

DISSERTATION

Charge Trapping and Single-Defect Extraction in Gallium-Nitride Based MIS-HEMTs

zur Erlangung des akademischen Grades

Doktor der technischen Wissenschaften

*eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik
von*

ALEXANDER GRILL

Matrikelnummer 0125698

Anton Baumgartner Straße 125/8/10
1230 Wien, Österreich

Betreuer: UNIV.PROF. DIPL.-ING. DR.TECHN. TIBOR GRASSER

Wien, im Oktober 2018

Abstract

Charge Trapping and Single-Defect Extraction in Gallium-Nitride Based MIS-HEMTs

The trend to use electricity as a primary energy source will ever increase as major sectors like the individual and public transport are going to be more and more electrified. To meet the international goals on carbon emission, not only the production and consumption but also the conversion efficiency of electricity is going to play a crucial role. The ever-growing interest in wide-bandgap semiconductors like gallium nitride (GaN) or silicon carbide (SiC) is primarily driven by their fundamental material properties, which allow building much more efficient power conversion systems when compared to silicon technology.

GaN-based transistors, commonly called high-electron-mobility transistors (HEMTs) or metal-insulator-semiconductor HEMTs (MIS-HEMTs), are among the most attractive candidates for novel power semiconductor device concepts because of their high breakdown voltage as well as their low specific on-resistance. Therefore, this work starts with an introduction of the material system, and state-of-the-art concepts for normally-on and normally-off devices. Although first devices are already commercially available, severe reliability issues – usually related to charge trapping – are still preventing them from a more widespread acceptance.

To improve the reliability of future GaN technology, a detailed physical understanding of the origin and the effects of these defects is of utmost importance. The focus of this work lies on the characterization and modeling of the defects responsible for bias temperature instability (BTI) degradation in GaN MIS-HEMTs, which usually suffer from large V_{th} instabilities already at nominal operating conditions. Therefore, well-established methods for BTI characterization and reliability modeling in silicon need to be checked carefully for their applicability to GaN technology.

One important phenomenon which is usually not considered in BTI studies for silicon technology is the electrostatic feedback of the trapped charges. Highlighting the importance of charge feedback effects on the observed BTI degradation of GaN/aluminium gallium nitride (AlGaN) MIS-HEMTs and its impact on defect modeling is one of the main findings presented in this thesis.

Another focus is the development of more robust methods for the calculation of single-defect parameters from random telegraph noise (RTN) measurements in nano-scale GaN/AlGaN fin-MIS-HEMTs. The presented methods are used to obtain the characteristic time constants from the stochastic charge capture and emission events of RTN producing defects. Repeating those extractions for different bias conditions and temperatures allows to calculate other defect parameters like the vertical defect position or the trap level by assuming the transitions being governed by the non-radiative multi-phonon (NMP) theory.

The insights into the physics of charge feedback caused by BTI degradation in large-area devices and the methods provided for the extraction of single-defect parameters from GaN/AlGaN MIS-HEMTs can be seen as a first step towards the identification of the microscopic defect structures responsible for BTI degradation in GaN technology. A promising path towards this goal would be the comparison of measurements on large-area and nano-scale devices with technological computer aided design (TCAD) and first-principle simulations. Therefore, this thesis can be a valuable contribution to future improvements of GaN technology.

Kurzfassung

Ladungsträgereinfang und Einzeldefektextraktion in Gallium-Nitrid Basierten MIS-HEMTs

Wesentliche Sektoren wie der Individualverkehr und der öffentliche Verkehr werden in Zukunft immer stärker elektrifiziert werden, deshalb wird auch der Trend Elektrizität als Primärenergiequelle zu nutzen zunehmen. Um die internationalen CO₂-Emissionsziele zu erreichen, werden nicht nur Produktion und Verbrauch, sondern auch die Umwandlungseffizienz von Elektrizität in Zukunft eine entscheidende Rolle spielen. Das ständig wachsende Interesse an Halbleitern mit großer Bandlücke, wie zum Beispiel Gallium-Nitrid (GaN) oder Silizium-Karbid (SiC), ist hauptsächlich durch ihre fundamentalen Materialeigenschaften begründet. Sie ermöglichen im Vergleich zur Siliziumtechnologie die Herstellung von wesentlich effizienteren Energieumwandlungssystemen.

GaN-basierte Transistoren, oft als HEMTs oder MIS-HEMTs bezeichnet, gehören zu den vielversprechendsten Kandidaten für neuartige Konzepte von Leistungshalbleitern, da sie eine hohe Durchbruchspannung bei gleichzeitig niedrigem Einschaltwiderstand aufweisen. Deswegen beginnt diese Dissertation mit einer Einführung in das Materialsystem und einer Erklärung der modernsten Konzepte für selbstleitende und selbstsperrende Transistoren. Gravierende Zuverlässigkeitsprobleme – üblicherweise verbunden mit dem Einfang von Ladungen in Defekten – verhindern derzeit eine noch größere Marktverbreitung, obwohl einige Bauteile bereits im Handel erhältlich sind.

Ein detailliertes physikalisches Verständnis des Ursprungs und der Auswirkungen dieser Defekte ist von äußerster Wichtigkeit um die Zuverlässigkeit zukünftiger GaN-Technologien zu verbessern. Der Schwerpunkt dieser Arbeit liegt daher auf der Charakterisierung und Modellierung von Defekten, die für die Degradation der Schwellspannung in GaN verantwortlich sind. Bereits unter normalen Betriebsbedingungen leiden diese Bauelemente oft schon unter großen Instabilitäten. Daher müssen auch etablierte Methoden für die Charakterisierung und Zuverlässigkeitsmodellierung von Defekten in der Siliziumtechnologie nochmals sorgfältig auf ihre Anwendbarkeit auf die GaN-Technologie überprüft werden.

Die elektrostatische Rückkopplung von eingeschlossenen Ladungen ist ein wichtiges Phänomen, das bei Studien über die Bias-Temperaturinstabilität in der Regel nicht berücksichtigt wird. Die Bedeutung dieser Effekte für die Degradation von GaN/AlGaN MIS-HEMTs und ihre Auswirkungen auf die Modellierung von Defekten ist daher eines der wichtigsten Ergebnisse dieser Arbeit. Ein weiterer Schwerpunkt liegt auf der Entwicklung von robusteren Methoden zur Berechnung der Parameter von Einzeldefekten in GaN/AlGaN Fin-MIS-HEMTs. Die vorgestellten Methoden werden anschließend auch verwendet um die charakteristischen Zeitkonstanten aus den stochastischen Ladungseinfang- und Emissionsereignissen zu erhalten. Die Auswertung der Messungen bei verschiedenen Spannungen und Temperaturen ermöglicht es auch, andere Parameter wie die vertikalen Defektpositionen oder das Energieniveau der Defekte zu berechnen. Hierbei wird angenommen, dass die Übergänge durch die NMP-Theorie beschrieben werden.

Die Physik der Ladungsrückkopplung durch Degradation und die Extraktion von Parametern aus Einzeldefekten kann als ein erster Schritt zur Identifizierung jener mikroskopischer Defekte gesehen werden, welche für Degradation in GaN-Technologien verantwortlich sind. Ein vielversprechender Ansatz für zukünftige Studien ist der direkte Vergleich von Messungen an großflächigen und nano-Bauelementen mit TCAD und First-Principle-Simulationen. Daher können die Ergebnisse dieser Dissertation einen wertvollen Beitrag zur zukünftigen Verbesserung von GaN-Technologien liefern.

Acknowledgements

First and foremost, I want to thank **Prof. Tibor Grasser** for teaching me how to become a researcher and for his continuous support throughout my PhD program. You surely are the most diligent and passionate researcher I ever met. You showed me that looking at the precise details can be the truly important part of research, and your talent to ask the right questions inspired large parts of this work. As my boss, I always knew that I could count on your support, even in situations which were very delicate for both of us.

Furthermore, I want to express my gratitude to be able to work in such a great place like the **Institute for Microelectronics** during the last years. Working in such a professional and enjoyable environment really helped me to carry on every time I was frustrated. Special thanks go to **MICHAEL WATL**: Without your efforts, none of the experimental single-defect studies would have been possible. I also want to thank **BERNHARD STAMPFER** for his great work on band-interactions within the **NMP** theory and for helping me to understand **RTN** and how to extract data from it.

I want to thank all of **my colleagues** for the valuable technical discussions but also for the distraction and the free time we were able to spend together. Thank you **WOLFGANG Gös** and **FRANZ SCHANOVSKY** for providing the basics to make my first months at the institute considerably easier. Thank you **MARKUS JECH**, **ALEXANDER MAKAROV**, **GERHARD RZEPA**, **STANISLAV TYAGINOV**, **BIANKA ULLMANN**, and **YANNICK WIMMER** for being also very good friends.

Without the support of **Infineon Villach** and the studies on **GaN** from **CLEMENS OSTERMAIER**, **ROBERTA STRADIOTTO**, and **PETER LAGGER**, this work would not have been possible at all. Thank you very much for sharing your data and also helping me to interpret them correctly. I also want to acknowledge the great work of **Global TCAD Solutions**, which provided the simulation platforms necessary for my investigations. The fast and unbureaucratic support of **CHRISTIAN KERNSTOCK**, **FERDINAND MITTERBAUER**, and **ZLATAN STANOJEVIĆ** really helped to keep my research on track. Last but not least, I want to express my gratitude to **KI-SIK IM** and **JUNG-HEE LEE** from **Kyungpook National University**, Daegu, Korea for providing the samples for my single-defect studies in **GaN**.

A big thank you also goes to all of **my friends** for their emotional support throughout the last years. Unfortunately, there is not enough space here to mention everyone, but I shall thank every single one of you in person. The good memories we share always help me to get back on track when I struggle.

Last but not least, my deepest gratitude goes to **my family** for supporting me in any decision of my life and for always standing behind me, no matter what happens. I doubt that I can find the right words to express how important all of you have been in order to reach this point.

Dear **RENATE**, **ALOIS**, **CHRISTINA**, **RONALD**, **CHRISTIAN**, and **SARAH**:
Thank you for all your support!

