Class-F mode in particular has a theoretical efficiency of 100%, when considering an infinite number of harmonics and $V_{knee} = 0$. Switch-mode class-E and class-F$^{-1}$ also feature ideal efficiencies [15], but with an operational principle assuming binary driving conditions the immediate use in amplitude modulated spectrally efficient signal transmission is limited until deployed in, for instance, RF Pulse-Width Modulation (PWM) or the polar architectures discussed in Chapter 3.

2.2.1 Class-B theory

In class-B mode, for which the simplified model in Section 2.1 is suitable, the current is assumed to be a HWR sinusoidal. The corresponding voltage waveform is assumed perfectly sinusoidal. Intrinsic time-domain waveforms and corresponding load lines (current versus voltage) are shown in Fig. 2.6. These waveforms are utilizing the full voltage and current swing capability of the transistor, which in turn maximizes the output power in this mode.

\[ \beta I_{\text{max}} \times \begin{array}{c} + \\ \cap \\ C_{ds} \end{array} \]
of operation [9]. There is a duality between waveforms and harmonic load impedances, where Fourier decomposition of the assumed current (going out of the transistor) and assumed voltage gives the corresponding harmonic load impedances. The fundamental load impedance to maximize the output power, for a drain supply voltage $V_{DC}$, is given by,

$$R_{opt} = \frac{V_{DC} - V_{knee}}{I_{max}/2},$$  \hspace{1cm} (2.1)

and $V_{DC} = (V_{br} + V_{knee})/2$ if to utilize the full transistor voltage swing capability and not enter breakdown. The load impedances at all higher harmonics are zero (short-circuited), which is logical if the output voltage is to be sinusoidal. Provided the transistor produces a HWR output current, these are the intrinsic impedances the load network must present to the current source through the transistor parasitics.

A first estimation of transistor $R_{opt}$ and effective $C_{ds}$ for matching network design is possible from (2.1) and small-signal capacitance extraction, respectively. However, real devices [16] and their models [17] exhibit temperature and dispersion effects when operating in the large-signal regime. For bare-die transistors with limited package parasitics, identification of $R_{opt}$ and effective $C_{ds}$ is made from fundamental load-pull characterization with short-circuited higher harmonics [A-D]. This approach is explained in Fig. 2.7. For added confidence the procedure is repeated at multiple fundamental frequencies.

Provided proper load network design to intrinsically produce ideal harmonic conditions, the class-B output power is given by,

$$P_{out} = \frac{\beta^2 I_{max} (V_{DC} - V_{knee})}{4},$$  \hspace{1cm} (2.2)

with a corresponding DC power consumption equal to,

$$P_{DC} = \frac{\beta I_{max} V_{DC}}{\pi}.$$  \hspace{1cm} (2.3)

This allows for calculation of the efficiency $\eta$ at which DC-power is converted into RF-power,

---

**Figure 2.6:** Left: Class-B mode voltage and current waveforms in time-domain. Right: equivalent load line.
Figure 2.7: Identifying transistor $R_{\text{opt}}$ and $C_{ds}$ from load-pull efficiency contours: adding a shunt capacitance $C_{\text{add}} = -C_{ds}$ at the transistor output moves the peak efficiency contour to $R_{\text{opt}}$ on the Smith chart real-axis (red line indicating the transformation).

Figure 2.8: Class-B efficiency versus output power back-off, with the PDF of a 6.7 dB PAPR WCDMA signal for reference.

$$
\eta = \frac{P_{\text{out}}}{P_{\text{DC}}} = \beta \frac{\pi}{4} \left( V_{\text{DC}} - V_{\text{knee}} \right). \quad (2.4)
$$

The ideal efficiency of class-B ($V_{\text{knee}} = 0$) is calculated to 78.5% at maximum output power ($\beta = 1$). However, when the output power is backed-off the efficiency decreases and is halved at 6 dB OPBO, as shown in Fig. 2.8. The probability density function (PDF) of a WCDMA signal is shown in the figure for reference. Average efficiency $\eta_{\text{avg}}$ for a given signal PDF function $p$ and PA efficiency characteristic $\eta(P_{\text{out}})$ can be calculated as,

$$
\eta_{\text{avg}} = \frac{\int_{0}^{P_{\text{max}}} P_{\text{out}} \cdot p(P_{\text{out}}) \, dP_{\text{out}}}{\int_{0}^{P_{\text{max}}} \eta(P_{\text{out}}) \cdot p(P_{\text{out}}) \, dP_{\text{out}}}. \quad (2.5)
$$
Dropping efficiency and high probability of low output power gives $\eta_{avg} = 25\%$ for the two characteristics in Fig. 2.8. This highlights the need for enhanced OPBO efficiencies, as modern base stations commonly handle signal PAPRs in the 6–12 dB range. This problem must be addressed at an architecture level as discussed in Chapter 3 and is not solvable by any mode of operation.

### 2.2.2 Harmonic resonators and tuning

The reactive conditions required in PA load network design, e.g. the short-circuited higher harmonics in class-B, are realized by resonators. At microwave frequencies in particular, the use of a shunt quarter-wave stub at suitable electrical length from the transistor output is common to present the required harmonic impedance. Multiple resonators may be cascaded to handle a finite number of harmonics, followed by a fundamental match. In a trade-off between losses, physical size and diminishing returns due to modeling inaccuracy, more than three harmonics are rarely handled in practice. For instance, a GaN class-F PA with two harmonic resonators is reported with a PAE of 85% at 2 GHz [18].

The mode of operation concept, although useful, is an idealization. Non-ideal transistor behavior make it impossible to reproduce exact waveforms of a mode. For this reason the quest for high efficiency in narrowband PAs is often referred to as harmonic tuning. The efficiency is optimized by multi-harmonic source- and load-pull simulations or measurements, with the objective of minimizing intrinsic transistor current and voltage waveform overlap. To achieve the highest possible efficiency, due to transistor feedback, this typically becomes an iterative procedure between harmonic source- and load-pull. The PA input and load matching networks are finally designed to realize the identified optimum impedances without consideration to any particular mode of operation. This type of load-pull based design methodology is equally suitable when using packaged devices. Harmonically tuned GaN PAs are reported with 80% and 70% PAE at 3.5 and 5.5 GHz, respectively [19].

### 2.3 Transistor technology

The PA designs in [A-D] exclusively use GaN HEMTs due the excellent power and high frequency performance offered. Despite the allure, GaN is still an emerging technology with fundamental reliability and operational problems related to dispersion [20–25]. It is therefore relevant to review the GaN HEMT together with other transistor technologies to quantify transistor performance in terms relevant to the theoretical work presented in this thesis.

#### 2.3.1 GaN HEMTs

The AlGaN/GaN heterostructure and the corresponding GaN HEMT have been under development for approximately 20 years [26,27]. This heterostructure of wide bandgap semiconductors forms a two-dimensional electron gas with high mobility, high saturation velocity and high carrier density without the need for intentional impurity doping. The combination of good electron
transport capabilities and high breakdown field in GaN compared to common semiconductors, enables the realization of high voltage, high power, high frequency transistors. Power densities as high as 40 W/mm have been reported [28]. Significant research is also being invested in GaN for power generation at millimeter frequencies. Recently reported $f_{\text{max}}$ of 400 GHz shows how GaN may play a part in bridging the THz-gap [29]. For these reasons the GaN HEMT is an attractive technology for both military and civil wireless applications [30]. The commendable review of GaN HEMT and MMIC technology in [31] serves as a useful introduction. There are already commercial suppliers of high power GaN HEMT devices for low-GHz mobile communication, where [A,D] and [B,C] use devices from Cree, Inc. (USA) and Mitsubishi Electric, Co. (Japan), respectively.

2.3.2 Comparing technologies

Typical parameters of various transistor technologies are summarized in Table 2.1. At top of the table the GaN HEMT is seen to boast the highest $W/mm$ power density. A high power density is advantageous as the physical transistor size is smaller and suffers less from distributed effects for a given output power level. The high $R_{opt}$ impedance level of GaN, compared to the GaAs technologies, allows for easier realization of both hybrid and MMIC high power PAs. A high $R_{opt}$ is also offered by LDMOS technology, however, at the price of much higher output capacitance. Realization of reactively matched wideband PAs, e.g. class-A/B and the resistively terminated PA in [A], relies on matching of the transistor output capacitance. Shown by Fano, the fractional bandwidth $\frac{\Delta \omega}{\omega_0}$ over which a match $|\Gamma|$ can be provided is fundamentally limited [32]. Applied to PA load matching, this gives,

$$\frac{\Delta \omega}{\omega_0} \ln \frac{1}{|\Gamma|} \leq \pi \frac{X_{Cds}}{R_{opt}}. \quad (2.6)$$

The $X_{Cds}/R_{opt}$ ratio, which is independent of device periphery, therefore captures the bandwidth potential of the transistor technology (a higher ratio is better). In [B] it is shown that $X_{Cds}/R_{opt}$ limits the maximum frequency to which efficient class-J DLM can be expected. Likewise, the bandwidth and OPBO efficiency performance of dual-input PAs is closely related to $X_{Cds}/R_{opt}$ [D]. In Table 2.1 the bandwidth potential of high-power GaN is seen comparable to the low-power GaAs technologies, while the reliable and robust LDMOS, the workhorse in today’s narrowband base stations, is far behind. The combination of high $R_{opt}$ and high $X_{Cds}/R_{opt}$ makes GaN exceptional. With expanding base station spectrum allocations and multi-standard requirements the pursuit of mature GaN technology for wideband transmitters is understandable.

2.4 Wideband PA design

Significant research efforts are invested in development of wideband PAs given the application flexibility they offer, e.g. for use in software defined radio, electronic warfare, multi-functional radar and mobile communication. For
Table 2.1: Transistor technology parameters [9], [D].

<table>
<thead>
<tr>
<th>Transistor Technology</th>
<th>$P_{\text{out}}$ (W/mm)</th>
<th>$V_{\text{DC}}$ (V)</th>
<th>$I_{\text{max}}$ (mA/mm)</th>
<th>$C_{ds}$ (pF/mm)</th>
<th>$R_{\text{opt}}$ (Ω·W)</th>
<th>$X_{C_{ds}}/R_{\text{opt}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaN HEMT</td>
<td>5.0</td>
<td>30</td>
<td>800</td>
<td>0.3</td>
<td>375</td>
<td>7</td>
</tr>
<tr>
<td>Si LDMOS</td>
<td>1.4</td>
<td>28</td>
<td>200</td>
<td>1.0</td>
<td>392</td>
<td>0.6</td>
</tr>
<tr>
<td>GaAs pHEMT</td>
<td>0.9</td>
<td>10</td>
<td>350</td>
<td>0.25</td>
<td>51</td>
<td>11</td>
</tr>
<tr>
<td>GaAs HBT</td>
<td>0.2</td>
<td>3</td>
<td>250</td>
<td>0.5</td>
<td>5</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 2.2: Wideband 10-W GaN PAs sorted by BW

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Freq. (GHz)</th>
<th>BW (%)</th>
<th>Gain (dB)</th>
<th>$P_{\text{out}}$ (dBm)</th>
<th>$\eta$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[43] (2011)</td>
<td>2.15–2.65</td>
<td>21</td>
<td>11–12</td>
<td>40.4–41.8</td>
<td>65–72</td>
</tr>
<tr>
<td>[44] (2013)</td>
<td>2.0–4.0</td>
<td>67</td>
<td>12–14</td>
<td>40</td>
<td>40–57</td>
</tr>
<tr>
<td>[45] (2011)</td>
<td>0.55–1.1</td>
<td>67</td>
<td>9.5–12</td>
<td>40</td>
<td>65–80</td>
</tr>
<tr>
<td>[46] (2010)</td>
<td>1.9–4.3</td>
<td>77</td>
<td>9–11</td>
<td>40–41.8</td>
<td>57–72</td>
</tr>
<tr>
<td>[47] (2011)</td>
<td>0.9–2.2</td>
<td>84</td>
<td>10–13</td>
<td>40–43</td>
<td>63–89</td>
</tr>
<tr>
<td>[48] (2012)</td>
<td>1.3–3.3</td>
<td>87</td>
<td>10–14</td>
<td>40</td>
<td>60–83</td>
</tr>
<tr>
<td>[49] (2012)</td>
<td>0.5–2.0</td>
<td>120</td>
<td>11–16</td>
<td>38–41</td>
<td>42–68</td>
</tr>
<tr>
<td>[50] (2011)</td>
<td>0.6–2.4</td>
<td>120</td>
<td>N/A</td>
<td>40–42</td>
<td>57–75</td>
</tr>
<tr>
<td>[51] (2009)</td>
<td>0.5–2.5</td>
<td>133</td>
<td>15</td>
<td>39.5–41.3</td>
<td>45–63</td>
</tr>
<tr>
<td>[A] (2011)</td>
<td>0.4–4.1</td>
<td>164</td>
<td>10–15</td>
<td>40–42</td>
<td>40–62</td>
</tr>
<tr>
<td>[52] (2009)</td>
<td>0.35–8.0</td>
<td>183</td>
<td>8–10</td>
<td>38–40</td>
<td>22–37</td>
</tr>
</tbody>
</table>

Energy sensitive applications with high PAPR signals, wideband PAs with high peak power efficiencies are attractive for deployment in OPBO efficiency enhancement architectures such as ET or polar (discussed in Chapter 3).

2.4.1 Comparing techniques

There are various approaches to wideband design. Distributed amplifiers can for instance circumvent the Fano-limit (2.6) and offer multi-decade performance, but tend to suffer from low transistor power utilization and low efficiencies due to inefficient power distribution and high quiescent power consumption [33–36]. Another idea [37] is to use electronically tunable components to enable wideband frequency reconfigurable matching and higher harmonic control [38–40]. There are successful reconfigurable designs reported both in the UHF [41] and the X-band [42]. The work in this thesis has focused on moderate bandwidths and higher efficiencies, a domain where reactive matching is a suitable approach. Table 2.2 lists a subset of recently published wideband GaN HEMT PAs. The PAs selected for the survey are in the 10-W and 1–2 GHz range, resulting in similar impedance levels and a more fair comparison.