Contents

Abstract	iii
Kurzfassung	v
Acknowledgements	vii
Acronyms	xiv
1 Introduction	1
1.1 Fundamentals of Power Semiconductor Devices	1
1.1.1 Overview	2
1.1.2 Figures of Merit	2
Baliga's Figure of Merit	3
Johnson's Figure of Merit	3
1.1.3 Comparison of Semiconductor Materials	3
1.2 Motivation and Outline	5
1.2.1 Motivation	5
1.2.2 Outline	6
2 Group III-Nitride Materials and Devices	7
2.1 Crystal Structure and Growth	7
2.2 Electronic Material Parameters	9
2.2.1 Band Structure Parameters	9
2.2.2 Transport Properties	10
2.2.3 Alloy Parameters	12
2.2.4 Polarization Charges	12
2.3 Fundamentals of GaN/AlGa _N Power Devices	14
2.3.1 Formation of the 2DEG	15
2.3.2 Depletion Mode High-Electron Mobility Transistors	17
2.3.3 Normally-Off Device Concepts	17
Barrier Recess Etch	18
Fluorine Implantation	18
Gate Injection Transistor	19
3 Reliability Issues in GaN/AlGa_N HEMTs	21
3.1 Defects in GaN/AlGa _N Devices	21
3.1.1 Bulk Defects	22
3.1.2 Interface Defects	23
3.2 Degradation Mechanisms in GaN/AlGa _N HEMTs	24
3.2.1 Current Collapse	24
3.2.2 Threshold Voltage Drift	25
3.2.3 Hot Carrier Degradation	26

4	Experimental Considerations	29
4.1	BTI Measurement Techniques	29
4.1.1	Hysteresis Measurements	30
4.1.2	Conductance Frequency Method	30
4.1.3	Measure-Stress-Measure Method	30
	Extended Measure-Stress-Measure Method	31
4.1.4	On-The-Fly Method	33
4.2	Single Defect Characterization	33
4.2.1	Random Telegraph Noise	33
4.2.2	Time Dependent Defect Spectroscopy	34
5	Defect Modeling	37
5.1	Capture-Emission Time Maps	37
5.2	Non-Radiative Multi-Phonon Model	39
5.2.1	Theoretical Background	39
5.2.2	Oxide Defects	40
	The NMP Four-State Model	40
5.2.3	Local Band Interactions	45
6	Extraction of the Characteristic Time-Constants	49
6.1	Markov Processes and the Hidden Markov Model	49
6.2	Two State Defects	51
6.3	Multi State Defects	53
6.4	Multiple Defects	54
6.5	Coupled Defects	56
6.6	Histogram-Based Methods	57
6.6.1	Amplitude Histograms	58
6.6.2	Lag Plots	59
6.6.3	Spectral Maps	60
	Step Detection and Clustering	60
	Time Constants Extraction	62
	Problems	63
6.7	Hidden Markov Model Training	64
6.7.1	Forward Algorithm	65
6.7.2	Backward Algorithm	66
6.7.3	Parameter Update	67
	Multiple Defects	69
	Multiple Sequences	70
	Calculation of the Defect Distributions	71
6.7.4	Baseline Estimation	72
	Local Regression	74
	Basis Splines Smoothing	75
	Asymmetric Least Squares Smoothing	76
6.7.5	Benchmarks	78
6.8	Conclusions	83
7	Charge Trapping Effects in GaN/AlGaN HEMTs	85
7.1	Threshold Voltage Drift in GaN/AlGaN MIS-HEMTs	85
7.1.1	Electrostatic Device Simulation	86
	Piezoelectric Charges	86
	Bulk Doping and Barrier Transport	87

7.1.2	Forward Bias Threshold Voltage Drift	88
7.1.3	Defect Feedback Mechanisms	89
	Electrostatic Charge Feedback	91
	Local Potential Perturbations	93
7.1.4	Conclusions	93
7.2	Single Defect Characterization in GaN/AlGa _N MIS-HEMTs	95
7.2.1	Motivation	95
7.2.2	Electrostatic Device Simulation	96
7.2.3	Extraction using Spectral Maps	97
7.2.4	Electrostatic Coupling of Defects	104
7.2.5	Extraction using the Hidden Markov Model	108
7.2.6	Conclusions	116
8	Conclusions and Outlook	119
8.1	Conclusions	119
8.2	Outlook	120
A	The Hidden Markov Model Library	123
	Bibliography	125
	List of Publications	139
	Curriculum Vitae	

Acronyms

V_{th}	threshold voltage
2DEG	two-dimensional electron gas
ADC	analog-digital converter
AFM	atomic force microscopy
Al	aluminium
AlGaN	aluminium gallium nitride
AlInN	aluminium indium nitride
AlN	aluminium nitride
arPLS	asymmetrically reweighted penalized least-squares
BFoM	Baliga's figure of merit
BTI	bias temperature instability
C	carbon
CC	configuration coordinate
CDF	cumulative distribution function
CET	capture emission time
DD	drift-diffusion
DFT	density functional theory
DG	density gradient
DLTS	deep level transient spectroscopy
EDMR	electrically detected magnetic resonance
EM	expectation maximization
EMF	electromotive force
eMSM	extended measure-stress-measure
ESR	electron spin resonance
Fe	iron
FET	field-effect transistor
Ga	gallium
GaAs	gallium arsenide
GaN	gallium nitride
H	hydrogen
HCD	hot-carrier degradation
HEMT	high-electron-mobility transistor
HFET	heterojunction field-effect transistor
HMM	Hidden Markov Model
IGBT	insulated gate bipolar transistor
In	indium
InGaN	indium gallium nitride
InN	indium nitride
JFET	junction field-effect transistor
JFoM	Johnson's figure of merit
LED	light emitting diode
LOWESS	locally weighted scatterplot smoothing
MAP	maximum a-posteriori
Mg	magnesium
MIS-HEMT	metal-insulator-semiconductor HEMT
MOS	metal-oxide-semiconductor
MoS ₂	molybdenum disulfide
MOSFET	metal-oxide-semiconductor field-effect transistor
MSM	measure-stress-measure
N	nitrogen
NBTI	negative bias temperature instability
NMP	non-radiative multi-phonon

O	oxygen
OTF	on-the-fly
PBTI	positive bias temperature instability
PDF	probability density function
RMS	root mean square
RTN	random telegraph noise
Si	silicon
SiC	silicon carbide
SiGe	silicon germanium
SILC	stress induced leakage current
SiN	silicon nitride
SiO ₂	silicon dioxide
SNR	signal to noise ratio
STM	scanning tunneling microscopy
TAT	trap-assisted tunneling
TCAD	technological computer aided design
TDDS	time-dependent defect spectroscopy
TEM	transmission electron microscopy
TMI	TDDS measurement instrument
UID	unintentionally doped
WKB	Wentzel–Kramers–Brillouin

Chapter 1

Introduction

During the last four decades, the world's total electricity consumption has dramatically increased from 6.1 PWh in 1973 up to 24.3 PWh in 2015, which is an overall increase of almost 400 % [1]. Despite considerable efforts to promote renewable energy sources (hydro, wind, solar, geothermal and others) during the past two decades, their total market share has increased only marginally from 21.5 % to 23.1 %. At the same time, the amount of electricity in the global energy consumption has almost doubled from 9.4 % to 18.5 %.

If the international goals to limit carbon dioxide emissions should be met, the trend to use electricity as a primary energy source will ever increase as sectors as for example individual and public transport are going to be more and more electrified in the near future. To significantly increase the amount of electricity from renewable sources, not only the production and consumption of electricity but also the conversion efficiency is going to play a crucial role in the future.

For that reason, there is an ever growing interest in creating more efficient power semiconductor devices, with a special focus on wide-bandgap materials like gallium nitride (GaN) and silicon carbide (SiC) because their fundamental material properties are far superior to silicon. To obtain a widespread acceptance of such devices, several technological issues have to be solved. Most importantly, the reliability of these devices is still a major concern in present technologies. For many device concepts, charge trapping in defects is a severe performance limiting factor which prevents manufacturers to actually exploit the superior material properties of SiC and especially GaN.

This chapter briefly explores the reasons for the ever-growing interest in wide-bandgap semiconductors by comparing them to other materials using various figures of merit. The comparison will mainly focus on power applications being one of the most promising applications in the future. The second part holds a brief motivation for this work, followed by a short description of the content and the focus of the different chapters.

1.1 Fundamentals of Power Semiconductor Devices

Power semiconductor devices are typically used to control the electrical power flow in switching mode power supplies. For that purpose, they are only operated under two different conditions, the on-state and the off-state. Unlike ideal switches which conduct current without any losses in the on-state and block any voltage in the off-state, power semiconductor devices consume energy in both of these states and can only block finite voltages before breaking down. In addition, real power switches also consume energy for switching between the on- and off-state. To make power

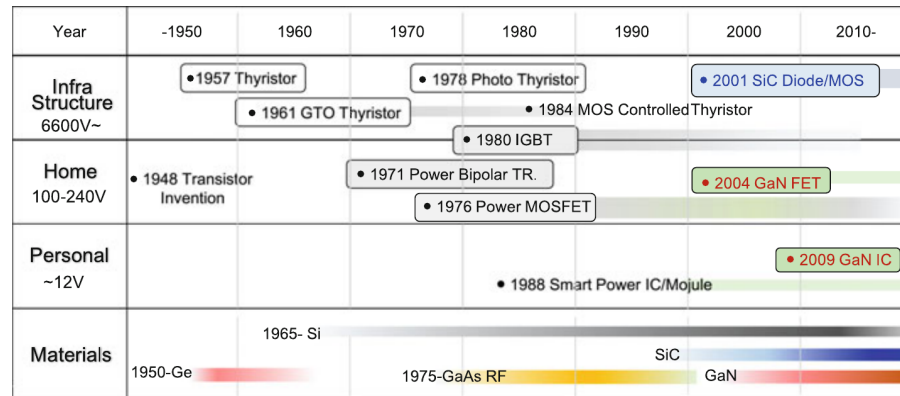


Figure 1.1: The evolution of power semiconductors for different applications for infrastructure, home, and personal usage. For decades silicon had been the predominant material, only recently competition started to emerge by the adoption of wide-bandgap materials (from [2]).

switches as efficient as possible, all of these three contributions to the overall losses need to be minimized.

1.1.1 Overview

Power conversion systems are used in a broad range of applications with nominal powers ranging from several watts (e.g., chargers for consumer electronics) to several gigawatts (e.g., high-voltage transmission systems for offshore wind parks). The applications are usually split into layers based on the required voltage range. The three major application layers, infrastructure for large voltages in the kV range, home for domestic power applications, and personal for integrated power conversion systems with low voltages and high currents, are shown in Figure 1.1. The personal layer has emerged recently, as the operating voltages of highly integrated circuits were scaled to values as low as 1 V for hundreds of millions of transistors. The usual application for this layer is a down-step converter from the battery voltage in a mobile device with tens of amperes of output current.

1.1.2 Figures of Merit

To understand the recent interest in wide-bandgap semiconductors, the main parameters of power devices and the trade-offs between them need to be discussed. The most important parameters of a power switching device are summarized as

- on-resistance R_{on} ,
- nominal current,
- breakdown voltage B_v ,
- switching capacitance.

Since all the parameters mentioned above strongly depend on the fundamental device concept and the device geometry, a fair comparison of different materials is not possible. The general suitability of different semiconductor materials for power applications should instead be calculated based on the fundamental electrical properties rather than device specific parameters. In the following, two important figures of merit for power semiconductors, namely Baliga's figure

of merit (BFoM) and Johnson's figure of merit (JFoM) are introduced and will later be used to compare different wide-bandgap semiconductors to silicon.

Baliga's Figure of Merit

For any material used in power devices, the specific on resistance as the main contributor to the conduction losses should be kept as low as possible with the breakdown voltage being as high as possible. Baliga's figure of merit combines these two parameters into a single figure of merit, which is the specific power density a given material can theoretically withstand. It is calculated with the permittivity ϵ_r of the material, the electron or hole mobility $\mu_{n,p}$, and the critical electric field strength for material breakdown E_c [3, 4].

$$\text{BFoM} = \epsilon_0 \epsilon_r \mu_{n,p} E_c^3 \text{ [W m}^{-2}\text{]} \quad (1.1)$$

Note that for a given material, a higher numbers means a lower specific on resistance and thus less conduction losses for a given technology. For unipolar devices in the one-dimensional limit, BFoM connects the specific on resistance of the device $R_{\text{on},\square}$ to the breakdown voltage B_v of the device.

$$R_{\text{on},\square} = \frac{4B_v^2}{\epsilon_0 \epsilon_r \mu_{n,p} E_c^3} \text{ [\Omega m}^{-2}\text{]} \quad (1.2)$$

Johnson's Figure of Merit

In modern switching mode power supplies, switching losses often make up for the largest contribution to the overall losses. Additionally, modern mobile base stations and radar applications requires power amplifiers in the microwave and millimeter-wave frequency range. For high-frequency applications, a basic trade-off between breakdown voltage and transit frequency is needed. This trade-off is reflected by JFoM [5]. In a first order approximation, the maximum transit frequency f_T of carriers through a metal-oxide-semiconductor field-effect transistor (MOSFET) channel is given by the channel length L and the saturation velocity for electrons or holes v_{sat} .

$$f_T = \frac{v_{\text{sat}}}{2\pi L} \quad (1.3)$$

Considering a uniform field across the channel, the breakdown voltage can easily be calculated from the critical field strength as $B_v = E_c l$. The product of these two parameters becomes constant and independent of the channel length. This constant is usually referred to as Johnson's limit or JFoM.

$$\text{JFoM} = B_v f_T = \frac{E_c v_{\text{sat}}}{2\pi} \quad (1.4)$$

1.1.3 Comparison of Semiconductor Materials

The interest in wide-bandgap materials can be understood best if their most important material properties are compared to those of other common semiconductors. In Table 1.1, the most fundamental material properties of common wide-bandgap semiconductors are listed together with those of silicon and gallium arsenide (GaAs) [2, 6–8]. Both figures of merit, BFoM for the conduction losses and JFoM for the switching losses show a large advantage for the wide-bandgap semiconductors when compared to silicon (Si) or GaAs. On top of that, the superior thermal conductivities offer the ability to ease some of the problems regarding the thermal design of devices.

Parameter	Si	GaAs	GaN	AlN	SiC(6H)	SiC(4H)
E_g [eV]	1.12	1.42	3.39	6.2	3.02	3.26
ϵ_r [1]	11.7	10.9	8.9	8.5	10	9.7
μ_n [$\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$]	1500	8500	1000	300	900	400
E_c [MV cm^{-1}]	0.3	0.55	3-5	1.2-1.8	3	3
v_{sat} [10^7 cm s^{-1}]	1	1.2	2.5	3.7	2	2
κ [$\text{W cm}^{-1} \text{K}^{-1}$]	1.5	0.46	2.1	2.85	2.3-4.9	2.3-4.9
BFoM [TW cm^{-2}]	0.42	13.6	504	7.62	215	92.8
JFoM [THz V]	0.48	1.05	15.9	8.24	9.55	9.55

Table 1.1: A comparison of different semiconductor materials used in power applications. Both figures of merit, **BFoM** for the conduction losses and **JFoM** for the switching losses show a large advantage for the material properties of wide-bandgap semiconductors when compared to silicon or gallium arsenide. (data from [2, 6–8])

scaling factor: $1/k$	
R_{on}	1
B_v	$1/k$
chip area	$1/k^2$
chip cost	$1/k^2$
capacitance	$1/k^2$
switching loss	$1/k^2$

Table 1.2: Scaling laws for different important technology parameters if the device dimensions are scaled by a factor of $1/k$. If the wide-bandgap devices are designed for the same voltage and current rating, an improvement of a factor of around 100 in chip area and switching losses can be obtained theoretically.

The improved critical field also helps to overcome the scaling limit for power conversion devices made of silicon. **Table 1.2** shows well-known scaling rules for essential parameters in silicon MOSFET technology for a scaling factor of $1/k$. Assuming a constant voltage design, the larger values of the critical field in wide-bandgap semiconductors offer quite some margin for further miniaturization. Note that the breakdown voltage scales linearly, while other key chip parameters like area, switching losses, and cost scale quadratically with the applied scaling factor.

The ratios between the breakdown fields of GaN or SiC and Si thus allow reducing chip area and switching losses by a factor of 100 while keeping roughly the same voltage and current ratings (neglecting thermal design). Of course, this is just a theoretical value as factors like the maturity of the technology, fabrication cost or process compatibility are often limiting factors preventing technologies from a broader market adoption.

To see the current status of realized GaN and SiC devices compared to silicon power devices, in **Figure 1.2** the values of R_{on} versus the breakdown voltage across different state of the art device concepts and technologies are shown [9]. Although the theoretical MOSFET limit of silicon can be stretched by using different concepts like super-junction transistors or insulated gate bipolar transistor (IGBT), the GaN and SiC devices clearly outperform silicon devices despite being an immature technology.

Overall, wide-bandgap materials like GaN and SiC offer superior material properties, both for high-power and high-frequency operation. In the high-power regime, this allows for significant improvements regarding the energy efficiency and required chip area due to scaling. The

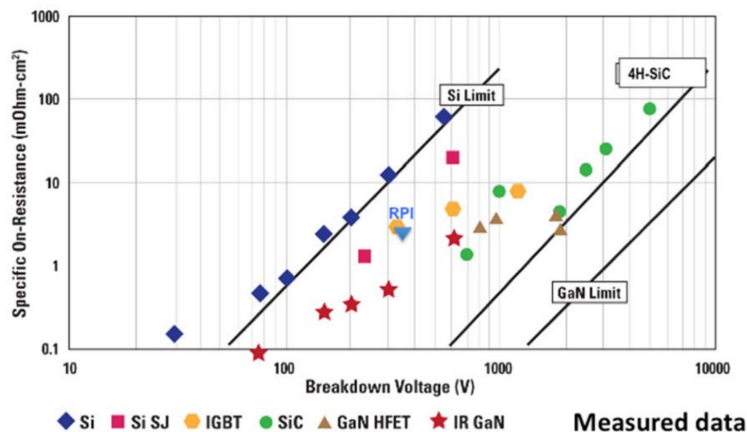


Figure 1.2: The specific on resistance versus the breakdown voltage across different devices and material systems. Despite being a quite immature technology, GaN and SiC clearly outperforms silicon based devices (from [9]).

improved high-frequency operation makes them ideal candidates for power amplifiers in the microwave and millimeter frequency regime. For applications in power switching, higher operating frequencies also allow to scale down passive components like transformers, capacitors or inductors. This helps to decrease costs for power conversion systems when using wide-bandgap semiconductors.

1.2 Motivation and Outline

This section provides the scientific motivation to write this thesis, followed by a brief description of the overall structure of this work.

1.2.1 Motivation

Despite their superior material properties, wide-bandgap semiconductor devices often suffer from severe reliability issues. Especially charge trapping phenomena prevent researchers and engineers from a further exploitation of the theoretical capabilities of the material systems.

This work primarily focuses on charge trapping in GaN devices, with a special focus on bias temperature instability (BTI) degradation. Unlike silicon technology, which is a very mature technology and thus has a relatively low degradation at nominal operating voltages, comparable bias conditions in GaN devices often already lead to large instabilities of the device parameters. This work aims to give an overview of the current understanding of the physical mechanisms behind those instabilities. Furthermore, some pitfalls and peculiarities of well-established methods for reliability characterization and defect modeling specific to GaN devices are highlighted.

Specifically when using technological computer aided design (TCAD) simulations, one very important phenomenon which is usually not considered for silicon is the charge feedback of the trapped charges. The effects of this mechanism and highlighting its importance for defect modeling of GaN devices is one of the main goals of this work. The second goal is the development of more robust methods for the extraction of different defect parameters from measurements. The focus here is to reliably extract the characteristic time constants of single-defects from random telegraph noise (RTN) signals, which are then used to calculate different parameters like the defect structure, the trap-levels, and the vertical defect positions. Although exclusively tested for

GaN devices, the proposed methods are formulated universally enough to be also transferable to other technologies.