The most wideband PAs in Table 2.2, [51,52] and [A], employ Fano-limited reactive matching. In [52] the transistor is biased in class-A for high linearity. Load-pull is used on a packaged device to identify fundamental matching ver-
sus frequency. Resulting efficiency is low, given class-A operation, but higher harmonics need not be considered in the wideband design. The deep class-AB MMIC design in [51] employs a lossy input match to flatten the gain. A near 50 $\Omega$ output impedance facilitates wideband performance, while no consideration is made to the termination of higher harmonics. In [A], as covered in Section 2.4.2, class-B gate bias and higher than class-B drain voltage enables a wideband resistive mode of operation. The corresponding load network design in [A] is simple and automatically handles higher harmonics, giving higher than class-A efficiency.

As mentioned in Section 2.2, high efficiency modes require rather specific higher harmonic impedances. This inherently limits the efficiency bandwidth of the PA due to the narrowband response of typical harmonic resonators. However, the relatively recent presentations of continuous modes has changed this outlook. Starting with [53], it is shown that class-B and class-J are part of the same waveform family. Transition from one mode to the other, by smoothly and appropriately changing harmonic impedances, is possible while maintaining high efficiency operation. The PA load network should therefore closely reproduce these continuous impedance trajectories versus frequency for wideband high efficiency operation. Presentations of other continuous modes have followed, e.g. class-E [54], class-F [55], class-F3 (class-F with 3rd harmonic freedom) [56], and class-F/J/F$^{-1}$ with reactive [57] and lossy 2nd harmonic [58].

Peak voltage swings within some modes approach four times the DC voltage, compared to two times in class-B. GaN HEMTs, with breakdown several times higher than typical operating voltages, are therefore highly suitable for continuous mode implementation. In [49] continuous class-J is implemented by brute-force optimization of load network topology parameters while catering for package parasitics. Two implementations of continuous class-F are reported in [43,45]. Although not recognized as a continuous implementation, the work in [47] is an optimization of wideband class-E performance. An engineered transition from one continuous mode (class-F) to another (class-F$^{-1}$) is used to extend the efficiency bandwidth in [48]. In [50] efficiency degradation in the transition from class-B to class-J mode is mitigated by active 2nd harmonic injection.

Continuous modes are of course, outside the environment of a load-pull measurement system [53], impossible to perfectly reproduce in practice. In general the modes only offer impedance design guidelines and compromises in terms of output power and efficiency must be made. Sensitivity analysis is important in wideband design, as it is typically more important to avoid the minima than to hit the maxima over the bandwidth. Empirical wideband design based on load-pull can therefore be successful [44, 46]. In part such success may be explained by the presence of continuous modes, relaxing the strict impedance conditions known from single mode theory, allowing for more circuit design freedom.

High efficiency modes of operation typically require highly reactive higher harmonic terminations. This poses an obvious bandwidth barrier, as the transistor is prevented from delivering power at these frequencies by the load matching network. With the use of reactive second harmonic impedances the bandwidth in most PAs is therefore constrained to less than one octave. This limit is extendable by push-pull PAs, common at lower frequencies [9],
2.4. WIDEBAND PA DESIGN

provided availability of high frequency wideband baluns with appropriate odd-mode impedances [59]. A decade bandwidth GaN push-pull PA with $\eta > 45\%$ was recently published [60].

2.4.2 Resistive harmonic loading

Presented in [A], results comparable to the best push-pull implementation are still achievable, in a willing efficiency compromise, by reactive matching. Consider the waveforms in Fig. 2.9, i.e. a HWR current and inverse-HWR voltage waveform. Here $V_{DC}$ is increased by a factor of $(1 - 1/\pi)/2 = 36\%$ relative to $V_{knee}$ for the same output power and peak voltage swing as class-B. The overlap between current and voltage gives an efficiency of $\eta = \pi^2/(8(\pi - 1)) = 58\%$. Fourier decomposition of the waveforms shows that all harmonics are equally loaded resistively by $R_{opt}$. This frequency independent loading condition is convenient. Similar to a class-A PA the load network simply needs to provide wideband matching of the output capacitance. Any harmonics falling within the load network bandwidth are thereby properly loaded by default. Although the 58\% efficiency is significantly lower than in class-B (78.5\%), it is still higher than in class-A (50\%). Further, problems of high quiescent power consumption and rapid efficiency roll-off versus OPBO are reduced by class-B compared to class-A gate bias. The price to pay is class-B power gain, which is 6 dB lower compared to class-A.

Based on this principle, a GaN HEMT demonstrator is designed in [A]. The output capacitance is resonated by a simple transmission line structure together with an RF-grounded drain supply inductor. Given a bare-die implementation, limited consideration of the drain bondwire inductance transformation is required. A wideband stepped-impedance Chebychev transformer is used to implement a wideband load resistance matched to 50 $\Omega$. It is often found in load-pull that an impedance $R_{pow}$, lower than the efficiency optimal $R_{opt}$, will provide higher output power (at reduced efficiency) [9]. Given that the wideband transformer inevitably will ripple due to the Fano-limit, it is here designed to ripple between $R_{opt}$ and the identified $R_{pow}$. Measured load network matching is shown in Fig. 2.10 (left).

Figure 2.9: Resistive mode of operation in [A]. Left: Voltage and current waveforms in time-domain when all harmonics are loaded by $R_{opt}$. Right: equivalent load line.
CHAPTER 2. POWER AMPLIFIER PRINCIPLES AND WIDEBAND DESIGN

Figure 2.10: Measured S-parameters of PA matching circuits over 0.4–4 GHz. Left: load network matching. Right: input network matching with the 55% PAE source-pull contours (PA in compression).

Figure 2.11: Static measurement results versus frequency and output power back-off.

The input matching is based on source-pull of the transistor with designed load network. A simple stepped-impedance transformer is enough to hit the overlapping high efficiency source-pull regions at different frequencies, as shown in Fig. 2.10 (right). The PA is made conditionally stable around 50 Ω, by the addition of a resistor in the gate feed.

Measurements of the PA were done using an LSNA to capture harmonic content at input and output. Below 600 MHz, due to LSNA frequency limitations, measurements were complemented by standard powermeter measurements. Results versus frequency are shown in Fig. 2.11. The PA delivers >10 W of output power over 0.4–4.1 GHz at 2–3 dB compression with drain efficiencies of 40–62%, for a decade of bandwidth. This compares well with the decade push-pull result in [60] and, in terms of an efficiency-bandwidth trade-off, the results fit nicely into Table 2.2. Given a limited stability effort and non-lossy input match, the PA return loss is 2–6 dB over the bandwidth.

Measured harmonic power levels in Fig. 2.12 show the signs of a resistive mode of operation, with the second harmonic approaching a relatively high -10 dBc at low frequencies. The general trend is decreasing harmonic power levels versus frequency. From 2–4 GHz the PA mostly relies on a decent fundamental match and arbitrary modes of operation, as the second harmonic
falls outside the load network bandwidth.

Modulated measurements on a few frequencies (0.9–3.5 GHz) within the PA bandwidth give a worst case adjacent channel leakage ratio (ACLR) of -31 dBC. The raw linearity of the class-B biased PA is not sufficient to pass the tough spectral requirements of wireless communications. After linearization by digital pre-distortion (DPD) the worst case ACLR is improved to -46 dBC, sufficient to pass the transmission standards. The output spectra before and after linearization are shown in Fig. 2.13. Signal characteristics with 7.0 and 8.5 dB PAPR result in an average PAE of 29% and 23%, respectively.

After introducing GaN HEMT technology, simplified transistor modeling, and the mode of operation concept, the wideband high efficiency results of the resistively loaded PA can be explained. This type of PA may be used with efficiency enhancement architectures (discussed in Chapter 3) to improve OPBO and average transmitter efficiency.
This chapter focuses on the realization of high efficiency transmitters. In this report, transmitter refers to the deployment of one or several PAs to amplify signals while simultaneously meeting linearity requirements, e.g. those in base stations. High transmitter efficiency implies the use of PA OPBO efficiency enhancement techniques. A large plethora of transmitter solutions, to a large extent enabled by digital processing power, for simultaneously high linearity and efficiency are proposed in literature. Recent state-of-the-art for various transmitter architectures report rather comparable efficiency performance, e.g. 40–50% PAE or overall efficiency for common 8–10 dB PAPR signals with 20–50 MHz bandwidth.

Compared to the simple linear PAs described in the previous chapter, increased transmitter complexity is typically motivated by added value, e.g. in terms of better efficiencies, linearity or wideband performance. However, of these competitive advantages only linearity is government regulated, making it non-negotiable. Linearity requirements are different depending on the application, being different in low power handsets compared to high power base stations and varying with transmitted signal standard, e.g. GSM ($-70$ dBc ACLR), WCDMA ($-45$ dBc ACLR), WiMAX ($-30$ dBc ACLR) in a base station. The choice of architecture for an application is based on system level analysis. For instance by comparing hardware requirements for phase-only modulation in LINC, quadrature IQ-modulation plus baseband control in varactor-based DLM [B,C], and double quadrature IQ-modulation in a dual-RF input PA [D]. A complex architecture requiring power hungry DPD, not yet salvaged by Moore’s law, may yield worse system performance than a simpler implementation. Trade-offs in hardware complexity and cost, in combination with varying linearity requirements, somewhat explains and motivates the diverse transmitter solutions proposed in literature.

Lately, in line with base station development, there is a pursuit of wideband efficiency enhanced transmitters. With the inclusion of carrier aggregation in LTE-A, concurrent multi-band operation, resulting in a confusing mix of frequency and time-domain concepts, is a step beyond typical transmitter architecture analysis. There are successful demonstrations of concurrent operation, e.g. in ET [61] and multi-band Doherty [62] transmitters. However, it is for example shown that proper multi-band Doherty design must
consider the combiner behavior also at the inter-band modulation frequencies [a]. In this thesis and most of literature, wideband therefore refers to CW performance, i.e. single narrowband transmission enabled over a wide carrier frequency range. Support for concurrent operation is not implied by default. The 2% instantaneous bandwidth 2–3 GHz piezo reconfigurable PA in [63], varactor and MEMS band reconfigurable Doherty PAs [64,65], and frequency reconfigurable DLM [66] are examples of this.

The survey of transmitter architectures in this chapter inevitably excludes some interesting concepts, e.g. combined ET and Doherty [67], all-digital polar [68], class-S [69,70], burst-mode [71] and pulsed load modulation [72]. Nevertheless, the most common architectures for efficiency enhancement, with basic principle of operation and literature review of recent research, are presented in the next few sections. More in-depth descriptions of varactor-based class-J DLM [B,C] and dual-RF input PAs [D] are also given.

### 3.1 Efficiency enhancement principles

Average PA efficiency under bandwidth-limited signal transmission can be estimated from the signal PDF by (2.5). Given that a base station PA is operating at 6 dB or deeper OPBO most of the time, where typical efficiencies are low, efficiency enhancements at reduced output powers are most essential to improve average transmitter efficiency.

The OPBO inefficiency of a class-B PA is illustrated by load lines in Fig. 3.1. At 6 dB IPBO the transistor gate voltage swing is halved, resulting in a corresponding halving of the output current compared to peak power operation. Given that the PA load resistance is fixed the output voltage swing is also halved, resulting in 6 dB OPBO and the expected linear class-B behavior. Inspecting the 6 dB OPBO load line, the reduced efficiency is attributed to an unnecessarily high supply voltage under the reduced voltage swing condition. Alternatively, the load resistance at this current level is too low, causing an under-utilization of the device voltage swing capability at the particular supply voltage.

It may be claimed that all published OPBO efficiency enhancement solutions are enabled by either or a combination of the dynamic supply modulation
3.2 Dynamic supply modulation

The basic DSM architecture is shown in Fig. 3.3. One of the main strengths of DSM is the carrier frequency independent nature of the concept. A wideband PA caters to the RF bandwidth while the efficiency enhancement is offloaded to a dynamic supply in an elegant separation of OPBO efficiency enhancement from PA operation. Ideally, design of wideband PAs with high peak efficiency, e.g. those based on continuous modes, combined with DSM is sufficient to realize wideband high efficiency transmitters. In reality there are many practical problems associated with DSM, making it an active research area and a heavily patented field [9].

The terminology used throughout literature tends to be inconsistent, with similar and dissimilar DSM architectures referred to as ET, EER and polar

(DSM) and dynamic load modulation (DLM) principles shown in Fig. 3.2. Under ideal conditions, if the supply voltage is changed so the voltage swing grazes the knee voltage the transistor efficiency can be improved without affecting the output power (Fig. 3.2 (a)). In a similar manner, if instead the load resistance is changed to achieve the same grazing effect the efficiency is recovered by a maximization of the voltage swing (Fig. 3.2 (b)). In this later case the output power is not kept constant without simultaneous coordination of load and output current control. Nevertheless, provided that the supply or load can be dynamically modulated with the output current at speeds comparable to the transmitted signal envelope (typical bandwidths in the 10–100 MHz-range), high PA and transmitter efficiencies can be maintained over significant OPBO dynamic range.

Examples of architectures encompassing DSM include ET and polar with continuously tracking supplies, and LINC popularly equipped with switched supplies. Load modulation architectures, the focus of this thesis work, encompass both varactor-based DLM and active load modulation (ALM), e.g. Doherty and Chireix. A review and basic operational principles of DSM, DLM, and ALM architectures are covered in the next few sections.

Figure 3.2: Output power back-off efficiency enhancement principles, (a) supply modulation, (b) load modulation.
transmitters. An effort to clear up these definitions is found in [73] and adopted here. In ET the transistor works like a current source, i.e. the way DSM is illustrated by a class-B PA in 3.2 (a). Changes in the supply level in ET has limited effect of the PA linearity and output power. In polar the transistor works like a voltage source, i.e. typically implying saturated or switched operation [9]. Unlike ET, where the output power is regulated by the input power, the output power in a polar transmitter is instead modulated by the supply with the input power held constant. EER, as originally presented by Kahn in 1952 [74], is a specific application of a polar transmitter with additional signal processing.