1.2.2 Outline

Chapter 1 gives a brief explanation about the recent interest in wide-bandgap semiconductors and compares their most important material properties to those of silicon using different figures of merit. Furthermore, the scientific motivation of this work and a brief overview on the structure of this work is given.

Chapter 2 discusses the fundamental properties of III-N based devices. Special attention is paid to the modeling of important electronic material parameters for device simulation. The current understanding of the formation of the native two-dimensional electron gas (2DEG) in GaN based devices together with state-of-the-art concepts for normally-on and normally-off devices is also introduced in this chapter.

Chapter 3 delivers an overview of the reliability issues in GaN devices. After a short introduction of common defects in the bulk material and the interfaces, three very important degradation mechanisms in GaN devices, namely current collapse, threshold voltage drift, and hot carrier degradation are presented.

Chapter 4 summarizes the experimental characterization of threshold voltage drift phenomena. Different measurement methods for large-area and nano-scale devices are introduced and discussed with respect to their applicability to the characterization of GaN technology.

Chapter 5 deals with the physical defect models used to simulate charge trapping throughout this work. After the description of capture emission time (CET) maps, which are a clever way to visualize BTI degradation, the non-radiative multi-phonon (NMP) model to calculate phonon assisted charge transitions in insulators and semiconductors will be discussed in more detail.

Chapter 6 is dedicated to the extraction of the characteristic time constants from stochastic charge capture and emission events. After a mathematical description of Markov processes and their application to different types of single defects, different methods to reliably estimate the expectation values of the capture and emission times are developed. First, the most common methods for extraction are compared to a new method based on spectral maps. Afterwards, a Hidden Markov Model (HMM) able to reliably extract the time constants from complex RTN signals is put forward. After a detailed description of all parts of the algorithms, different benchmarks are used to test the robustness of the presented model.

Chapter 7 is split in two main parts. The first part investigates the relevance of charge feedback mechanisms on the threshold voltage (V_{th}) drift of large-area GaN devices. By comparing TCAD simulations to measurement data for different voltages, the impact of these effects on defect modeling, the observed time constants and the simulations itself are discussed in detail. The second part is about single-defect characterization in nano-scale GaN devices. It uses both extraction methods developed in the previous chapter to a) derive different defect structures from the observed RTN signals, b) extract their characteristic time constants, c) calculate different important defect parameters from these results and d) try to deduce the most likely defect candidate.

Chapter 8 gives a brief summary of the achievements of this thesis and some suggestions on future improvements of the presented methods.

Chapter 2

Group III-Nitride Materials and Devices

This chapter investigates the fundamental properties of gallium nitride-based devices. It starts with an overview on the crystal structure and fundamental properties of the available group-III nitrides being gallium nitride, aluminium nitride, and indium nitride. In [Section 2.2](#), essential material parameters for electrical device simulation and the consequences of the electrical polarization in strained and unstrained nitrides are reviewed. The last section of this chapter starts with a discussion about the physics behind the formation of the 2DEG in GaN/AlGaN heterostructures. Afterwards, the fundamental design concepts of different normally-on and normally-off transistors are reviewed.

2.1 Crystal Structure and Growth

Group-III nitrides generally can crystallize in three different structures. The most important one is the hexagonal wurtzite structure being the thermodynamically stable structure at room temperature [7, 10, 11]. The second form is the zinc blende structure which can be stabilized on cubic substrates like Si, SiC or GaAs on the $\{0\ 1\ 1\}$ crystal plane [12–14]. The rock salt structure can only be formed at high pressures and has so far only been investigated in fundamental research in material science and theoretically using ab-initio calculations.

Since the hexagonal structure is the thermodynamically stable form for all III-N semiconductors and most commonly used in fabrication, the other two forms are not going to be discussed throughout the remainder of this thesis. Consequently, all material parameters presented in this chapter and the results and conclusions presented throughout this work refer to wurtzite GaN.

As can be seen in [Figure 2.1](#), the wurtzite crystal structure is formed by two interpenetrating hexagonal closed-packed sublattices for each of the atom types. Due to the hexagonal structure of the unit cells, it is described by two lattice constants, a for the in-plane direction and c for the perpendicular direction (i.e., the distance between two planes). The offset between the cells of metal and nitride ions in c -direction is $5c/8$. Another important property of III-N materials is the lack of an inversion plane perpendicular to the c -axis. This causes two different forms of the crystal, one terminating with the metal ions being in plane as in [Figure 2.1](#) and one with the nitrogen ions on top. These two forms are usually referred to as Ga-face and N-face GaN in literature. The gallium atoms in [Figure 2.1](#) can be replaced by other metals such as aluminium (Al) or indium (In), forming the other two important III-N materials, aluminium nitride (AlN) and indium nitride (InN) [7, 11].

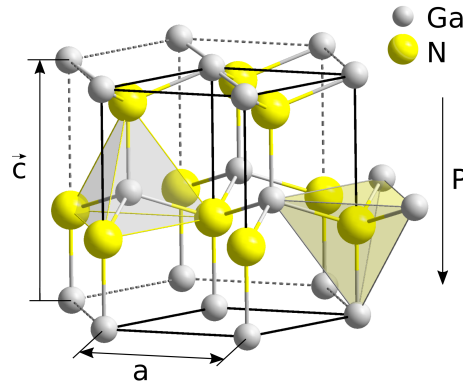


Figure 2.1: The ideal wurtzite crystal structure of group III nitrides consists of two hexagonal closed-packed sublattices, one for each type of atom. The sublattices for metal and nitrogen atoms are interlaced with an offset of $5c/8$ in along c -direction. The separation between the two lattices along the c -axis causes a spontaneous electrical polarization pointing from the metal to the nitrogen atom (from [15]).

Because of the distance between nitrogen (cation) and metal (anion), all nitride compound semiconductors show spontaneous and piezoelectric polarization. The importance of the metal-face and N-face crystals lies in the direction of the polarization which is always pointing from the anion to the cation across the c -direction. A thorough discussion of the polarization effects in nitrides is given in [Section 2.2.4](#).

Like in many other III-V compound semiconductors, one key feature of many GaN based devices relies upon the formation of heterostructures enabled by the ability to engineer the bandgap and – if applicable – the polarization of the materials by using ternary alloys. In the case of group III nitrides, an unprecedentedly large range of energies is available, ranging from about 0.77 eV for InN to 6.28 eV for AlN [2]. The mismatch in the lattice constants between the substrate (usually GaN) and the pseudomorphically (i.e. grown without lattice relaxation) grown alloy on top thereby induces mechanical strain which causes an additional piezoelectric polarization in the strained layer. The relationship between the bandgap energies and the lattice constant c is given in [Figure 2.2](#).

The choice of substrate materials for GaN based devices strongly depends on the device architecture. Up to date, most efforts in GaN epitaxy concentrate on foreign substrates like Si, SiC, and sapphire [16, 17] because of the absence of native substrates. The foreign substrates thereby introduce mechanical strain due to mismatches in the lattice constants and thermal expansion coefficients of the different materials. The best available substrate material is SiC with a lattice mismatch of only 3.1 % and a linear expansion coefficient of $4.4 \times 10^{-6} \text{ K}^{-1}$ as compared to $5.6 \times 10^{-6} \text{ K}^{-1}$ for GaN. The high thermal conductivity of the SiC substrate additionally allows for efficient heat dissipation in power devices. Nonetheless, the very high price compared to silicon wafers limits the usage of SiC as a substrate material.

Despite having the largest lattice constant mismatch (-17%) and the largest mismatch of the thermal expansion coefficient ($2.6 \times 10^{-6} \text{ K}^{-1}$) of all common foreign substrates compared to GaN, the wide range of available sizes, lower costs and its compatibility to existing fabrication processes makes silicon a very attractive base substrate for GaN devices. Special strain management techniques and nucleation layers are needed to prevent crack formation and help to obtain low dislocation densities on the order of 10^8 cm^{-2} [17].

Sapphire as a base material suffers from similar problems as silicon regarding mechanical strain and crack formation [18]. The particularly low thermal conductivity of Al_2O_3 can be mitigated

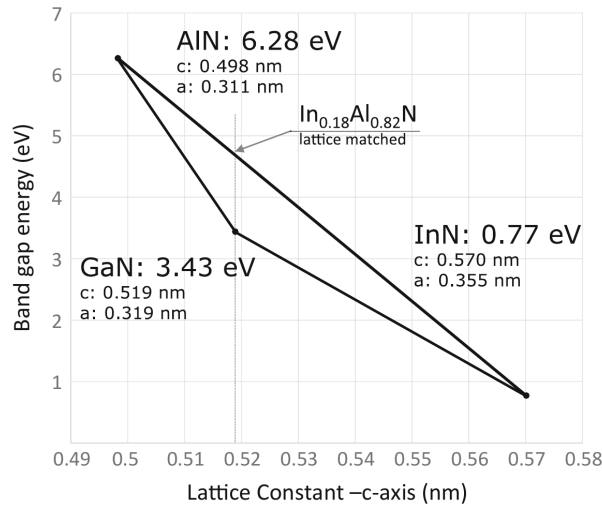


Figure 2.2: The relationship between the bandgap energies and the lattice constant in c-direction of GaN, AlN and InN. The mismatch in the lattice constants between the substrate and a pseudomorphically grown layer on top induces mechanical strain which cause an additional piezoelectric polarization in the strained layer (from [2]).

by flip-chip bonding of the devices onto a material with high thermal conductivity [19]. Sapphire is especially attractive for high-frequency devices because it is naturally insulating and thus has a much higher resistivity compared to silicon wafers.

Native GaN substrates currently can be grown by different techniques up to sizes of 3" [20]. Superior dislocation densities, typically two to four orders of magnitude better than films grown on the other substrate materials, help to lower leakage currents and raise the breakdown voltage of the devices. They also enable the design of vertical power devices which obviously cannot be manufactured on foreign substrates. Despite many improvements during the last years regarding cost-effective large-scale production of high-quality native GaN wafers, many challenges still need to be overcome. Native substrates are thus by far the most expensive choice among all the discussed materials.

2.2 Electronic Material Parameters

This section covers the most important material parameters for device simulation for all three group-III nitrides and their alloys. As many parameters used throughout the literature still show a significant amount of scatter, the set of parameters used by the device simulator Minimos-NT [21] for the simulations in the remainder of this work is listed and justified. A more detailed discussion of the electronic material parameters in GaN, AlN, InN and their alloys can be found in [7, 8, 22].

2.2.1 Band Structure Parameters

One of the most important electrical parameters defining semiconducting materials is their energy gap and the alignment of the conduction and valence band edges of the different materials to each other. The dependence of the bandgap energy on the lattice temperature T is often described

Parameter	GaN	AlN	InN	Ref.
E_g [eV]	3.4	6.2	0.9	[22]
E_{off} [eV]	–	-1.12	0.58	[22]
α [meV/K]	0.909	0.018	0.414	[22]
β [K]	800	1462	454	[22]
M_c [1]	1	1	1	[8]
m_n^* [1/m ₀]	0.2	0.4	0.04	[8, 22]
m_p^* [1/m ₀]	1.0	7.26	1.15	[8]
ϵ_r [1]	8.6	8.5	15.3	[8, 23]

Table 2.1: The most important band structure parameters for the three binary nitrides used in the device simulations.

empirically by [7]

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{\beta + T}. \quad (2.1)$$

The alignment of different materials can be expressed in different ways. One of the most common ones is to use the electron affinity χ , which in solid state physics is defined as the energy difference between the conduction band edge and the vacuum energy. In Minimos-NT [21], the energy alignment is expressed using offset energies relative to the valence band edge. The band edges are then calculated as

$$E_v = E_{\text{off}} \quad (2.2)$$

$$E_c = E_v + E_g \quad (2.3)$$

Other important parameters are the relative permittivity ϵ_r of the materials and the effective carrier masses which are used to calculate the effective density of states by

$$N_c = 2M_c \left(m_n^* \cdot \frac{2\pi k_B}{\hbar^2} \cdot \frac{T}{300 \text{ K}} \right)^{3/2}, \quad (2.4)$$

$$N_v = 2 \left(m_p^* \cdot \frac{2\pi k_B}{\hbar^2} \cdot \frac{T}{300 \text{ K}} \right)^{3/2}, \quad (2.5)$$

with M_c being the number of equivalent conduction band minima. The parameters used for the device simulations are listed in [Table 2.1](#).

2.2.2 Transport Properties

For drift-diffusion simulations, the most important material parameters are the carrier mobilities $\mu_{n,p}$ and carrier saturation velocities $v_{n,p}^{\text{sat}}$. A simple power law is typically used to empirically describe the temperature dependence of the low field electron mobility μ_n^L .

$$\mu_n^L = \mu_{n,300}^L + \left(\frac{T}{300 \text{ K}} \right)^{\gamma_{0,n}} \quad (2.6)$$

Parameter		GaN	AlN	InN	Ref.
$\mu_{n,300}^L$	[cm ² V ⁻¹ s ⁻¹]	1405	683	10400	[22, 25]
$\gamma_{0,n}$	[1]	-2.85	0.8	-3.7	[22, 25]
$\mu_{n,300}$	[cm ² V ⁻¹ s ⁻¹]	80	29	500	[22, 25]
$\gamma_{1,n}$	[1]	-0.2	-3.21	2.39	[22, 25]
$N_{\text{ref},300}$	[cm ⁻³]	7.78×10^{16}	5×10^{17}	3.4×10^{17}	[22, 25]
$\gamma_{2,n}$	[1]	1.3	1.21	-0.33	[22, 25]
$\alpha_{n,300}$	[1]	0.71	-0.18	0.7	[22, 25]
$\gamma_{3,n}$	[1]	0.31	0.31	0.31	[25]
β_n	[1]	1	0.45	1	[22, 25]
$v_{n,300}^{\text{sat}}$	[cm s ⁻¹]	2.5×10^7	3.7×10^7	5×10^7	[8, 22]
$v_{p,300}^{\text{sat}}$	[cm s ⁻¹]	7×10^6	7×10^6	7×10^6	[8]
μ_p	[cm ² V ⁻¹ s ⁻¹]	170	14	220	[2, 8]

Table 2.2: The most important transport parameters for the three binary nitrides used in the device simulations.

Mobility reduction due to ionized impurity scattering for III-V semiconductors is often calculated using the empirical formula of Caughey and Thomas [24] with temperature dependent coefficients.

$$\mu_n^{\text{LI}} = \mu_n^{\text{min}} \frac{\mu_n^L - \mu_n^{\text{min}}}{1 + \left(\frac{N}{N_{\text{ref}}}\right)^{\alpha_n}} \quad (2.7)$$

The temperature dependent parameters for the Caughey-Thomas equation are again calculated with simple power laws.

$$\mu_n^{\text{min}} = \mu_{n,300}^{\text{min}} \left(\frac{T}{300 \text{ K}}\right)^{\gamma_{1,n}} \quad (2.8)$$

$$N_{\text{ref}} = N_{\text{ref},300} \left(\frac{T}{300 \text{ K}}\right)^{\gamma_{2,n}} \quad (2.9)$$

$$\alpha_n = \alpha_{n,300} \left(\frac{T}{300 \text{ K}}\right)^{\gamma_{3,n}} \quad (2.10)$$

The high-field mobilities are calculated in dependence of the driving force F_n and the saturation velocity v_n^{sat} of the electrons using [21]

$$\mu_n^{\text{LIF}} = \frac{\mu_n^{\text{LI}}}{\left(1 + \left(\frac{\mu_n^{\text{LI}} F_n}{v_n^{\text{sat}}}\right)^{\beta_n}\right)^{1/\beta_n}} \quad (2.11)$$

Because of their low mobilities, holes are of minor interest for transport in most GaN power devices. Therefore the hole mobilities are typically modeled by constant values. The transport parameters for the three binary nitrides are listed in Table 2.2. If some material parameters like the saturation velocity for holes and some of the mobility parameters cannot be determined for AlN and InN, the respective values of GaN are used.

Parameter	AlGaN	InGaN	Ref.
E_g [eV]	-1.33	1.4	[8, 22]
μ_n [$\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$]	40	1×10^6	[8]
v_n^{sat} [cm s^{-1}]	-3.85×10^7	0	[8]
m_n^* [1]	4.8×10^{-3}	0	[8]
ε_r [1]	0	0	[25]

Table 2.3: The bowing parameters for mixing of the different electrical parameters of AlGaN and InGaN alloys.

2.2.3 Alloy Parameters

The material parameters for ternary alloys like aluminium gallium nitride (AlGaN) or indium gallium nitride (InGaN) can usually be calculated by linear or quadratic interpolation of the respective binary compounds based on the alloy parameter x . For most of the parameters mentioned in Sections 2.2.1 and 2.2.2, the equation reads:

$$\eta^{\text{ABN}} = \eta^{\text{AN}}(1 - x) + \eta^{\text{BN}}(x) + C_\eta(1 - x)x \quad (2.12)$$

Here, η^{AN} and η^{BN} are the parameters of the binary compounds and C_η is called bowing parameter. There are two exceptions to this rule, namely for the band energy offsets E_{off} and the electron mobilities. If the ratios between conduction band offset and valence band offset is assumed to be constant over the whole composition range, for the energy offset the mixing equation reads [26]

$$E_{\text{off}}^{\text{ABN}} = \frac{E_{\text{off}}^{\text{AN}}(E_g^{\text{ABN}} - E_g^{\text{BN}}) - E_{\text{off}}^{\text{BN}}(E_g^{\text{ABN}} - E_g^{\text{AN}})}{E_g^{\text{AN}} - E_g^{\text{BN}}}, \quad (2.13)$$

while the effective low-field mobility is calculated by

$$\frac{1}{\mu^{\text{ABN}}} = \frac{1 - x}{\mu^{\text{AN}}} + \frac{x}{\mu^{\text{BN}}} + \frac{(1 - x)x}{C_\mu}. \quad (2.14)$$

Table 2.3 lists the bowing parameters for the different quantities used in the simulations for the two alloys AlGaN and InGaN.

2.2.4 Polarization Charges

As mentioned previously, all group-III nitrides show spontaneous and piezoelectric polarization because of the distance between the planes of the nitrogen and the metal ions in the crystal structure. The sign and magnitude of spontaneous polarization depend on the difference in electronegativity between the metal and the nitride. For ternary alloys like AlGaN, the macroscopically observed polarization is a function of the alloy parameter x between the two binary semiconductors. As can be seen in Figure 2.2, also the lattice constants a and c depend on the used alloy composition.

If for example a layer of AlGaN is grown pseudomorphically on top of an unstrained GaN buffer, the mismatch of the lattice constants a cause tensile strain in the AlGaN layer, which induces an additional amount of piezoelectric polarization on top of the spontaneous polarization. A schematic picture of the different terms of polarizations present in a GaN/AlGaN heterostructure can be seen in Figure 2.3.

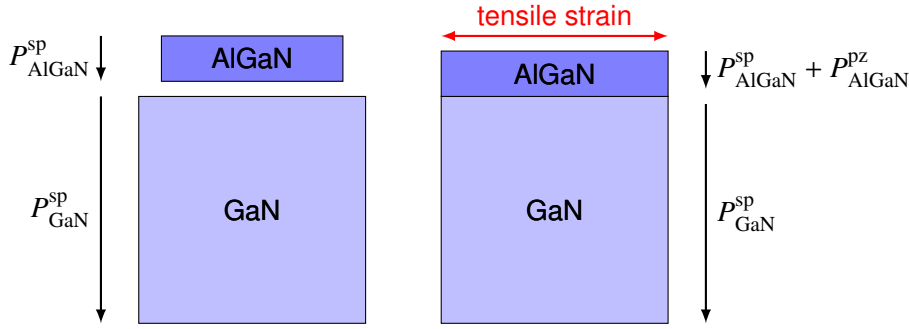


Figure 2.3: Schematic of spontaneous and piezoelectric polarization induced by a pseudomorphically grown AlGaN layer on top of a relaxed GaN substrate (from [6]).