Total transmitter efficiency is a product of the supply and PA efficiencies. A typical DC supply is highly efficient. In a DSM transmitter the typical DC supply is replaced by a dynamic supply (or envelope amplifier) which should provide all power required by the PA. This power is not restricted to DC, however, but extends to frequencies several times the transmitted signal bandwidth. The design of the supply immediately translates into another amplifier problem. A review of supply solutions and publications is given in [75] and a clear trade-off between peak power handling and bandwidth (slew-rate) performance is apparent. In literature hybrid supplies with efficiencies in the 70-85% range are common, where typically a high efficiency low frequency switched supply is complemented by a low efficiency high frequency linear amplifier [76–85]. This combination appears to provide the best overall compromise between supply efficiency and bandwidth. Alternative supply solutions include for instance multi-level supplies [86,87].

There are significant differences in the operation of ET and polar architectures. In polar the output power is proportional to the supply voltage, putting stringent requirements on accurate supply control to ensure transmitter linearity. Critically, the polar transmitter bandwidth is limited by the supply bandwidth. This is not the case in ET, where inaccurate tracking can be accepted, as the linearity can be recovered by input power control. For this reason ET appears as the common solution in linear high bandwidth signal transmission. Significant research efforts have been focused on input power and supply control signal shaping to maximize the efficiency in ET under reduced supply slew-rate conditions [78,79,88–94]. Although there are demonstrations of pure polar [95], hybrid ET-polar transmitters, with the addition of input power control at low powers, appear to be more common [81,96,97].

One complication in PA design for DSM are detuning effects arising from
the variable supply operation. A class-AB biased transistor will for instance move towards class-A efficiency when the supply is reduced, due to higher relative quiescent power consumption. Further, transistor output capacitance tends to be highly nonlinear versus supply, which detunes the optimum load matching. In most transistor technologies the output capacitance varies by a ratio of 1:2–1:4 under DSM, e.g. GaN [81,97], CMOS [80], and LDMOS [98]. This effect is reportedly less severe in GaAs HBT technology [99].

Despite the ideal separation of PA and supply, the best ET performance is achieved when supply and PA are co-designed. A supply is optimized differently depending on requirements, e.g. voltage dynamic range, peak power, and switching frequency depending on expected signal PDF [100]. In the same way, due to the detuning effects, the PA can be optimized differently depending on the PDF, both in matching trade-offs [81,97,98,101,102] and by input control shaping [80,99]. Further, the inability to add bandwidth limiting bypass capacitors in the drain bias network, causes stability concerns unless the PA is designed to consider the output impedance of the supply [73].

DSM has been demonstrated in many demanding applications. For instance, 20-W 10 MHz 9.6 dB PAPR WCDMA signal transmission at 2.14 GHz with an overall system PAE exceeding 50% [77]. Proving useful in handsets, with a supply capable of handling battery degradation over time, an overall system PAE of 32.3% for a 10 MHz 25.8 dBm LTE signal at 2.535-GHz is reported in [103]. The technique is indeed power scalable to both Watt-level micro [82] and hundred-Watt macro base stations [86,102]. Signal transmission bandwidths in the 20 MHz range are also reported [78,84,104].

### 3.2.1 Linear amplification using nonlinear components

Linear amplification using nonlinear components (LINC), introduced by Cox in 1974 [105], is not an efficiency enhancement technique. However, lossy combiner LINC produces interesting results when combined with dynamic supply techniques. The Chireix amplifier, described in Section 3.4.2, is a predecessor but an example of LINC using a lossless combiner. While the Chireix is intentionally designed to achieve an ALM effect, most LINC solutions employ lossy combiners to oppositely ensure no unintentional load modulation. Provided there is good branch balance, this allows for linear amplification by outphasing two (identical) amplifiers, i.e. the output power is set by the outphasing angle. Transmission fidelity is, due to the isolation, determined by the quality of the phase modulation, similar to the supply modulation in polar transmitters, and not by amplifier linearity. Any linear or nonlinear components can therefore be used in LINC, hence the name.

Employing LINC outphasing at low powers complemented by linear or auxiliary amplification at peak power can give an interesting compromise between linearity and efficiency [106–108]. In [109] a handset transmitter is implemented with a slow supply to reconfigure peak power and improve efficiency. LINC combined with ET allows for efficiency enhancement while potentially reducing supply bandwidth requirements compared to linear ET or saturated polar [110]. Discrete peak power reconfiguration is possible by symmetrically switching amplifiers in and out in both LINC branches [111]. In [112] four parallel LINC branches, in combination with full, half, or disabled supplies,
enables several output power configurations.

The LINC efficiency drops faster than class-B in OPBO, i.e. versus out-phasing angle, due to increasingly out-of-phase power combining. However, saturated and switched amplifier output power decreases with the supply voltage. By employing multi-level supplies the outphasing angle can be decreased for the same output power level, improving the average efficiency [113–116]. Methods to optimize the discrete multi-levels for a given PDF are developed [117]. In an extension, allowing the amplifiers to simultaneously have different supply levels enables further reduction in the outphasing angle by the combination of asymmetric power vectors [118–120]. In [121] discrete RF-PWM duty-cycle control in combination with asymmetric multi-level LINC adds additional artificial power levels to reduce the outphasing angles further.

Efficiency enhanced LINC is demonstrated in various technologies, e.g. CMOS [110, 115, 118, 119] and Bi-CMOS [111]. The 1.95-GHz asymmetric 4-level GaN transmitter in [120] has 41% overall efficiency for a 36-dBm 40-MHz 16-QAM signal with 8 dB PAPR.

### 3.3 Dynamic load modulation

Despite the schematically close resemblance to DSM, i.e. input power and baseband control, the DLM architecture in Fig. 3.4 is radically different. By dynamically tuning reactive components in the matching network, typically a single varactor, the PA efficiency is optimized in OPBO. The control voltage op-amp can be made highly efficient and high-speed given the limited power requirement. This is the main advantage of DLM compared to DSM, i.e. the baseband control is by voltage and not by power. Total transmitter efficiency is therefore dominated by the load modulated PA performance. This is also the complicating factor in DLM design, as the OPBO performance, e.g. high efficiency dynamic range, frequency agility, and instantaneous signal bandwidth, are all set by the PA mode of operation and load network complexity. Involving a varactor semiconductor technology or similar in the PA design, especially on the high power side of the transistor, is a further complication. Considerations on tunable component technologies related to DLM are discussed in Chapter 4, where the development of dedicated SiC Schottky diode varactors is presented.

The efficiency enhancement principle by DLM under class-B operation (Fig. 3.2 (b)) is conceptually simple. Assuming some fixed parasitic output capaci-
3.3. DYNAMIC LOAD MODULATION

tance, the transistor optimum DLM trajectory follows a constant susceptance trajectory. Conductance/resistive modulation, however, is not realized with low complexity using reactive components. Although there are earlier demonstrations of electronically tuned load networks for output power reconfiguration, e.g. [122], some of the first DLM results under modulated signal conditions are as recent as 2003 [123]. Here the output capacitance of a MOSFET is used for approximate conductance tuning in a class-E PA, trading perfect DLM for a lower complexity load network with a single control voltage. On the other hand, demonstrated by the GaAs HEMT PA in [124], accurate class-B DLM requires two varactor controls. The mode of operation is therefore closely related to the load network complexity. In [B] it is theoretically shown that high OPBO efficiencies are achievable in class-J by pure reactance modulation. Described in more detail in Section 3.3.1, this enables low-complexity varactor load networks with a single control voltage [B,C].

DLM is a relatively new architecture and pre-studies by simulation and load-pull hinted that Watt-level microwave implementations would be possible [125,126]. At the time, a record 7-W LDMOS implementation at 1-GHz followed, approaching DLM like a tuner design problem [127]. In this solution the DLM matching network is designed in a 50-Ω environment and connected external to a pre-existing PA. Other publications propose a similar tuner approach [124,128–130]. However, due to significant electrical distance between PA and tuner, the modular approach suffers from bandwidth limitations. Enabling wider bandwidths, later DLM publications have instead adopted to integrate the tunability directly in the PA output matching. Corresponding DLM network design has tended to be entirely empirical and based on load-pull methods [66,131–133]. Problematically, this allows for limited understanding of the DLM operation and how it relates to the choice of load network topology, e.g. in terms of OPBO performance, bandwidth, power scaling, and varactor requirements. The class-J DLM theory in [B] is suitable for such analysis. Similarly, a comprehensive theory for efficient class-E RF-PWM PA design with DLM is given in [h].

Similar to ET, the transistor in DLM operates like a current source, allowing adjustment of PA linearity and efficiency by shaping the input power and load network control functions. In [134] a simple inverse model, based on static measurements, is applied for basic transmitter linearization. More complete DPD methods to meet base station linearity requirements are reported in [135,136]. The effects of time-misalignment and bandwidth reduction schemes are reported in [90], where a reduction of the control signal to 2x the signal bandwidth is found to have negligible effect on the efficiency.

DLM has been demonstrated in a variety of applications. Several sub-W CMOS DLM transmitters utilizing switched capacitors and linearized to standard have been published [137–139]. Descriptions of both class-D and class-F DLM PAs are given in [140,141]. An overall transmitter efficiency of 60% with a 3.8 MHz 6.7 dB WCDMA signal is reported for a CMOS-GaN class-E RF-PWM DLM PA at 2 GHz in [f]. A quad-band PA, between 0.9–2.1 GHz, with DLM functionality utilizing silicon-on-glass pn-diode varactors is described in [142]. Octave bandwidth DLM performance is reported in [66], with 30–45% efficiency at 10 dB OPBO over 1–1.9 GHz. A record, by an order of magnitude, 86-W peak power DLM PA at 2.14 GHz is enabled by the
power scaling of class-J DLM theory [C]. In [131] a transmitter PAE of 44% is reported for a bandwidth-upscaled 38 MHz WCDMA-type signal with 6.7 dB PAPR.

### 3.3.1 Class-J varactor-based DLM

The class-J mode of operation is interesting for PA implementations at microwave frequencies, as the inevitable parasitic transistor output capacitance is utilized to enable high efficiency operation [9]. This mode also turns out to be highly suitable for varactor-based DLM.

Suppose a low-pass filter is connected directly to the transistor output, presenting an extrinsic open circuit above the carrier frequency. The intrinsic current source, assuming the simplified transistor model in Section 2.1, is thereby capacitively loaded by the output capacitance at the higher harmonics. A few examples of class-J output power and efficiencies for different combinations of $X_{Cds}/R_{opt}$ and intrinsic fundamental load are tabulated in [9] (p. 71). Interestingly, it turns out that class-B output power and efficiencies are still attainable under these conditions, by intrinsic fundamental matching that includes a reactive component. Mentioned in the previous section, resistive class-B DLM essentially requires a varactor load network with two control voltages. The prospect of a simpler load modulation alternative motivated the class-J investigation in [B].

The listing in [9] offers little insight regarding PA behavior under extrinsic DLM. This is a major contribution in [B], where output power and efficiency contours are calculated for different $X_{Cds}/R_{opt}$ versus extrinsic fundamental load impedance, as shown in Fig. 3.5. These contours expose the required load network behavior to enable efficient class-J DLM. Optimum DLM trajectories are marked by the OPBO labels in Fig. 3.5. The peak power load resistance is seen to increase with the $X_{Cds}/R_{opt}$ ratio. At low ratios ($\approx 0.5$) the optimum trajectory immediately bends towards low load resistances, whereas at high ratios ($\approx 4$) a curving trajectory is observed. However, at intermediate ratios ($\approx 2$) significant high efficiency OPBO dynamic range is achievable by pure load reactance modulation.

Presented differently in [B], there is significant overlap between the high efficiency regions of different $X_{Cds}/R_{opt}$. Given that the transistor operating frequency is proportional to $(X_{Cds}/R_{opt})^{-1}$, a solution for frequency reconfigurable DLM appears. After empirically choosing a fixed load resistance $R_L = R_{opt}/2$, OPBO and efficiency contours versus frequency and load reactance can be calculated, as shown in Fig. 3.6. More than 70% efficiency over 6 dB OPBO dynamic range is seen for $(X_{Cds}/R_{opt})^{-1} = \{0.4, 0.8\}$, i.e. over an octave of bandwidth.

For implementation, a load network consisting of a varactor-tunable series L-C resonator between transistor and $R_L$ is a perfect fit, where the bandpass characteristic conveniently provides increasing impedance at higher harmonics. The octave DLM implementation in [66], empirically designed based on load-pull, employs a series L-C resonator and appears to operate in class-J. This type of frequency reconfigurable DLM highlights the need for varactors with high effective tuning range. At the high frequency peak output power the varactor is tuned at $C_{min}$, where it can handle significant RF-swing.
3.3. DYNAMIC LOAD MODULATION

Figure 3.5: Class-J output power and efficiency prior to compression versus fundamental load impedance for different $\frac{X_{Cds}}{R_{opt}}$. Optimum DLM, keeping high efficiency, follows the output power back-off label placements.

Figure 3.6: Output power back-off efficiency versus load network reactance and inverted $\frac{X_{Cds}}{R_{opt}}$ when $R_L = \frac{R_{opt}}{2}$, (a) contours, (b) cross-sections.
CHAPTER 3. HIGH EFFICIENCY TRANSMITTERS

Figure 3.7: Theoretical calculation (dashed) and nonlinear transistor model simulation (solid) of class-J operation at $R_L = 17\Omega$. Efficiencies are listed in the legend (theoretical results in parentheses), (a) intrinsic waveforms, (b) intrinsic load lines.

ever, at the low frequency peak output power the varactor must be biased at an increased tuning ratio, where the voltage handling is significantly reduced. In [66] this is solved by a low supply voltage, reconfigured versus frequency. Varactor effective tuning range therefore limits the PA output power and frequency tunability, further motivating the tailored SiC varactor design in Chapter 4.