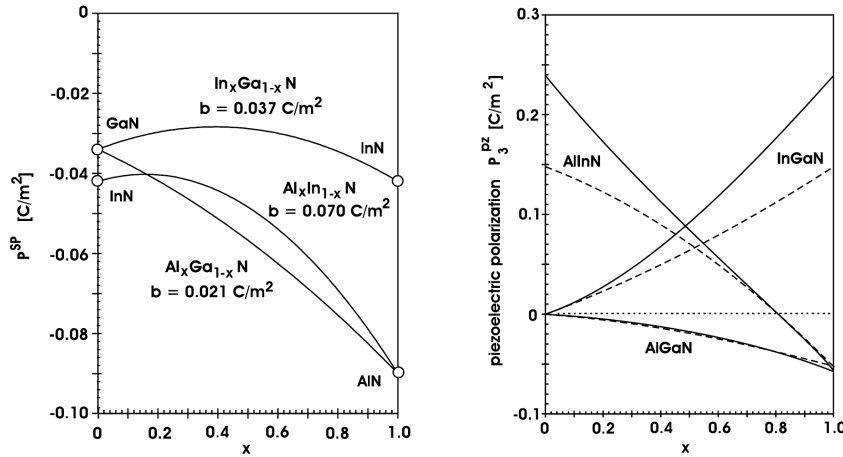


Figure 2.4: Spontaneous and piezoelectric polarizations in binary and ternary nitrides. The values of the piezoelectric polarization are calculated with a set of density functional theory (DFT) parameters from [28] for a pseudomorphically grown alloy on top of a relaxed GaN buffer. The dashed lines are calculated using a linear interpolation between the strain parameters of the binary compounds. For highly strained layers containing large amounts of indium, the nonlinearities in terms of strain should be taken into account (solid lines) (from [27]).

In [Figure 2.4](#), the amount of spontaneous and piezoelectric polarization of different nitrides in dependence of their alloy composition can be seen. For the calculation of the piezoelectric polarization, the alloy is assumed to be grown on top of a relaxed GaN buffer. The dashed lines for the piezoelectric polarizations are calculated with a linear interpolation between the strain parameters of the binary compounds. Especially for highly strained layers containing large amounts of indium, the nonlinearities due to the strain should be taken into account (solid lines) [27].

The values in [Figure 2.4](#) for the spontaneous and piezoelectric polarizations for the most important binary and ternary alloys can be calculated in dependence of x [27]:

$$P_{\text{AlGaN}}^{\text{SP}}(x) = -0.090x - 0.034(1-x) + 0.021x(1-x) \quad [\text{C/m}^2] \quad (2.15)$$

$$P_{\text{InGaN}}^{\text{SP}}(x) = -0.042x - 0.034(1-x) + 0.037x(1-x) \quad [\text{C/m}^2] \quad (2.16)$$

$$P_{\text{AlInN}}^{\text{SP}}(x) = -0.090x - 0.042(1-x) + 0.070x(1-x) \quad [\text{C/m}^2] \quad (2.17)$$

The piezoelectric polarization obviously depends on the used buffer material. For weakly strained ternary alloys grown on an unstrained GaN buffer, the following equations hold [27]:

$$P_{\text{AlGaN/GaN}}^{\text{pz}}(x) = -0.0525x + 0.0282x(1 - x) \quad [\text{C/m}^2] \quad (2.18)$$

$$P_{\text{InGaN/GaN}}^{\text{pz}}(x) = +0.1480x - 0.0424x(1 - x) \quad [\text{C/m}^2] \quad (2.19)$$

$$P_{\text{AlInN/GaN}}^{\text{pz}}(x) = -0.0525x + 0.1480(1 - x) + 0.0938x(1 - x) \quad [\text{C/m}^2] \quad (2.20)$$

For highly strained InGaN and aluminium indium nitride (AlInN) layers, the linear interpolation in (2.18)-(2.20) severely underestimates the amount of piezoelectric polarization (solid lines in Figure 2.4). A refined approach for the calculation in those cases can be found in [27]. The polarization-induced bound charge density can be calculated by taking the gradient of the total polarization in space.

$$\rho_{\text{pol}} = -\nabla \vec{P} = -\nabla(\vec{P}^{\text{sp}} + \vec{P}^{\text{pz}}) \quad (2.21)$$

As mentioned before, nitride compound semiconductors only have a polarization across the *c*-axis with its sign depending on the surface termination of the crystal surface (gallium (Ga)-face or nitrogen (N)-face). For the material interfaces in heterostructures as shown in Figure 2.3, the value of the polarization also jumps at the GaN/AlGaN interface because of the different values of spontaneous and piezoelectric polarization. The abrupt change of polarization across the *c*-direction thus causes fixed two-dimensional charge densities at these interfaces. For the GaN/AlGaN heterostructure the values read as:

$$\sigma_{\text{AlGaN}} = \vec{P}_{\text{AlGaN}}^{\text{sp}} + \vec{P}_{\text{AlGaN}}^{\text{pz}} \quad (2.22)$$

$$\sigma_{\text{GaN/AlGaN}} = \vec{P}_{\text{GaN}}^{\text{sp}} - (\vec{P}_{\text{AlGaN}}^{\text{sp}} + \vec{P}_{\text{AlGaN}}^{\text{pz}}) \quad (2.23)$$

$$\sigma_{\text{GaN}} = -\vec{P}_{\text{GaN}}^{\text{sp}} \quad (2.24)$$

2.3 Fundamentals of GaN/AlGaN Power Devices

Historically, the interest in GaN based semiconductor devices was sparked by the development of blue light emitting diodes (LEDs) by Akasaki, Amamano, and Nakamura [29]. Soon after the discovery of the formation of a native 2DEG in unintentionally doped (UID) GaN/AlGaN heterostructures [30], already the first heterojunction field-effect transistor (HFET) was built [31]. Because of the large sheet carrier density of about 10^{13} cm^{-2} and the high electron mobilities in the channel, this kind of devices are usually also called high-electron-mobility transistor (HEMT).

Up to date, the majority of the HEMTs are built on Ga-face GaN and show normally-on behavior, meaning that a negative voltage has to be applied in order to turn them off. However, due to safety reasons in case of driver failures, normally-off devices are the preferred choice for power applications. Since the 2DEG is formed natively, the design of normally-off devices with comparably high performance is still an essential task for research groups around the world.

This section briefly addresses the physics behind the formation of the 2DEG due to the polarization charges as well as the compensation mechanism of the electrons in the channel. Afterwards, the design principles of current normally-on and normally-off HEMTs are discussed in Sections 2.3.2 and 2.3.3.

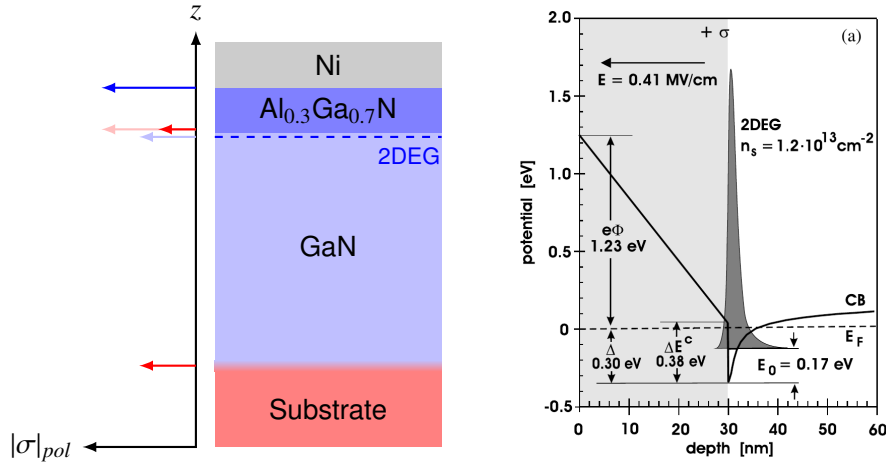


Figure 2.5: **Left:** The schematic of a Ga-face GaN/Al_{0.3}Ga_{0.7} heterostructure with the sign and position of polarization induced sheet carrier densities. The red arrows stand for positive and the blue ones for negative sheet charges. **Right:** Self-consistent solution of the structure on the left with the polarization charges and the Schottky barrier of nickel for layer thicknesses of 2000 nm and 30 nm, respectively (from [27]).

2.3.1 Formation of the 2DEG

The formation of an intrinsic 2DEG at a high-quality GaN/AlGaN interface with improved carrier mobilities was first discovered by Khan in 1991 [30]. During the following years, the source of the electrons forming this electron channel was a matter of intense debate. To understand the confusion about the source of the electrons, it is advisable to come back to the polarization induced charges presented above. In Figure 2.5 a complete heterostructure including the induced sheet carrier densities is shown together with a self-consistent solution of the conduction band edges from [27].

The source of the free electrons cannot be found in polarization charges, as their net charge across the whole device is zero. In addition, the energy per unit area in a polarized material with the thickness t is given by [6]

$$E = \frac{\sigma_{\text{pol}}^2}{2\epsilon_0\epsilon_r} t. \quad (2.25)$$

According to this equation, the energy stored in the crystal structure scales linearly during growth. Due to energy minimization, the polarization charges will be neutralized if the energy exceeds the bandgap of the material. This value is usually called the critical thickness of polarized materials. It can be seen as an upper limit beyond which the polarization charges are going to be neutralized by some kind of defects during growth. For free-standing GaN, this critical thickness is about 10 nm, clearly favoring a local compensation of the polarization induced sheet charges at the surface region between the substrate and GaN. The same argument holds for the AlGaN layer, usually also referred to as the barrier region, as the local energy minimization at the AlGaN surface can be fulfilled either by adsorbates for freestanding surfaces or the next metal or passivation layers for fully processed devices.

Up to date, the most widely accepted model to explain the source of the electron channel is the surface donor model, proposed by Ibbetson in 2000 [32]. It follows a chain of simple charge neutrality arguments which can be outlined as follows:

- In the absence of an external electric field, the sum of all space charges present in any as-grown device needs to be zero. Since the polarization induced sheet charges form a dipole,

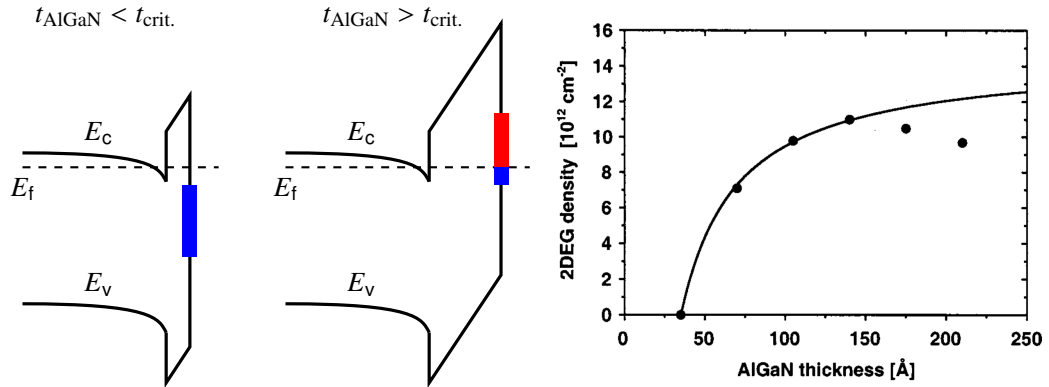


Figure 2.6: The surface donor model. **Left:** If the barrier thickness is below a critical thickness, all surface donors are filled with electrons and no 2DEG is created. **Middle:** The surface donors are only filled partially and donated electrons to the channel. **Right:** Experimentally observed 2DEG density as a function of barrier thickness in a GaN/ $\text{Al}_{0.34}\text{Ga}_{0.66}$ heterostructure. The deviation from the theoretical values can most likely be explained by partial strain relaxation for thicker AlGaN barriers (from [32]).

their sum also has to be zero.

- According to the critical thickness argument discussed before, the positive polarization charges at the substrate are compensated locally. The fact that the electrons are confined in a 2DEG without external fields rules out the same mechanism for the GaN/AlGaN interface.
- If the electrons were primarily generated thermally in the GaN buffer due to unintentional doping or point defects, the 2DEG would also not be confined.
- The confinement of electrons in the 2DEG is, however, experimentally well proven by CV profiles [33].
- Thus, the only two charge contributions left to compensate for the channel electrons are bulk defects in the barrier layer and surface defects. The dominant part of these two contributions can only be the surface states. For typical barrier thicknesses of 10 nm to 20 nm, the effective barrier doping would need to be on the order of 10^{19} cm^{-3} , which would significantly degrade the channel mobility.
- The dominantly donor-like behavior of the defects stems from the fact that their net charge needs to be positive when ionized in order to compensate for the electrons.
- The polarization induced barrier field introduces a dependence of the surface potential on the barrier thickness. Thus the position of a pre-existing defect distribution is a function of the barrier thickness. That means if the barrier is below a certain critical thickness, all donors are below the Fermi level and no 2DEG is going to be formed (see also Figure 2.6).
- Conversely, the defects could also be formed during growth above a certain threshold due to energy minimization following (2.25) and the following discussion. The consequences for the induced electron channel would essentially be the same as for pre-existing defects.

Despite being the widely accepted model for the formation of the 2DEG, the surface donor model is a purely phenomenological model. It provides no information about the structure, distribution and dynamic behavior of the defects. If ionized donors in the barrier layer are neglected, the surface donor model gives a lower bound to the defect density at the interface, which must then correspond to the sheet carrier density in the channel.

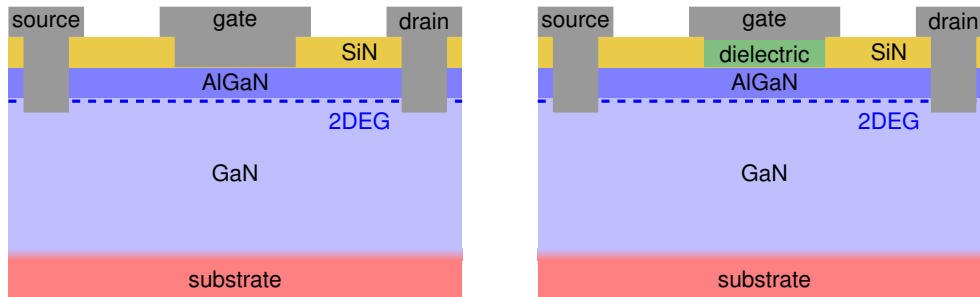


Figure 2.7: Schematic pictures of a GaN/AlGaN HEMT using a Schottky gate (left) and a MIS-HEMT with the gate being insulated from the barrier by an additional dielectric (right). The SiN passivation helps to mitigate current dispersion due to surface defects.

2.3.2 Depletion Mode High-Electron Mobility Transistors

The first GaN/AlGaN HEMT was built using a 100 nm thick barrier layer with 14% Al content on top of a 600 nm thick GaN layer grown on sapphire [31]. The ohmic contacts in HEMTs are manufactured to have direct contact to the 2DEG with no need for additional doping implants. The native 2DEG thus is used as the transistor channel. The gate was initially placed directly on the AlGaN layer forming a Schottky contact. A device schematic of such a device is shown in Figure 2.7 (left).

During the following years, devices with ever-increasing microwave performance were demonstrated [34, 35]. All of these devices were suffering from severe current dispersion effects (i.e. the temporal reduction of the on-current) due to charge trapping in the buffer and at the AlGaN surface. The first strategy to address this problem was to minimize dislocation density in the buffer by optimizing the substrate nucleation layers. Further progress in the suppressing the dispersion was achieved by depositing silicon nitride (SiN) passivation layers on the surface of the devices [36–38] and using field plates to reduce the electric fields at the drain side of the gate during high voltage operation [39, 40].

There are, however, downsides in using Schottky contacts for the transistor gate. The most prominent one is the increased gate leakage currents especially in the off-state at high reverse gate biases [41, 42]. Another problem in switching applications is called back electromotive force (EMF) coming from the load, which potentially can reverse the gate diode for a short timespan and damage the device. An insulating layer placed between the barrier and the gate contact addresses these problems. Such devices are usually called metal-insulator-semiconductor HEMT (MIS-HEMT) and were first demonstrated by Khan in 2003 [43]. A schematic representation of a GaN MIS-HEMT can be seen in Figure 2.7 (right).

The additional AlGaN/dielectric interface underneath the gate nevertheless introduces an additional problem: Due to the Fermi level pinning of the gate in Schottky devices, the effect of surface donors is mostly limited to the access regions of the device causing current dispersion. Because of the insulator underneath the gate, these defects now become relevant in terms of V_{th} shift, especially under forward bias conditions [6, 44].

2.3.3 Normally-Off Device Concepts

Normally-off devices are the preferred choice in power switching applications because of safety reasons in the case of driver failures. Since regular GaN HEMTs naturally show normally-on behavior, special concepts are needed to fabricate enhancement mode transistors. This section

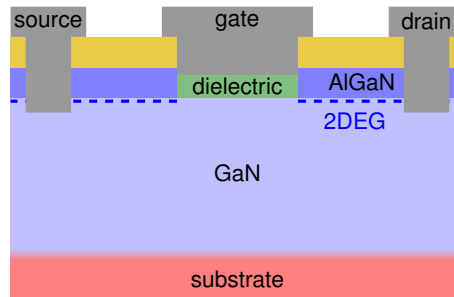


Figure 2.8: Schematic picture of a fully recessed normally-off MIS-HEMT. The removal of the barrier underneath the gate leaves behind a regular MOSFET structure. Because of the removal of the heterointerface, such devices often suffer from poor channel conductivity in the on-state. More recent concepts thus work with partly recessed barriers in conjunction with fluorine implants.

briefly introduces the three most common concepts together with their individual benefits and drawbacks.

Barrier Recess Etch

One of the first approaches to create normally-off devices was to perform a recess etch of the barrier layer before the deposition of the gate [45, 46]. For the first demonstrated devices, the barrier layer underneath the gate was removed completely, leaving behind a regular GaN field-effect transistor (FET) based on the unintentional doping of the buffer layer. If a Schottky gate is used directly on this channel, the achievable threshold voltage is smaller than 1 V with limited gate swing due to turn on of the gate diode.

These problems can be mitigated by the formation of a regular MOSFET structure, see Figure 2.8. The main drawback of fully recessed devices is that the barrier layer needs to be sacrificed to obtain normally-off operation. Such devices thus cannot utilize one of the main benefits of the GaN/AlGaN structures, which is the high-density, high-mobility 2DEG at the heterointerface. Consequently, these devices often suffered from poor performance regarding the channel resistance. The design of devices with shorter channel lengths to lower the channel resistance only solved parts of the problems as trade-offs against the breakdown voltage needed to be made. More recent concepts use hybrid approaches between a partly recessed barrier layer and an additional fluorine implantation step [47, 48].

Fluorine Implantation

Another concept to realize enhancement mode HEMTs is to treat regular depletion mode devices with an additional fluorine implant prior to the formation of the gate, see Figure 2.9 and [49–51], or partly recess etched devices together with implantation to enlarge V_{th} and enhance reliability [47, 48].

To be effective, the implantation has to incorporate the fluorine atoms into the barrier layer. Fluorine mostly stabilizes in interstitial positions in the crystal where it tends to capture a free electron due to its large electronegativity. Thus it can be seen as a fixed negative charge which compensates parts of the positive polarization charges and therefore depletes the 2DEG underneath the gate.

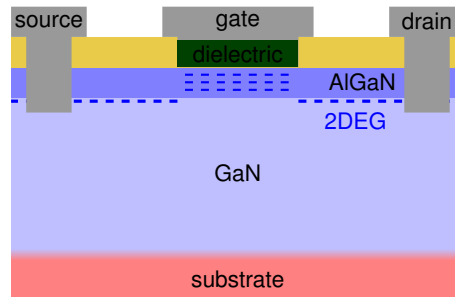


Figure 2.9: Schematic picture of a fluorine implanted enhancement mode MIS-HEMT. The fluorine implants tend to stabilize in interstitial positions in the barrier and can be seen as fixed negative charges in the barrier layer. They compensate part of the polarization charges at the heterointerface and therefore help to deplete the 2DEG. The amount of V_{th} shift can be reliably controlled by the implantation time.