The demonstrator design in [B] is a pre-study for the high power implementation in [C]. In both circuits the same 0.7-μm gate length GaN HEMT technology from Mitsubishi Electric, Co. is employed. A nonlinear Angelov/Chalmers transistor model of a 3-mm unit-cell is developed based on pulsed I(V), DC, and multi-bias S-parameter extractions. Good agreement is observed in the comparison between theoretical and nonlinear transistor model waveforms, output powers, and efficiencies under class-J operation in Fig. 3.7. This GaN HEMT technology offers $X_{C_{ds}}/R_{opt} = 3$ at the design frequency of 2.14 GHz and $V_{DC} = 30$ V.

A schematic of the load network topology adopted in the designs is shown in Fig. 3.8. The transmission line stubs in the low-pass filter provide an open circuit at the second harmonic, while being transparent at the fundamental frequency. This network therefore provides reactance load modulation at the transistor reference plane by a shunted varactor, inverted through a quarter wave line. Indeed, this topology is narrowband, but provides some useful properties compared to a series L-C realization. The transistor is easily biased through the load network, as there is no DC-blocking function. Unlike a series L-C topology, where the output load resistance is dictated by the transistor technology, the $R_p$ load resistance in this topology is related to the choice of low-pass filter $Z_c$. More importantly, this degree of freedom in impedance level allows for adjustment of the voltage swing over the varactor (parallel with the load resistance).

Load network parameters for the two designs are shown in Table 3.1. Scaling from 1-cell [B] to 8-cells [C] is straightforward, with $R_L$ and $X_L$ decreasing.
by a factor of 8. Skipping an output transformer \((R_p = 50\Omega)\) the design in [B] ideally needs an anti-series connection of 1.5 pF varactor devices. Only 3.0 pF devices were available at time of assembly, which reduces the peak output power of the PA by 3 dB. In [C] \(R_p\) is reduced to 12.5 \(\Omega\) to handle the voltage swing at high output powers. This quarter reduction in \(R_p\) requires a fourfold increase in the effective varactor capacitance, realized by two parallel 3 pF anti-series stacks.

The PA matching networks are realized on substrates with high dielectric constants \((\epsilon_r = 10–150)\). This allows for downscaling, specifically to fit the high power DLM PA into a pre-existing package. Three-dimensional EM simulations of the transitions between substrates, connected by bondwires, are made to ensure proper handling of edge effects. This is not critical in [B], consisting of one input and one output substrate in alumina. However, the design in [C] involves 10 substrates, excluding transistor, varactors, and microchip capacitors.

A photo of the assembled PA in [B] is shown in Fig. 3.9 together with measured CW power sweeps at different varactor control voltages. Peak output power is 38 dBm, which is 3 dB lower than the expected 41 dBm with appropriate varactor values. Tuning the varactor control from 30 to 4 V, with appropriate input power control, enhances the efficiency at 8 dB OPBO from 20% to 45%. Observe that the varactors in these designs are biased relative to the transistor drain, i.e. \(V_c = 30\) V implies a varactor bias of \(-60\) V. The assembled high power DLM PA in [C] is shown in Fig. 3.10 together with pulsed measurements of output power and efficiency under optimum control at different drain supply voltages. With a 40 V supply the peak output power is 86 W. Efficiency enhancements in the order of 10–15 percentage-units are observed in OPBO and good agreement with simulations is observed.

Optimum PAE control functions for the PA in [C] are identified from pulsed measurements at \(V_{DC} = 40\) V and shown in Fig. 3.11 (a). These functions
CHAPTER 3. HIGH EFFICIENCY TRANSMITTERS

Figure 3.9: DLM PA [B], (a) photo (CuW carrier: 35 mm x 20 mm), (b) power sweeps at 2.08 GHz for different control voltages.

Figure 3.10: DLM PA [C], (a) photo (package: 40 mm x 20 mm), (b) measured (solid) and simulated (dotted) efficiency at 2.14 GHz for $V_{DC}$: 20 V (blue), 30 V (green), 40 V (red), given optimum control. Power sweeps at $V_c - V_{DC} = -60$ V (dashed) for reference.
are fitted by polynomials to form a static inverse model (SIM) of the amplifier and given as input to the DPD algorithm. A vector switched generalized memory polynomial (VS-GMP) model is used for behavioral modeling [143]. This type of model allows for different behaviors at different signal levels, making it suitable in DLM where control voltage modulation is only active at high powers [f]. For a 17-W 3.84-MHz WCDMA signal with 6.7 dB PAPR at 2.14 GHz the transmitter efficiency is 34%. Spectra are shown in Fig. 3.11 (b), with a linearized ACLR below -46 dBc to pass spectral requirements.

The demonstrators show the applicability of class-J DLM theory in practical design, simultaneously proving that DLM architectures are scalable towards the higher output powers required in macro base stations. Discussed in [B], efficiency numbers below state-of-the-art are mainly attributed to substrate ohmic losses in the packaged implementation and should be redeemable by hybrid implementation on lower loss substrates.

### 3.4 Active load modulation

Active load modulation refers to the use of transistors to realize time-varying loads for PA OPBO efficiency enhancement. This is similar to DLM, except that active instead of passive devices are used and the load modulation is controlled by RF input power and phase instead of baseband voltage. A strength of ALM is the familiar use of transistor technology. Exclusive use of active devices also means that all expensive semiconductor components can contribute to the PA output power. ALM operation is conceptually more complicated than DLM and design can be challenging. A main complication is that active devices consume power, implying that high transmitter efficiency requires all incorporated transistors to simultaneously and efficiently load modulate each other versus OPBO. Popular and proven ALM architectures for transmitters include the Chireix and Doherty topologies. Intricacies of implementations at microwave frequencies, especially the extension to wide bandwidths, has made this an active research area over the last decade.
CHAPTER 3. HIGH EFFICIENCY TRANSMITTERS

Figure 3.12: Active load modulation principle.

Figure 3.13: Active load modulation architecture. Current source amplitude and phase functions versus OPBO are co-designed with the matching networks.

The ALM principle is explained by the circuit in Fig. 3.12. Currents $I_1$ and $I_2$ are entering a two-port containing a shunted resistive load $R$, e.g., an output transformer or the load connected to the PA output. From Kirchhoff’s law, the current through $R$ is the sum of the port currents and the output power is given by the in-phase components,

$$P_{out} = \frac{R|I_1 + I_2|^2}{2}. \quad (3.1)$$

Now, the voltage over $R$ is proportional to the complex load current $(I_1 + I_2)$, allowing calculation of the impedances $Z_{in1}$ and $Z_{in2}$ looking into each port,

$$Z_{in1} = R\left(1 + \frac{I_2}{I_1}\right), \quad (3.2)$$

$$Z_{in2} = R\left(1 + \frac{I_1}{I_2}\right). \quad (3.3)$$

The port impedance expressions are different and functions of both currents. If the current amplitudes and phase relations are modulated over time, so are the impedances, resulting in an ALM effect. It is important to note that many combinations of $I_1$ and $I_2$ result in the same output power, but different load modulation. From previous discussions on modes of operation, corresponding transistor efficiency is dictated by the loading of the intrinsic current source. Addition of intermediate matching networks can ensure that the ALM at $R$ is transformed to the proper load modulation at the transistor current sources. The architecture is shown in Fig. 3.13. Successful co-design of matching networks and current amplitude and phase functions versus OPBO have resulted in the well-known Doherty and Chireix architectures [9].
3.4. ACTIVE LOAD MODULATION

3.4.1 Doherty

The ingenious Doherty amplifier was proposed in 1936 [144]. An assumption of class-B conditions is made, as short-circuited higher harmonics ensures no complicating or efficiency degrading voltage components. Matching networks to fit into the ALM architecture in Fig. 3.13 are shown in Fig. 3.14. The so-called carrier amplifier provides the $I_1$ current, proportional to the PA output voltage. At low output powers the other, so-called peak amplifier, is turned off and $Z_{in1}$ remains fixed. By proper design the quarter wave line in MN$_1$ transforms this impedance to $2 \cdot R_{opt}$. At half the maximum carrier current, i.e. at 3 dB carrier OPBO and at 6 dB PA OPBO, the carrier reaches full voltage swing and the PA efficiency peaks. The peak amplifier is then turned on to deliver $I_2$ current. With the peak connected directly to the load, as both $I_1$ and $I_2$ are increasing, the $Z_{in2}$ impedance decreases from open-circuit to $R_{opt}$ at peak power. A $90^\circ$ phase-shift (at the peak input) is applied to the peak current to ensure in-phase combining and purely resistive load modulation. The carrier amplifier $Z_{in1}$ is seen to increase as $I_2$ increases after the peak kicks in. Mitigating this unwanted effect the quarter wave line inverts the impedance, instead producing a decrease from $2 \cdot R_{opt}$ to $R_{opt}$ where the PA efficiency peaks again with both carrier and peak working at their maximum output power.

With ideal current control the Doherty is a linear amplifier. However, implementing the Doherty as a single-ended PA is highly desirable. The required nonlinear current control functions are therefore approximated in typical transistor implementations. Power splitting from a single input is followed by appropriate phase delay to the peak amplifier. The carrier is typically class-AB, while the peak is biased in class-C to control the power turn-on level. Class-C operation is not ideal, however, as it leads to soft current turn-on by conductance angle modulation. Commonly the peak device is selected to be larger than the carrier to provide the required higher current slope. A larger device implies that uneven power splitting may be necessary, wasting additional input power when the peak is off. Design guidelines for efficiency and linearity, e.g. in terms of power splitting, device size selection, and bias, are investigated in [145–147].

The original Doherty assumes linear sub-amplifier operation. Implementation of the class-F Doherty, with theoretically higher peak efficiencies, is therefore possible [148, 149]. Even employing saturated class-F$^{-1}$ amplifiers is possible after theoretical modifications [150–152]. Other extensions of the Doherty, adding more sub-amplifiers, are reviewed in [153], e.g. in balanced, parallel, asymmetric/distributed N-way [154–156], and multi-stage [155] con-
3. High Efficiency Transmitters

figurations.

In the quest for wider bandwidth operation, recent efforts have focused on addressing bandwidth limitations in the microwave Doherty. Bandwidth limitations in compensation elements, e.g. the common use of offset lines and phase delay lines, are investigated in [157]. Reduction in the output transformer ratio can reportedly improve bandwidth performance [158, 159]. Synthesis of wideband impedance inverting networks by the real-frequency technique is investigated in [160]. Bandwidth limitations on absorbing transistor output capacitances in the impedance inverter is analyzed in [161].

The realization of single-ended Doherty PAs has significant merit. However, the engineering complexity to compensate for nonideal operation, e.g. gate bias modulation on the class-C peak amplifier [162,163], encourages reflection on the single-ended motivation. A dual-RF input Doherty, with individual control of carrier and peak amplifiers, allows for closer realization of the original architecture. Digitally the input power division can be made perfectly efficient with optimum current control and phase-delay, e.g. for compensation of nonideal AM/PM and reverse leakage effects. At the cost of increased hardware complexity (e.g. two quadrature IQ-modulators), a dual-input solution can allow for improved efficiency and linearity with increased signal bandwidths [164]. Dual-input Doherty PA bandwidth extension and efficiency improvements by phase-compensation are reported in [165,166]. Modification of the classical Doherty impedance levels in a dual-input implementation improves the bandwidth further [c]. In fact, covered in Section 3.4.3, a dual-input PA with 100% fractional bandwidth is possible with a different combiner topology [D]. This PA operates like a Doherty in part of the bandwidth.

Doherty PAs achieve excellent performance. A 2.14 GHz triple-input 3-Way GaN implementation gives a PAE of 53% for a 38 dBm 11.5 dB PAPR WCDMA signal [154]. Combined with ET a CMOS Doherty for handset applications achieves a PAE of 39% for a 24 dBm WiMAX signal [167]. The GaN MMIC implementation in [168] demonstrates 27.5 dBm 10 MHz 256-QAM with 7.8 dB PAPR at higher than 35% PAE across 6.8-8.5 GHz. At higher frequencies still, a 23 GHz MMIC realization is reported with 27% PAE for a 22 MHz 256 QAM signal [169]. A 500 W GaN Doherty proves the architecture power scalability [170]. Wideband signal transmission is demonstrated at 3.5 GHz with 40% efficiency for a 50 MHz LTE-A signal [171].

3.4.2 Chireix

The Chireix amplifier was conceived in 1935 [172]. Different from the transistor current source description adopted in this thesis, operation of the Chireix assumes transistors acting as constant voltage sources. Given this constant voltage source assumption, the architecture is suitable for use with heavily saturated or switched amplifiers [9]. Although detailed analysis is complicated, basic operation is intuitive enough. Consider two saturated class-B amplifiers connected to a combiner. These amplifiers are susceptible to resistive load modulation. In fact, provided enough saturation the output power is essentially proportional to the load resistance, with full voltage swing to ensure high efficiency. Now, if the phase between the amplifiers is modulated the load impedances are given by,
3.4. ACTIVE LOAD MODULATION

\[ Z_{in1} = R(1 + \cos(\phi) + j \sin(\phi)) = Z_{in2}^*, \]  \hspace{1cm} (3.4)

where \( \phi \) is the outphasing between \( I_1 \) and \( I_2 \) (equal in amplitude) at the \( R \) reference plane. Clearly outphasing results in both resistive and reactive ALM. Typical matching networks to form a Chireix combiner suitable for microwave implementation, as suggested in [173], are shown in Fig. 3.15. Quarter wave impedance inverters ensure decreasing resistive load modulation towards \( R_{opt} \) at full inphasing and peak power. The reactive contributions, however, cause efficiency degradation. Addition of shunt compensation elements \( \pm B \) can cancel this unwanted effect, but only at a single outphasing angle. The choice of compensation value therefore affects the efficiency characteristic by moving the point of perfect resistive loading (high efficiency) versus OPBO. Optimum choice of \( B \) is therefore dependent on the signal PDF. This theoretically simple approach to trading peak and OPBO efficiencies makes the Chireix technique highly interesting in transmitters.

In practice, as the Chireix combiner is non-isolating, issues with phase and amplitude imbalance arise. Analytical analysis of the typical transmission line implementation reveals that different reactive loading of the amplifiers is a main source of nonlinearity [174,175]. Choice of amplifier mode of operation also affects the transmitter linearity and this has been studied, e.g. class-D/E/F are compared in [176], class-AB/F in [177], saturated class-B in [178, 179].

Design of the combiner is central in Chireix research. Analytical transmission line based design equations are given in [174]. Studies of different combiners, e.g. differently compensated Chireix, lossy hybrid, Wilkinson, and harmonic suppressing combiners are reported in [180–184]. A general conclusion is that better linearity is offered by lossy combiners, especially when operated offset from center frequency. Various novel Chireix combiners are also presented, e.g. based on right/left-handed transmission lines [185], coupled-lines [186], transformers [187], and baluns [188]. Making the combiner reconfigurable is a reoccurring Chireix theme, e.g. by varactor-based adaptive power combining [189] and tunable compensation elements [190–192]. Not surprisingly there are patents to be found [193]. A presentation of lossless combiner design for four or more outphased amplifiers, resulting in multiple OPBO efficiency peaks, is given in [194].

Due to the above mentioned imbalances, it is not always sufficient to apply phase-only pre-distortion in Chireix amplifiers [195]. In fact, by switching to linear operation at the resistive OPBO peak efficiency point higher back-off efficiencies are achievable [190,196,197]. The dual-RF input PA in [D] operates
 CHAPTER 3. HIGH EFFICIENCY TRANSMITTERS

in outphasing mode in parts its bandwidth, i.e. outphasing being a special case of dual-input control.

Beautiful Chireix efficiency curves are demonstrated by class-E at 1 MHz and 1.95 GHz in [186,198], respectively. CMOS class-D [199,200] and 5 GHz GaAs MMIC class-F [201] circuits are also reported. The digitally compatible 19-W peak power CMOS-GaN class-E Chireix in [186] gives 42% total transmitter efficiency for a 9.6 dB WCDMA signal. Average efficiency of 50% for the same signal is reported for a 90-W peak power GaN implementation [202].

3.4.3 Dual-RF input Doherty-outphasing continuum

In the previous sections the Doherty and Chireix are presented as two successful examples of matching network and transistor drive co-design for the enhancement of back-off efficiency. Similar to the continuous PA modes of operation (Section 2.4.1), it turns out there is a continuum of co-designs producing comparable performance. Mentioned previously, a PA implementation with two independent RF inputs is advantageous, allowing for frequency optimum current control to maximize the efficiency enhancement and improve linearity. This section treats the ideal dual-input Doherty-outphasing continuum described in [203], featuring 100% fractional bandwidth with high average efficiency. The effort in [D] to maintain performance while considering transistor parasitics and non-ideal higher harmonic conditions is then presented.

A simple way to evaluate proposed ALM matching networks is to record a large set of output power and corresponding efficiency for different combinations of transistor current amplitudes ($\beta_1$ and $\beta_2$) and relative phase ($\phi$). Some of these combinations will produce the same output power, but at different efficiency. Optimum current control ($\beta_{1,\text{opt}}(P_{out})$, $\beta_{2,\text{opt}}(P_{out})$, $\phi_{\text{opt}}(P_{out})$), that maximizes the efficiency versus output power with the networks, can then be identified. Average efficiency, provided optimum control, for a certain signal PDF can then be estimated using (2.5) to quantify performance. This approach is adopted in [D] to optimize the networks for wideband performance.

Consider the rather generalized matching networks in Fig. 3.16. Evaluating the performance of a Doherty realization ($\theta_1 = 90^\circ$, $\theta_2 = 0^\circ$, $Z_{c1} = Z_{c2} = R_{\text{opt}}$ and $R_L = R_{\text{opt}}/2$), assuming short-circuited higher harmonics, produces the output power and efficiency cloud in Fig. 3.17 (top left). When increasing the number of combinations the optimum currents in Fig. 3.16 (bottom left) converge to reproduce the currents well-known from Doherty theory [9]. Now, repeating the procedure with the same impedances, but changing the transmission line lengths ($\theta_1 = 114^\circ$, $\theta_2 = 57^\circ$) gives the outphasing-type results in
Fig. 3.17: Doherty (left column) and outphasing (right column). Clouds of efficiency and output power results for different current amplitude and phase combinations (gray dots), with maximized drain efficiency indicated (red lines) under optimum current control.

Fig. 3.17 (right column). Interestingly, although the optimum current control arguably appears more complicated, high efficiency is maintained over a larger OPBO dynamic range compared to the Doherty. This outphasing functionality is explainable. Making the transmission lines $\approx 90^\circ \pm \Delta$ approximates the addition of Chireix compensation susceptances, with the peak power inphasing angle shifting to $\phi = \theta_1 - \theta_2$. Discussed in Section 3.2.1, the efficiency of outphasing drops rapidly in deep back-off. The optimum control therefore keeps one transistor turned off in this region to operate in the more efficient linear mode. At higher powers the transistors are outphased with equal amplitudes, driven to avoid compression (not handled in the simplified model) unlike original saturated Chireix. However, this current amplitude modulation conserves input power and simultaneously allows for smooth transition into linear operation. Observe, however, that the nonlinear optimum current control will cause signal bandwidth expansion at the PA inputs. It is therefore conceivable that a dual-input transmitter becomes limited by the digital bandwidth. Bandwidth reduction schemes similar to those in ET and DLM are then applicable, trading efficiency for linearity.

Average efficiency for a 6.7 dB WCDMA signal, provided optimum current control, is calculated as function of $\theta_1$ and $\theta_2$ in Fig. 3.18. It turns out many combinations of electrical lengths result in high average efficiency. For reference, average efficiency for this signal and ideal class-B is calculated to 25% in Section 2.2.1. In the efficiency contours there is diagonal symmetry due to the symmetry of the matching networks and a periodicity of $180^\circ$ due the periodic behavior of transmission lines, while the optimum control repeats every $360^\circ$. All symmetric and periodic solutions equivalent to the Doherty and outphasing in Fig. 3.17 have been marked for clarity in Fig. 3.18. Observe that the peak output power remains constant with perfect class-B efficiency independent of the electrical lengths. Now, the electrical length of a transmission line is pro-
portional to the operating frequency. The frequency dependent performance of the PA is therefore captured by drawing a line from the design frequency \( f_0 \) point to the origin. This is done for two choices of \( \theta_1 \) and \( \theta_2 \) in Fig. 3.18. An intelligent choice of periodic Doherty point \( (\theta_1 = 90^\circ, \theta_2 = 180^\circ) \) makes the PA pass through two outphasing points. This forms a Doherty-outphasing continuum with more than 65% average efficiency over 100% fractional bandwidth. For reference, the efficiency bandwidth for the usual Doherty point \( (\theta_1 = 90^\circ, \theta_2 = 0^\circ) \) is seen to be significantly lower. Observe that the bandwidth of a single-ended Doherty, without frequency optimum current control, is narrower still.

The Doherty-outphasing continuum results are both high efficiency and wideband. However, the results are calculated under the assumption of short-circuited higher harmonics. From the discussion on wideband design in Section 2.4, this is a practical impossibility for large bandwidths. Further, there is no consideration to microwave transistor parasitics, such as the output capacitance, which inevitably will affect performance. The idea in [D] is to include transistor parasitics and calculate the proper frequency dependent response of the matching networks at the higher harmonics. This is done by a linear multi-harmonic calculation method solving, using the simplified transistor model in Section 2.1, voltage waveforms, output power and efficiency. Proposed matching networks to include basic transistor parasitics, i.e. output capacitance and drain bondwire inductance, are shown in Fig. 3.19.

The matching networks are optimized in a brute force evaluation of a large number of realizations. Specifically, performance of the 14 million non-degenerate combinations of the parameters in Table 3.2 are evaluated in [D]
3.4. ACTIVE LOAD MODULATION

$$\mathcal{F} \equiv \mathcal{F}_\mathcal{V} G \chi \mathcal{F}$$

Figure 3.19: Proposed matching networks incorporating transistor parasitics, with no assumption of higher harmonic short-circuits.

Table 3.2: Circuit parameter combinations.

<table>
<thead>
<tr>
<th>$X_{Cds1/2}$ @ $f_0$ ($R_{opt}$)</th>
<th>$Z_{c1/2} &amp; R_L$ ($R_{opt}$)</th>
<th>$\theta_{1/2} @ f_0$ ($^\circ$)</th>
<th>$X_{L1/2}$ @ $f_0$ ($R_{opt}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, ..., 6</td>
<td>0.25, 0.5, 1, 2, 4</td>
<td>0, 5, ..., 180</td>
<td>0.125, 0.25</td>
</tr>
</tbody>
</table>

for $f/f_0 = 1$–$3$ at 11 frequency points and 5048 output current conditions. After solving the waveforms the output power and average efficiency is evaluated under varying signal PAPR and supply conditions. These calculations run on a computing cluster (C3SE), where a single processor core solves 140 circuits per minute, occupying 70 cores for a total of 24 hours. Results are then post-processed to identify the highest performing circuits. A constraint on maximum output power variation of 2 dB is imposed as the circuits are sorted by highest minimum average efficiency for a given bandwidth.

Performance of a few interesting circuit realizations are shown in Fig. 3.20, with network parameters listed in Table 3.3. Empirical investigation indicates that wideband and high PAPR performance correlates with high transistor $X_{Cds}/R_{opt}$. Similar to the ideal continuum, circuit-A features more than 100% fractional bandwidth.

Implementation of the circuit-A GaN demonstrator in [D] is straightforward, as all parasitics and higher harmonics are catered for. The load network simply consists of two transmission lines connected to a broadband stepped-impedance transformer to 50 Ω. Wideband input matching is more empirical, however, to ensure stability without sacrificing high frequency gain. Theoretical performance of the circuit is shown in Fig. 3.21 and compares well with

Table 3.3: Circuit (A-D) parameters and efficiency performance

<table>
<thead>
<tr>
<th>Circuit</th>
<th>Signal (PAPR)</th>
<th>$X_{Cds}/R_{opt}$ @ $f_0$</th>
<th>$Z_{c1}/R_{opt}$</th>
<th>$Z_{c2}/R_{opt}$</th>
<th>$\theta_1 @ f_0$</th>
<th>$\theta_2 @ f_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>WCDMA (6.7 dB)</td>
<td>6.0</td>
<td>1.0</td>
<td>1.0</td>
<td>31°</td>
<td>73°</td>
</tr>
<tr>
<td>B</td>
<td>WCDMA (6.7 dB)</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>5°</td>
<td>25°</td>
</tr>
<tr>
<td>C</td>
<td>LTE (9 dB)</td>
<td>6.0</td>
<td>0.5</td>
<td>1.0</td>
<td>135°</td>
<td>60°</td>
</tr>
<tr>
<td>D</td>
<td>LTE (9 dB)</td>
<td>1.0</td>
<td>0.5</td>
<td>1.0</td>
<td>10°</td>
<td>20°</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Circuit</th>
<th>$X_L/R_{opt}$ @ $f_0$</th>
<th>$R_L/R_{opt}$</th>
<th>$V_{ds2}/V_{ds1}$</th>
<th>$\eta_{avg} &gt; 50%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.125</td>
<td>0.5</td>
<td>1.0</td>
<td>108</td>
</tr>
<tr>
<td>B</td>
<td>0.25</td>
<td>0.25</td>
<td>0.7</td>
<td>56</td>
</tr>
<tr>
<td>C</td>
<td>0.125</td>
<td>0.5</td>
<td>0.7</td>
<td>58</td>
</tr>
<tr>
<td>D</td>
<td>0.25</td>
<td>0.25</td>
<td>0.5</td>
<td>42</td>
</tr>
</tbody>
</table>
Figure 3.20: Average efficiency and output power versus frequency for the circuits in Table 3.3 provided optimum current control.

layout-ready simulation of the GaN demonstrator. This proves the usefulness of the simplified transistor model and linear multi-harmonic calculation approach. Further, it shows that 100% fractional bandwidth with significant back-off efficiency enhancement is possible in a practical dual-input realization.

Accurate characterization of the dual-input PA is enabled by vector modulators to adjust input powers and relative phase. A large signal network analyzer (LSNA) is used to monitor the waves at the PA input reference planes. Shown in Fig. 3.22, sweeping different input power conditions produces a cloud of PAE versus output power and the optimum driving conditions can be identified. This optimum control can differ significantly from the theoretical predictions, requiring phase compensation due to nonlinear transistor effects, e.g. transistor feedback and AM/PM. Not all simulated points have been measured, but good agreement is observed in the optimum PAE behaviors.

A summary and comparison of simulated and measured PA performance at peak power and 6 dB OPBO is shown in Fig. 3.23. Measured maximum output power is $44 \pm 0.9$ dBm from 1–3.1 GHz, which is approximately 0.5 dB lower than simulated. Corresponding drain efficiencies are 68–45%, where the minimum at 1.7 GHz is stronger compared to the simulation. At 6 dB OPBO the
3.4. ACTIVE LOAD MODULATION

Figure 3.22: PAE versus output power for different input conditions and frequencies. Maximized PAE shown as red lines, (a) layout-ready simulations, (b) measurements and photo of amplifier (166 mm x 81 mm).

drain efficiencies are 68–48% and specifically 53% at 1.7 GHz. For modulated signal transmission the efficiency at OPBO is more critical than at maximum power. The measured gain is generally higher than simulated up to 3 GHz and the PAE at 6 dB OPBO exceeds 45% over the 1–3 GHz bandwidth. Unfortunately, a discrepancy between measured and simulated PA S-parameters at high frequencies indicates a problem with the realized input matching network. This causes an abrupt drop in the gain above 3 GHz, preventing continued characterization due to lack of sufficient pre-drivers. However, even when limited to a 1–3 GHz bandwidth, the PA in [D] is one of the most wideband back-off efficiency enhanced PAs reported to date.

Similar to the class-J DLM results, the dual-input PA demonstrator proves the validity of simplified transistor modeling and calculations methods for wideband efficiency enhanced design. The hardware implementations of these architectures are rather different, suggesting the solutions are complementary and can co-exist, with the choice of architecture depending on system specific requirements. It should be emphasized, however, that the results in this thesis are first efforts. The full potential of these architectures is yet to be realized.
Figure 3.23: PA performance under optimum input power conditions, (a) layout-ready simulations, (b) measurements.
Chapter 4

SiC varactors for dynamic load modulation

A critical part in DLM is the tunable component technology. The technology must allow for fast modulation and be capable of handling the power levels at the PA output. To ensure justifiable efficiency enhancement the technology simultaneously needs to be low loss. Different tunable components feature in DLM publications, e.g. pn-diodes [127, 142], MOS transistors [123, 131], switched CMOS capacitors [66,137,139], and capacitor banks [141].

Wide bandgap materials such as SiC, with high critical electric field, are especially suitable for varactor applications [204]. Further, as concretized by the class-J DLM theory in Section 3.3.1, varactor effective tuning range limits the power handling and frequency tunability of DLM transmitters. SiC Schottky diode varactors specifically tailored with large effective tuning range for DLM applications are therefore developed as part of this thesis work [E,F]. These devices have been successfully used in DLM PA [B,C,h], impedance tuner [g] and phase-shifter [d] circuits.

This chapter therefore shifts focus from amplifiers and transmitters to semiconductor technology. First, varactor considerations in microwave power applications and the benefits of wide bandgap SiC material properties are covered. Based on [E,F], epitaxial design decisions, device layout, fabrication process development, physical device modeling, DC and small-signal device characterization are presented. Nonlinear varactor characterization by multi-harmonic active source- and load-pull and considerations for high power tunable circuit design follows [G].

4.1 Microwave power varactors

The most common varactor figure of merit is the Q-factor, which measures the ratio between electrically stored power and power loss in the device [205]. An ideal varactor is purely reactive with infinite Q-factor. For the intrinsic Schottky varactor the small-signal Q-factor is given by,

$$Q_{\text{linear}}(V) = \frac{1}{2\pi f_0 R_{\text{drift}}(V) C_{\text{dep}}(V)},$$  \hspace{1cm} (4.1)
where $f_0$ is the fundamental excitation frequency, $C_{depl}(V)$ the depletion capacitance, $R_{drift}(V)$ the drift resistance through the undepleted part of the epi, and negligible leakage current is assumed. At high powers the Q-factor is also load dependent due to power loss to higher order harmonics [G]. Studies of nonlinear operation are covered in Section 4.6.

Another common varactor figure of merit is the small-signal capacitive tuning range,

$$T_{linear} = \frac{C_{max}}{C_{min}} = \frac{t_{j,max}}{t_{j,min}},$$ \hspace{1cm} (4.2)

where $C_{max}$ and $t_{j,min}$ are the capacitance and depletion region thickness at zero bias, while $C_{min}$ and $t_{j,max}$ are at the punch-through or breakdown voltage. This classical definition of tuning range lacks in the description of tunability at high power levels.

A C(V) shape typical for a uniformly doped semiconductor layer is shown in Fig. 4.1, with most of the tuning range concentrated in the low reverse bias region. However, there are additional constraints on the tuning range when superimposing a significant voltage amplitude. For a varactor diode the voltage waveform must not be allowed to enter the forward conduction and reverse breakdown regions, as this leads to high DC power dissipation and hazardous operation [127]. The waveforms in Fig. 4.1 indicate that the effective tuning range of the varactor quickly decreases with increasing voltage amplitude. A definition of effective tuning range (using two different peak-to-peak voltages) is used in [206] to explore epitaxial design trade-offs. A simpler definition of effective tuning range is adopted here and in [E],

$$T_{eff}(V_{RF}) = \frac{C_{max,eff}}{C_{min,eff}} = \frac{C(-|V_{RF}|)}{C(V_{BD} + |V_{RF}|)},$$ \hspace{1cm} (4.3)

where $|V_{RF}|$ is the superimposed voltage amplitude, and $V_{BD}$ the breakdown voltage. The effective tuning range is always lower than the small-signal tuning range ($|V_{RF}| \rightarrow 0$). A simplifying assumption is made, i.e. that the large
signal time-average capacitance equals the small-signal capacitance. Based on nonlinear measurements this approximation is fair for smooth C(V) characteristics [G]. However, depending on the application, e.g. narrowband filters, the resulting detuning may be critical and both varactors and circuit topology can be optimized to mitigate this [207].

A large effective tuning range is intuitively achievable by realizing a flat C(V) at high reverse biases, to minimize the increase in $C_{\text{min,eff}}$, together with a slowly decaying C(V), to maintain high $C_{\text{max,eff}}$. An complementary way to improve the effective tuning range, independent of varactor technology, is to stack devices in anti-series configuration. This effectively halves the voltage swing over both devices. Anti-series configuration is also desirable for its linearity benefits [206,208,209]. The cost of stacking is a reduction of absolute reactance, however, requiring larger devices which may result in yield or parasitic problems.

The bulk properties of various semiconductor materials are shown in Table 4.1. For SiC the 4H crystal polytype is listed as it has the highest electron mobility, e.g. compared to the 6H and 3C polytypes, and 4H-SiC is used for all device fabrication in this work.

| Table 4.1: Semiconductor Bulk Properties (300 K) [30,210,211] |
|----------------|-------|-------|-------|
| $E_g$ (eV)     | 1.1   | 1.42  | 3.39  | 3.26  |
| $E_{cr}$ (MV/cm) | 0.3   | 0.4   | 3.3   | 3.0   |
| $\mu_n$ (cm$^2$/Vs) | 1350 | 8500  | 1405  | 1150  |
| $\varepsilon_r$ | 11.8  | 13.1  | 9.0   | 9.8   |
| $k$ (W/cm·K)   | 1.5   | 0.43  | 1.3   | 3.3-4.5 |
| $FOM_{\text{limit}}$ (THz-V) | 10   | 108   | 1217  | 823   |

The wide bandgap semiconductors (GaN and SiC) are seen to support an order of magnitude higher critical electric field ($E_{cr}$), due to their large bandgap ($E_g$), compared to the more narrow bandgap materials (Si and GaAs). The low-field electron bulk mobilities ($\mu_n$) at low doping concentrations are comparable between the materials, except for the significantly higher mobility in GaAs. In 4H-SiC the mobility is anisotropic and highest parallel to the crystal c-axis (the tabulated value) [211], which coincides with the intrinsic varactor current transport direction. The mobility in GaN is often treated as isotropic, supposedly due to the piezoelectric effect [212]. For high power devices, where device cooling is crucial, the excellent thermal conductivity ($k$) of SiC is another reason for its use in DC power devices [213], and as substrates for GaN HEMT epitaxy [30].

A varactor figure of merit $FOM_{\text{limit}}$ proposed in [204], which only considers material parameters, is listed in the table and given by,

$$FOM_{\text{limit}} = \frac{\mu_n E_{cr}^2}{4\pi} \geq V_{BD} \cdot T_{\text{linear}} \cdot f_{c,\text{min}} = FOM_{1D},$$

(4.4)

where $f_{c,\text{min}}$ is the cut-off frequency (where the Q-factor equals one) at $C_{\text{max}}$. According to this material properties set an upper limit for the product of breakdown voltage, small-signal tuning range, and cut-off frequency (Q-factor),
i.e. there is a trade-off between these quantities. The wide bandgap materials are seen to feature a superior figure of merit. This is attributed to higher critical fields, which enables higher drift layer doping concentrations and lower series resistance for the same C(V). In reality (4.4) is a simplification, as $\mu_n$ and $E_{cr}$ are doping dependent [211,214]. Furthermore, the figure of merit does not consider effective tuning range (the shape of the C(V)), but maximizes the small-signal quantities. However, it emphasizes that the critical electrical field is a most important varactor material property.

The bulk material properties of SiC and GaN are similar. Conductivity models for both materials, including incomplete ionization and crystal directional dependence [215], suggest the materials offer comparable varactor performance. However, the technology required to engineer tailored doping profiles is more mature in SiC and a deciding factor in the choice of material in [E,F]. Seemingly few publications specifically on SiC RF and microwave varactors are otherwise reported in literature [216–218].

4.2 Epitaxial design

Varactor specifications were established in [215] for an intended DLM application at 1 GHz, to replace a commercial silicon device. The SiC varactor was to be designed for a minimum breakdown voltage of $-60\,\text{V}$, a minimum effective tuning range of 2, and a series resistance lower than $1\,\Omega$ for a device with $C_{\text{min}} = 2\,\text{pF}$.

Various varactor C(V) functions, where analytical doping profile functions are assumed, have been investigated in literature. This includes the common $C(V) \propto (-V)^{-n}$ functions [219] encompassing the abrupt (uniform doping, $n = 0.5$, used in anti-series for low distortion [208, 209]) and hyper-abrupt (graded doping, $n = 2$, used in VCOs for linear frequency control) varactor types. The doping profile for minimization of intrinsic series resistance versus breakdown voltage is close to $n = 2/3$ [220]. Exponential C(V) functions have also been found to have good linearity properties (IM3 cancellation), with better effective tuning range than the abrupt varactor types [206].

For the varactor in this work a parameterized polynomial C(V) function was optimized numerically by employing a physically based device model to meet the required specifications. C(V) and epitaxial design by numerical methods has drawbacks compared to analytical modeling, being slower to evaluate and more difficult to understand for various device trade-offs. However, while analytical expressions can give improved understanding they can only hint at final device performance. Limits on lithographical dimensions and the minimum device parasitics can for instance affect the optimal epitaxial design [221]. There are other practical constraints as well, e.g. area utilization, realizable doping profiles, and epitaxial layer thicknesses for practical processing. Ultimately, epitaxial optimization is often done by iterative investigation, which allows for more complex (numerical) evaluation models.

The resulting C(V) from the work in [215] is shown in Fig. 4.2 and has a cubic voltage dependence, saturating before punch-through. In typical DLM networks the maximum voltage swing (at peak PA output power) occurs when the varactor is biased at $C_{\text{min}}$. It is therefore essential that the C(V) is flat over
Figure 4.2: Varactor C(V) and corresponding $N_D(x)$ doping specification developed in [215]. The spacer layer extends for 220 nm closest to the Schottky contact. The doping minima around 1 μm is the result of smooth C(V) saturation before punch-through. Electric field profiles are shown for three different junction voltages ($V - V_{bi}$).

a large bias range to avoid an increase in $C_{min,eff}$ that will otherwise reduce the dynamic range of the DLM network. A flat, implying linear, C(V) behavior at high power will also reduce the power loss to harmonic generation. This flat region is realized by making the breakdown voltage significantly larger than the punch-through voltage. The epitaxial design utilizes a uniformly doped spacer layer nearest the Schottky contact, where the spacer thickness sets the zero-bias per-area capacitance (per-area $C_{max}$). Use of a spacer layer is practical as the diode on-resistance is irrelevant for a varactor. Appropriate choice of spacer layer doping concentration ensures that the spacer layer is fully depleted by the built-in junction voltage. Although it has been shown that a delta-doping after an undoped spacer layer is more effective at minimizing the electrical field [222], such a solution is more prone to epitaxial error. Further, given that the spacer layer is the lowest doped part of the device, experiencing the highest electrical field, the spacer doping dictates the nominal device breakdown voltage. A peak electric field of 2.5 MV/cm, equal to the critical electrical field at $N_D = 10^{16}$ cm$^{-3}$ [214], is expected at the Schottky/spacer interface for a bias of $-220$ V. Traditional varactors are not operated beyond punch-through and will typically feature a sharp C(V) kink at saturation. In order to avoid the worst of this nonlinearity the C(V) smoothly levels out before punch-through [E]. This smoothing results in a concentration minimum in the graded part of the doping profile. If the cost of reducing the discontinuity is deemed to high in terms of additional series resistance, the doping profile may be cropped [F].

Besides the improved effective tuning range, the slow decay of the C(V) reduces the device temperature sensitivity. A change in temperature effectively shifts the C(V) versus bias voltage, as the built-in junction voltage is
temperature dependent. For an abrupt type of varactor, where much of the
tuning range is concentrated in a small reverse bias range, even small voltage
shifts can significantly affect the varactor capacitance at a fixed bias. The SiC
varactor is therefore highly suitable for close integration in a PA, where the
temperature may change with time and become significantly higher than room
temperature.

It should be emphasized that the epitaxial specification in Fig. 4.2 is an
application-specific optimum. In another application operating the varactor
beyond punch-through may not be necessary, or the effective tuning range
requirements may be different. This allows for other design trade-offs, e.g. in
breakdown voltage and losses.

4.3 Fabrication

Varactor development has involved 2” semi-insulating Si-face $8^\circ$ off c-axis 4H-SiC substrates with tailored epi grown at Linköping University (LiU) [E] and
Norstel AB [F]. Fabrication, as described in this section, mostly corresponds
to that in [F]. The substrates are typically diced into 16 mm x 16 mm chips and
cleaned in standard RCA1 ($H_2O:H_2O_2(30\%):NH_4OH(25\%)$ 5:1:1) and RCA2
($H_2O:H_2O_2(30\%):HCl(37\%)$ 6:1:1) solutions with an intermediate HF (10\%)
dip. Nickel is used heavily throughout the fabrication process as a CHF$_3$
reactive ion etching (RIE) mask, having a 1:20 etch ratio to SiC. Photolithography
is used for all pattern transfer processes. A Ni mask is patterned for recessing
down to the high doped access layer (Fig. 4.3 a). Recesses are formed for ohmic
contacting and partial device isolation, originally two separate etch steps in
[E]. This merging is possible when the high doped access layer thickness is
thinner than the depletion layer. Isolation is only partial, as the full isolation
is completed during the subsequent self-aligned Schottky mesa etching step. A
sacrificial oxide is grown by thermal dry oxidation at 1240 $^\circ$C under the flow of
$N_2O$ gas for 6 hours. This results in approximately 40 nm of $SiO_2$, consuming

---

**Figure 4.3:** Cross-section view of device fabrication steps: a) ohmic
recess and partial isolation etch, b) sacrificial dry oxidation, c) ohmic
contact definition and annealing, d) Schottky contact definition and
ohmic contact protection, e) self-aligned Schottky mesa etch and iso-
lation completion.
20 nm of SiC (Fig. 4.3 b). A precleaning using trichloroethylene (TCE) is used to release Cl$^-$ ions into the quartz furnace, binding detrimental ions such as Na$^+$ and K$^+$, which improves the oxide quality by reducing the number of near interface traps [223, 224]. Nickel metal contacts (1200 Å) are patterned to overlap the ohmic recesses by 1 µm (Fig. 4.3 c), following a 1 min HF dip. Annealing in an Ar atmosphere for 5 min at 1000°C results in ohmic contacts. Schottky contacts (500 Å/200 Å/2000 Å/700 Å Ni/Pt/Au/Ni) are then evaporated (Fig. 4.3 d), where nickel with its higher metal work function substitutes the titanium Schottky in [E] to reduce the leakage current [221]. Prior to the Schottky metal evaporation another 1 min HF dip is performed. It is reported that this sacrificial oxide method can nearly unpin nickel on 4H-SiC [225]. The Schottky metal stack is simultaneously evaporated on top of the annealed ohmic contacts as etch protection, done in a separate evaporation step in [E]. The entire chip is etched by RIE, with the contact metal stack Ni top-layer working as an etch mask (Fig. 4.3 e). This results in self-aligned Schottky contact mesas with sharp side-walls, completing the partial device isolation commenced earlier in the process. Self-aligned mesas are intended as an edge termination technique, confining the electric fields laterally, preventing locally higher fields, and making the depletion region expand more one-dimensionally. A side-effect of the process is the unconventional topology surrounding the ohmic contacts. To avoid yield issues evaporation of the 1$^{st}$ metal layer in [E,F] is best skipped as the device topography makes lift-off difficult at this stage. Electrodes, pads, and air-bridges to the Schottky anode contacts are finally electroplated with 3 µm gold.

### 4.4 Layout and physical device modeling

Devices fabricated in this work typically feature an interdigitated layout, where Schottky (anode) and ohmic (cathode) fingers are interleaved in order to reduce the parasitic series resistance. A cross-section view of a single device Schottky finger, with one of its adjacent ohmic fingers visible, is shown in Fig. 4.4. All major device layout dimensions and semiconductor resistance components have been indicated. The unique topology surrounding the ohmic contact is the result of the self-aligned fabrication process.

The bias dependent drift resistance through the undepleted part of the epitaxial depletion layer is calculated using a conductivity model for SiC,

\[
R_{drift}(V) = \frac{1}{W \cdot L} \int_{t_{depl}}^{t_{j}(V)} \frac{1}{\sigma(N_D(x), T, \theta)} dx, \tag{4.5}
\]

where \( W \) is the width of the Schottky finger, \( L \) the length of the finger, \( t_{depl} \) the thickness of the depletion layer with doping profile \( N_D(x) \), and \( \sigma(N_D, T, \theta) \) the doping, temperature, crystal direction dependent SiC conductivity. Specifically, the conductivity model in [226] is used together with the mobility model in [211].

Incomplete ionization is an important consideration in wide bandgap physical modeling. The donor (acceptor) impurity energy levels are typically much further from the conduction (valence) band than in narrow bandgap materials. Given that the required excitation energy is large for these deep levels,
only a fraction of the impurity atoms are ionized at room temperature. This incomplete ionization decreases the number of free electrons, which must be considered when modeling the material conductivity [226].

The parasitic access resistance in the high doped buried layer is not calculated by a conductivity model due to degenerate doping, but from sheet and ohmic contact resistances extracted from TLM measurements, and is given by,

\[
R_{\text{access}} = \left( \frac{R_{\text{contact}}}{L} + \frac{R_{\text{sheet}}}{L} \left( D + \frac{W}{6} \right) \right) / 2, \tag{4.6}
\]

where \( R_{\text{contact}} \) is the ohmic contact resistance (\( \Omega \text{mm} \)), \( R_{\text{sheet}} \) the sheet resistance of the high doped buried layer (\( \Omega \text{/sq} \)), and \( D \) the anode-cathode contact spacing. The spreading resistance under the Schottky mesa extends the effective anode-cathode distance [227]. Etching is assumed perfect, i.e. \( t_{\text{etched}} = 0 \) and \( R_{\text{etched}} = 0 \). Given two adjacent ohmic fingers, the access resistance is halved by the factor of 2.

The electroplated (gold) electrodes running along the length of the fingers are not shown in Fig. 4.4. Metallization resistance, assuming a uniform current distribution [G], is given by,

\[
R_{\text{metal}}(\omega) = \frac{L}{t_{\text{metal}} \cdot \sigma_{\text{metal}}} \left( \frac{f(\omega, W)}{2W} + \frac{f(\omega, W_{\text{ohm}})}{4W_{\text{ohm}}} \right), \tag{4.7}
\]

where \( \sigma_{\text{metal}} \) is the metal conductivity, \( t_{\text{metal}} \) the metal thickness, and \( f(\omega) \) describes frequency dependent skin depth and proximity effects [228]. Only the frequency dependent skin depth effect is considered here.

The total varactor series resistance is therefore both bias and frequency dependent,

\[
R_s(V, \omega) = \frac{R_{\text{drift}}(V) + R_{\text{access}} + R_{\text{metal}}(\omega)}{N}, \tag{4.8}
\]

where \( N \) is the number of Schottky fingers.
4.5 Characterization

Four-point resistance measurements on dedicated TLM structures allows for extraction of sheet and contact resistances. Despite excellent tailored epitaxy the sheet resistivity of the LiU material [E] is non-uniform with \( R_{\text{sheet}} \approx 380-1400 \Omega/\text{sq} \) and \( R_{\text{contact}} \approx 0.3-0.5 \Omega \text{mm} \). The uniformity of the Norstel material [F] is excellent, with \( R_{\text{sheet}} \approx 100 \Omega/\text{sq} \) over an entire 16 mm x 16 mm chip. Contact resistivity on this material is not reliably extracted, however, due to the optimized varactor fabrication process making the realized TLM structures non-ideal [F].

DC measurements in the reverse and forward directions for a single-finger 5 \( \mu \text{m} \times 400 \mu \text{m} \) device [E] is shown in Fig. 4.6. The Schottky ideality factor is close to unity. It is important that the leakage current remains low, given the high voltage operation. The DC power dissipation should be insignificant to not cause loss or affect long-term device reliability. The breakdown of this particular device is \(-160 \text{V}\). A breakdown study of 17 different devices from [E] is shown in Fig. 4.7. Although typically in excess of \(-100 \text{V}\), in this early batch significant spread in the breakdown voltage is observed.

Auto-probed DC measurements of the latest batch of devices on Norstel [F] material is shown in Fig. 4.8. The figure shows the leakage current at
CHAPTER 4. SIC VARACTORS FOR DYNAMIC LOAD MODULATION

Figure 4.7: Breakdown voltage distribution (material [E]).

Figure 4.8: Leakage current auto-probed across an entire chip (material [F]), where 90% of the devices have a leakage current less than 0.5 mA at −100 V.

−100 V, both as mapped spatially over the chip (Fig. 4.8 (a)) and summarized in a histogram (Fig. 4.8 (b)). Devices with various layouts (5–14 μm wide with 1–14 fingers) are included on the chip, with layouts being identical along rows. These results therefore indicate no apparent correlation between device layout and yield. Device yield, defined as less than 0.5 mA of leakage current at −100 V, is 90%. A SEM image of a 14-finger $C_{\min} = 3$ pF device is shown in Fig. 4.9.

Devices are typically characterized in two-port series coplanar configuration by on-wafer vector network analyzer (VNA) measurements. In the low-GHz frequency range the devices can be represented by a Π-network [10], where capacitance, resistance and Q-factor are extracted from $-1/Y_{12}$. This is similar the procedure adopted in [G]. A comparison between identical layout LiU [E] and a Norstel [F] material devices is shown in Fig. 4.10. It is seen that the per-area capacitance of the Norstel material is higher than for the LiU material. This implies a reduction in intrinsic series resistance and a lower breakdown voltage. The device breakdown performance has yet to be characterized, but is nominally calculated to −150 V. From the corresponding plot
of normalized capacitance, the effective tuning range of the Norstel material is slightly reduced compared to the LiU material, as the 2:1 tuning bias has shifted from $-30\, \text{V}$ to $-26\, \text{V}$. The excellent access layer ($R_{\text{sheet}} = 100\, \Omega/\text{sq}$), however, approximately doubles the Norstel device Q-factor. Successful epitaxial engineering in both materials has distributed the C(V) in a non-abrupt way, enabling high effective tuning range. The Q-factor is competitive compared to other published junction varactors [E] and the effective tuning range (4.3) shown in Fig. 4.11 is the largest of the devices when the RF amplitude exceeds 5 V. This highlights that the shape of the C(V), not merely a high breakdown voltage (e.g. the GaN varactor [229]), is important for tunable microwave power applications. A comparison of varactor performance is given in Table 4.2.

In [F] island type layouts are evaluated and compared to interdigitated. The normal interdigitated layout divides the varactor current equally between two adjacent ohmic fingers, effectively halving the parasitic series resistance. On the same notion, the resistance is more than halved again by forming square Schottky mesa islands, interconnected by air-bridges, surrounded by ohmic contacts on all sides. This layout has a poorer active area utilization however (by a factor of 4). Devices with 14/7/5 $\mu$m wide Schottky contacts (total Schottky areas of 784/784/800 $\mu$m²), with 7/7/5 $\mu$m wide ohmic contacts, and a contact spacing of 6/4/4 $\mu$m, were fabricated in both finger and island layout forms. SEM images of the six layouts are shown in Fig. 4.12. Extracted C(V), Q-factors, and R(V) for the devices are summarized in Table 4.3. The island

Table 4.2: Comparison with other diode varactors.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Tuning Range</th>
<th>4:1/3:1/2:1</th>
<th>Punch- thr. (V)</th>
<th>Break- dn. (V)</th>
<th>$Q_{\text{min}} / Q_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[218] SiC 2 GHz 5.6:1</td>
<td>1 / 2 / 6</td>
<td>15</td>
<td>40</td>
<td>8 / 45</td>
<td></td>
</tr>
<tr>
<td>[230] GaAs 2 GHz 9:1</td>
<td>5 / 7 / 9</td>
<td>15</td>
<td>28</td>
<td>22 / 150</td>
<td></td>
</tr>
<tr>
<td>[E] SiC 2 GHz 6:1</td>
<td>8 / 18 / 30</td>
<td>60</td>
<td>150</td>
<td>20 / 160</td>
<td></td>
</tr>
<tr>
<td>[F] SiC 2 GHz 6:1</td>
<td>8 / 15 / 26</td>
<td>50</td>
<td>&gt;100</td>
<td>34 / 600</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.10: For identical devices on Norstel and LiU materials, Left: C(V), Middle: $C_{\min}$-normalized C(V), Right: Q-factor versus normalized C(V).

Figure 4.11: Effective tuning range of $[E]$ at fixed voltage swing, compared with other references: SiC [218], GaAs [230], GaN [229].
4.6 NONLINEAR VARACTOR CHARACTERIZATION

Figure 4.12: From the top row and down: SEM images of 14 μm-, 7 μm-, and 5 μm-wide finger and island devices of comparable Schottky contact area.

Table 4.3: Comparison between interdigitated finger (f) and island (i) type device layouts (Q-factor at 2 GHz)

<table>
<thead>
<tr>
<th>Device</th>
<th>$T_{\text{linear}}$</th>
<th>Q(0V)</th>
<th>Q(3:1)</th>
<th>$C(-70V)$</th>
<th>R(0V)</th>
<th>R(-70V)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>μF/μm²</td>
<td>μF/μm²</td>
<td>(pF/μm²)</td>
<td>(μm)</td>
<td>(μm)</td>
</tr>
<tr>
<td>14 μm (f)</td>
<td>5.6</td>
<td>18</td>
<td>35 (-16 V)</td>
<td>0.10</td>
<td>9.7</td>
<td>4.9</td>
</tr>
<tr>
<td>14 μm (i)</td>
<td>5.3</td>
<td>29</td>
<td>53 (-15 V)</td>
<td>0.12</td>
<td>5.7</td>
<td>1.9</td>
</tr>
<tr>
<td>7 μm (f)</td>
<td>5.6</td>
<td>33</td>
<td>63 (-15 V)</td>
<td>0.11</td>
<td>5.2</td>
<td>1.6</td>
</tr>
<tr>
<td>7 μm (i)</td>
<td>4.8</td>
<td>42</td>
<td>73 (-13 V)</td>
<td>0.14</td>
<td>3.7</td>
<td>0.7</td>
</tr>
<tr>
<td>5 μm (f)</td>
<td>5.5</td>
<td>34</td>
<td>64 (-15 V)</td>
<td>0.11</td>
<td>4.9</td>
<td>1.4</td>
</tr>
<tr>
<td>5 μm (i)</td>
<td>4.6</td>
<td>44</td>
<td>73 (-11 V)</td>
<td>0.16</td>
<td>3.1</td>
<td>0.5</td>
</tr>
</tbody>
</table>

layouts reduce the parasitic series resistance R(-70V) by ≈60%. The high R(0V)/R(-70V) ratios of some layouts suggest the devices are approaching intrinsic Q-factor performance. However, the parasitic parallel capacitance between air-bridges and ohmic contacts counteracts this reduction in series resistance and there is no Q-factor gain when scaling from the 5 μm to the 7 μm island layout. In fact, this scaling also decreases the allowable RF amplitude at the 3:1 tuning ratio from 13 V to 11 V.

4.6 Nonlinear varactor characterization

The varactors fabricated in this work are intended for use in high power circuits. However, the device characterization in Section 4.5 is limited to DC and S-parameter measurements. To study varactor operation under large-signal conditions, active source- and load-pull measurements are performed in [G]. Specifically this allows for the capture of device current and voltage waveforms at the ports under varying power and circuit loading conditions. This information can be used for circuit design purposes and to study effects of harmonic generation and power-loss by nonlinear C(V) and R(V). These measurements may also serve as a verification of nonlinear device models.
In [G] the main focus is to determine the Q-factor under nonlinear distortion and to explore an alternative way of extracting the bias dependent drift layer resistance. A modified version of the active load-pull system in [132] is used for the measurements. An LSNA (Maury/NMDG MT4463) measures the harmonic content of incident \((a_1, a_2)\) and reflected \((b_1, b_2)\) voltage waves in a calibrated reference plane at the probe tips. A schematic of the measurement system is shown in Fig. 4.13. A single signal source is used to generate the fundamental frequency \((f_0)\). This signal is then split and a doubler is used to generate the second harmonic signals \((2f_0)\), with perfect phase coherence. Fundamental and second harmonic signals are combined by filters and injected outside the reflectometers. Vector modulators, capable of controlling amplitude and relative phase of the injected signals, are used to produce desired harmonic load impedances at the ports. Synthesizing the port-1 (source) and port-2 (load) impedances is an iterative procedure handled by an optimization routine to simultaneously satisfy,

\[
\Gamma_{1,2} = \frac{a_{1,2}}{b_{1,2}}, \quad \Gamma_{2,1} = \frac{a_{2,1}}{b_{2,1}}, \quad \Gamma_{2,2} = \frac{a_{2,2}}{b_{2,2}},
\]  

where \(a_{1,2}\) and \(b_{1,2}\) are the port 1 second harmonic, \(a_{2,1}\) and \(b_{2,1}\) the port 2 fundamental harmonic, and \(a_{2,2}\) and \(b_{2,2}\) the port 2 second harmonic voltage waves. A harmonic impedance environment can thereby be presented to the DUT while measuring the port voltage and current waveform response.

A single finger coplanar abrupt SiC varactor, fabricated on a uniformly doped Cree material, is characterized in [G]. An equivalent circuit topology for the device under reverse bias is shown in Fig. 4.14. Short-circuiting port 2 at the fundamental harmonic, using the active load-pull system, eliminates the parasitic pad capacitance \(C_{p2}\). The fundamental varactor current \(I_{v,1}\) then equals the negative fundamental port 2 current \(I_{2,1}\), effectively eliminating \(C_{p1}\).

\[
Z_{eq} = \frac{V_{1,1}}{-I_{2,1}} = R_{eq} + \frac{1}{j\omega_0 C_{eq}},
\]

where \(V_{1,1}\) is the port 1 fundamental harmonic voltage, \(R_{eq}\) and \(C_{eq}\) are the equivalent resistance and capacitance, respectively. At low power \(R_{eq}\) is should converge to the drift layer resistance \(R_{drift}(V)\) in (4.5).

S-parameter measurements are used to extract the series inductance \(L_s\), while \(C_{coupl}\) and \(R_{sch}\) are assumed negligible. A Q-factor expression valid in
4.6. NONLINEAR VARACTOR CHARACTERIZATION

The presence of nonlinear distortion is derived,

\[ Q = \text{Im}\left(\frac{V_{1,1}I_{2,1}^*}{2} + j\omega_0 L_s |I_{2,1}|^2\right)/\text{Re}\left(\frac{V_{1,1}I_{2,1}^*}{2}\right). \]  (4.11)

This expression encompasses the linear Q-factor expression in (4.1) when the available power is low. However, as the available power increases more power is lost to the generation of higher order harmonics. The total varactor power loss therefore equals, based on energy conservation, the sum of resistive power loss and power delivered out of the varactor at all higher order harmonics,

\[ P_{\text{loss}} = P_{R_{s,1}} + \sum_{n=2}^{\infty} \left(P_{R_{s,n}} + \sum_{m=1}^{M} |\text{Re}(P_{m,n})| \right), \]  (4.12)

where \( m \) is the port index, \( M \) the number of ports, \(|\text{Re}(P_{m,n})|\) the delivered \( n^{th} \) harmonic power at the \( m^{th} \) port. \( P_{\text{loss}} \) must equal the delivered fundamental power \(|\text{Re}(P_{1,1})|\) at port 1. The resistive power loss is given by,

\[ P_{R_{s,n}} = \frac{1}{2} R_{s,n} |I_{v,n}|^2, \]  (4.13)

where \( I_{v,n} \) is the \( n^{th} \) harmonic current through the varactor, and \( R_{s,n} \) is the \( n^{th} \) harmonic frequency (e.g. skin depth [228]), bias, and power dependent series resistance component. The harmonic varactor currents \( I_{v,n} \) are the result of the circuit loading at (possibly) all higher order harmonics, e.g. due to the harmonic generation and mixing by the device nonlinearities. Using the active source- and load-pull system it is therefore possible to study the Q-factor dependence on the second harmonic loading conditions.

The device was fundamentally short-circuited at port 2, with higher harmonic port impedances set by the system (\( \approx 50\Omega \)). Equivalent resistance and capacitance versus bias at 3GHz, when sweeping the available power, are shown in Fig. 4.15. Both \( C_{eq} \) and \( R_{eq} \) are seen to increase with increasing fundamental voltage amplitude, especially when the varactor is biased in the most nonlinear region of the \( C(V) \). This leads to an increase in the generation of higher order harmonics and the Q-factor decreases. This type of capacitance shift can also lead to circuit detuning [207].

The results in Fig. 4.15 are specific to the loading at all harmonics. An equivalent circuit of the varactor at the second harmonic, pumped at the fundamental, is shown in Fig. 4.16. The intrinsic varactor junction is loaded by \( L_s, R_{s,2} \), and the pad capacitances in parallel with the controllable second harmonic port impedances. Sweeping \( \angle(\Gamma_{L,2}) = 1 \) (see Fig. 4.16) results in
CHAPTER 4. SiC VARACTORS FOR DYNAMIC LOAD MODULATION

Figure 4.15: $C_{eq}$ and $R_{eq}$ versus bias for different voltage amplitudes.

Figure 4.16: Equivalent varactor circuit at the second harmonic.

the $C_{eq}$ and Q-factors shown in Fig. 4.17 (left). At 158° the equivalent second harmonic frequency resonates, causing the Q-factor to drop. When the varactor is tuned, the resonance phase will change. In Fig. 4.17 (right) all values of $\Gamma_{L,2}$ causing resonance are highlighted. This device specific region should be avoided in tunable network design.

Varactors feature inherently nonlinear behavior affecting tunable network performance. In DLM PAs there is reason to examine bias and load networks to avoid degrading resonant effects. The low loss and high power tunable SiC varactor technology presented in this chapter, combined with an understanding of nonlinear behavior, are contributions towards improved frequency agility and efficiency in high frequency systems and wireless transmitters.

Figure 4.17: Left: Q-factor and $C_{eq}$ versus purely reactive $\Gamma_{L,2}$. Right: device $\Gamma_{L,2}$ region to avoid in a tunable network at 3 GHz.
Chapter 5

Conclusions

This thesis work contributes theory and technology to the development of wideband and high efficiency transmitters to meet future wireless capacity demand at reduced cost and environmental impact.

There are different ways to implement efficiency enhanced transmitters, each with different merits depending on application constraints. In this thesis different solutions, such as wideband resistive loading suitable for envelope tracking, varactor-based dynamic load modulation, and dual-RF input active load modulation, are evaluated and developed. Common to all is the use of simplified transistor modeling to enable understanding of the operational principles. This is especially important in order to determine how key transistor technology parameters affect amplifier performance, such as operating frequency, output power, efficiency, and bandwidth. Indeed, although the demonstrators are realized in wideband and high power GaN HEMT technology, translation of results to a different technology or specification should therefore be straightforward.

High power microwave frequency DLM amplifiers, involving an interplaying varactor device, introduces an additional semiconductor technology challenge. Development of a suitable varactor in wide bandgap SiC, by tailored epitaxial design, enables the realization of a device with low loss and large effective tuning range. Theoretical analysis of transistor class-J DLM operation exposes the relationship between transistor technology and tunable load network requirements. This is a step away from typical empirical design, as load network topologies can be proposed in a power-scalable and bandwidth conscious way. It clearly appears that varactor effective tuning range critically limits amplifier output power and frequency tunability. Further, the demonstrators in this work successfully prove the scalability the of DLM concept towards macro base station power levels. These results also highlight that limited Q-factor of components other than the varactor in the load network, e.g. transmission lines or inductors, may limit the final efficiency enhancement.

Dual-RF input PA transmitters can be made both wideband and highly efficient. The additional degree of freedom offered by two independent RF-inputs, compared to a single-ended implementation, enables optimum output current shaping to ensure efficient active load modulation over large bandwidths. Ignored in common theoretical treatments, nonideal higher harmonic conditions
and transistor parasitics are major obstacles to realize truly wideband implementations. Optimizing a dual-input PA in a nonlinear circuit simulator is difficult, given the large number of input conditions, operating frequencies, bias conditions, and combiner parameters to evaluate. However, the simplified transistor models and linear multi-harmonic calculations presented in this thesis work facilitate such optimization. The validity of this approach is proved by the realization of an unprecedented back-off efficiency enhanced dual-input GaN demonstrator with 100% fractional bandwidth.

This thesis presents important contributions to improve the performance of wireless systems. Although the focus has been on base station transmitters, the results are clearly applicable to related wireless applications such as radar, sensing, or jamming. The varactor developments in particular are relevant to all systems that may benefit from reconfigurability, including receivers.

5.1 Future work

DLM transmitters present challenges and research opportunities. Theoretical development of frequency reconfigurability is an important next step. Another ambitious goal, supported by preliminary simulations, could be to extended the high efficiency (>50% PAE) OPBO dynamic range towards 10–13 dB by the minimization of load network losses. Wider signal bandwidths, approaching 100 MHz LTE-A, is another potential target which undoubtedly will require focusing on linearity and bias network design.

There are plenty of opportunities in fundamental varactor research. Specific to the SiC varactor in this work, studies on the effectiveness of self-aligned edge termination, stress testing, long-term reliability, and layout scaling studies are of interest. The later may require development of more accurate Q-factor measurement methods (for $Q \gtrsim 100$). With relatively high yield in the current process, fabrication of larger devices and stacked configurations are highly relevant. For non-dynamic applications, integration of mesa resistors for bias decoupling would be useful. The present varactor design is no global optimum and improved epitaxial design could, with the tools presented in this thesis, be carefully derived for a given DLM application.

Dual-RF input amplifiers show great potential. Realization of demonstrators for higher PAPR signals and the evaluation of alternative load network topologies are of interest. Suitable DPD methods should also be developed to prove the transmitter functionality. Related, digital bandwidth could possibly be introduced as a constraint in circuit optimization/post-processing.

There are other open transmitter research problems. Concurrent multi-band operation does not guarantee proper efficiency enhancement functionality, warranting fundamental architecture research. Digital processing power consumption affects the overall efficiency as the transmitter power level is downscaled. Reducing the digital load by improving raw architecture linearity is relevant. Further, to eliminate the need for output isolators, especially in future MIMO-type arrays, it might be possible to handle output mismatch in combination with DLM.
Acknowledgment

I want to thank the people that made this work possible:

My examiner Prof. Herbert Zirath for the opportunity at a PhD degree. It is a privilege to be a PhD student at MEL.

My supervisor Dr. Niklas Rorsman and co-supervisor Dr. Christian Fager for their open-mindedness, support and patience. It feels good to have combined both semiconductor technology and circuit research.

Prof. Jan Grahn for handling of IP and inviting me to MEL.

All co-authors through the years, especially David Gustafsson, Mustafa Özen, Dr. Mattias Thorsell, and Roberto Quaglia.

For growing the SiC materials: Prof. Anne Henry, Sven Andersson, and Prof. Erik Janzén at Linköping University. Björn Magnusson and Niklas Henelius at Norstel AB.

The open and generous people at the Mitsubishi Electric Co. Amplifier Group. In particular Dr. Koji Yamanaka, Dr. Masatoshi Nakayama, and Dr. Yoshihito Hirano, for supervision and the internship opportunity at the Information Technology R&D Center.

All co-workers making it fun to go to work and for helping me at one time or another (or many).

Dedicated to Miyu, my friends and family.

This work has been performed as part of several projects:

Microwave Wide Bandgap Technology project financed by Swedish Governmental Agency of Innovation Systems (VINNOVA), Swedish Energy Agency (STEM), Chalmers University of Technology, Ericsson AB, Furuno Electric Co., Ltd., Infineon AG, Norse Semiconductor Laboratories AB, Norstel AB, NXP Semiconductors BV, and Saab AB.

"Advanced III-Nitrides-based electronics for future microwave communication and sensing systems” financed by the Swedish Foundation for Strategic Research (SSF).

"ACC” and "EMIT” within the GigaHertz Centre, joint projects financed by VINNOVA, Chalmers University of Technology, Ericsson AB, Infineon Technologies Austria AG, NXP Semiconductors BV, Saab AB, Comheat Microwave AB, Mitsubishi Electric Corporation, SP Technical Research Institute of Sweden, and United Monolithic Semiconductors.

My internship at Mitsubishi was supported by grants from the Sweden-Japan Foundation and Chalmerska Forskningsstiftelsen.
Bibliography