The main benefit of this method is that comparably large values of V_{th} up to 4 V can be achieved. Because it only consists of one additional processing step, normally-off devices can be created using the same device schemes as regular HEMTs or MIS-HEMTs. Unlike other methods, the high-quality 2DEG is mostly preserved although the scattering of channel electrons can potentially be increased due to the fluorine charges. A judgment about the amount of fluorine induced scattering is tricky, since normally-on and normally-off devices from the same technology need to be compared. In [50], a g_m degradation of about 10 % was reported while the devices in [49] showed no degradation at all after annealing.

Gate Injection Transistor

Another concept to obtain enhancement mode devices is called gate injection transistors, which are shown in Figure 2.10. They use a p-doped GaN layer underneath the gate to deplete the 2DEG in this region, allowing to obtain normally-off operation with threshold voltages of about 1 V to 2 V. The devices typically possess two regions of operation, one being similar to a conventional junction field-effect transistor (JFET) for low voltages and an additional increase of conductivity at higher voltages caused by hole injection into the GaN/AlGaN interface from the p-GaN layer [52, 53].

Because of charge neutrality, every hole injected into the channel produces an equal amount of electrons in the 2DEG. Due to the larger electron mobility, the holes tend to remain around the gate while the electrons contribute to the drain current. Since the AlGaN barrier blocks electron injection into the gate, these devices allow low gate leakage currents while maintaining high drain currents.

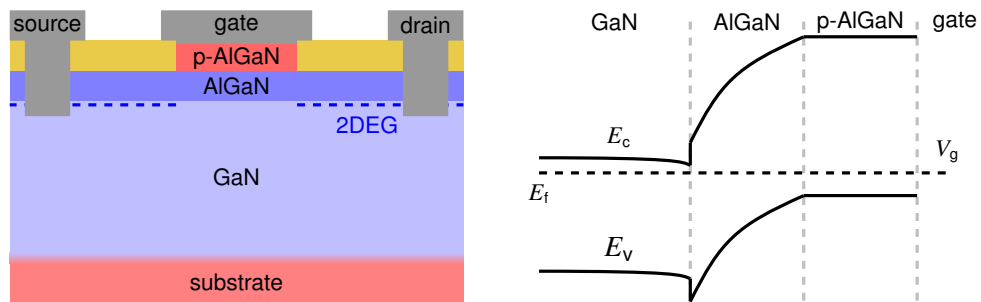


Figure 2.10: **Left:** Schematic picture of a gate injection transistor. An additional p-doped AlGaN layer forms a pn-junction underneath the gate and depletes the 2DEG in order to achieve normally-off operation. **Right:** The corresponding band edges for a vertical cut through the gate. Once the gate voltage exceeds the built-in voltage of the diode, the conductivity peaks again because of holes being injected into the channel.

Chapter 3

Reliability Issues in GaN/AlGaN HEMTs

Despite some GaN HEMTs and MIS-HEMTs being already commercially available, there are still concerns regarding their reliability. Especially charge trapping phenomena prevent researchers and engineers from a further exploitation of the theoretical capabilities of the material system.

This chapter starts with an introduction of the most common defects present in GaN/AlGaN devices. Afterwards, the mechanisms and the state of current research regarding the three major reliability issues, all of them caused by charge trapping, are discussed. These are drain current or R_{on} dispersion which are usually referred to as current collapse, V_{th} drift or negative bias temperature instability (NBTI)/positive bias temperature instability (PBTI) and hot carrier related effects.

3.1 Defects in GaN/AlGaN Devices

Due to a variety of reasons, mainly the lack of native substrates for epitaxy and other growth-related issues like thermally induced mechanical stress, the intrinsic defect density in GaN technology is higher as compared to silicon. Moreover, the used substrate, as well as growth technology and processing, greatly influence the type of the observed defects as well as their densities. Due to that reason, the undoped, as-grown bulk material is usually referred to as unintentionally doped (UID) GaN which commonly appears to be an n-type semiconductor.

Usually, the observed defects are categorized into four different classes:

- **Native Defects** are present in bulk GaN due to crystallographic imperfections. Based on the type of defect, these can be missing Ga or N atoms called vacancies (V_{Ga} and V_N), atoms at the site of their counterpart called antisites (N_{Ga} and Ga_N), or atoms present in the space between the crystal sites called interstitials (N_i and Ga_i).
- **Impurities** can be brought into the structure intentionally as doping (magnesium (Mg), iron (Fe), carbon (C), Si) or unintentionally via residuals in the reactor or during processing via the atmosphere (hydrogen (H), oxygen (O)) or precursor gases (C, H). The species often substitute one type of lattice atom and subsequently creates unwanted energy levels within the bandgap. On a side note, the presence of dopants always distorts the crystal lattice and can thus potentially increase the density of native defects.
- **Surface defects** are of special importance in GaN due to its polar nature. At the interface between crystalline and amorphous materials, the periodicity of the lattice stops abruptly, leading to the formation of a multitude of dangling bonds and other electrically active

defects. In this case, due to the polarization charges, this trend is even more pronounced as the polarization charges forming the 2DEG are widely thought to be compensated by donor states at the surface (see [Section 2.3.1](#)).

- **Dislocations** are extended defects which are primarily formed due to the imperfect surface on the substrate or mechanical stress induced by thermal stress or a mismatch in the lattice constants between two materials (various types of dislocations or stacking faults). Other types of extended defects are formed by molecules of different species (usually oxidized impurities) which tend to form complexes with dislocations, obscuring their theoretically predicted electrical response.

3.1.1 Bulk Defects

There is a large variety of literature dealing with the identification, classification and electrical characterization of point defects in bulk GaN and AlGaN. This section briefly summarizes the most common defect levels of native defects and impurities in bulk GaN. A more detailed introduction into the defect physics of nitride semiconductors is given in Chapter 4 in [7], Chapter 9 in [54], and [55].

First principle simulations using DFT combined with quantum molecular dynamics were used to provide the formation energies and electronic structure of a large variety of defects [56, 57]. There is a considerable spread in the provided transition levels, mostly due to the type of applied bandgap correction and the used DFT functionals. Another problem is that often unstrained crystal structures are used for the calculations, introducing additional errors.

The results of the first principle simulations are then compared to measurements by a multitude of different techniques. The most important one among the different methods is deep level transient spectroscopy (DLTS) [58, 59] as well as its variation using optical excitation of defects [59, 60]. Other commonly used methods are photoluminescence spectroscopy [54, 61], scanning capacitance microscopy and positron annihilation spectroscopy [7].

Despite considerable efforts to attribute defect levels within the GaN bandgap to certain kind of defects, only some of them could be assigned with sufficient confidence. This is partly due to measurement uncertainties or systematic errors introduced by the particular method used. A good summary of the most common defect levels observed in GaN and their ascribed types is given in Figures 3.1 and 3.2. The corresponding Arrhenius plots containing the capture cross sections of the various defects can be found in [55].

In the case of bulk AlGaN, the same structural defects in general appear with another activation energy as compared to GaN [62]. This is simply justified by the fact that even if the same ionization energies to the vacuum level are assumed for a certain defect, the energy differences to the conduction and valence band edges have to change with the alloy composition. On the other hand, there is evidence for other defect levels appearing in AlGaN which are either related to Al or other defects that are probably outside the energy gap in the case of GaN [63, 64]. Finally, some additional defects have been found that only seem to appear on fully processed devices [62, 63].

Of all the defects presented in [Figure 3.1](#), the dominating intrinsic defect is thought to be the nitrogen-vacancy together with its complexes. The *n*-type conduction present in as-grown GaN samples is usually attributed to this defect because of its shallow donor behavior. The most prominently investigated impurities are carbon which is commonly used for co-doping to obtain a semi-insulating buffer as well as magnesium which is mainly used as an acceptor in *p*-type

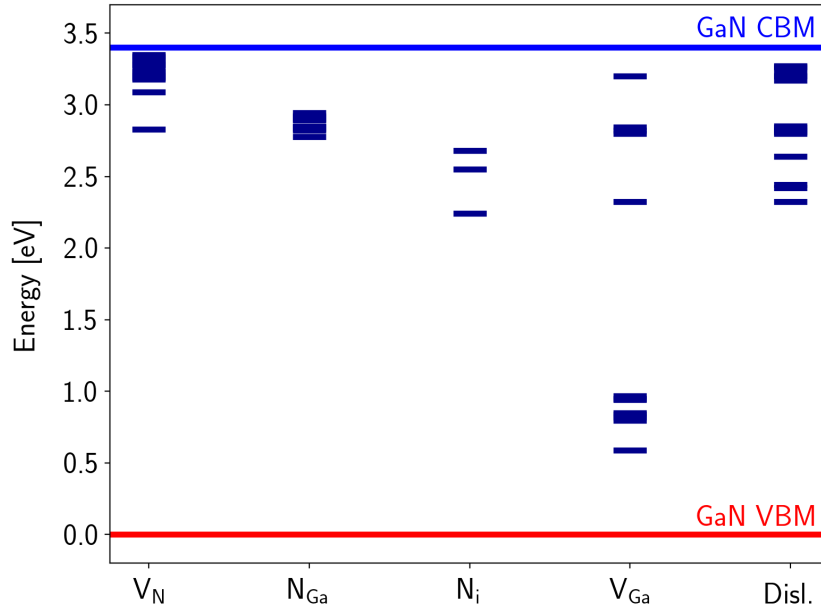


Figure 3.1: Energy levels formed by intrinsic point and extended defects and their complexes. The red and blue lines are the valence and conduction band edges in GaN (data taken from [55]). The listed defect types are the nitrogen vacancy (V_N), the nitrogen antisite (N_{Ga}), the nitrogen interstitial (N_i), the gallium vacancy (V_{Ga}), and various extended dislocation type defects.

GaN. The interest in investigating hydrogen as an impurity is primarily due to its role in forming complexes with other defects creating deep donors as well as deep states around mid-gap [55].

It has to be noted that the vast number of different electrically active states within the bandgap of GaN/AlGaN devices obfuscates the allocation of an electrical response to a certain microscopic defect. The high intrinsic defect densities together with co-doping schemes introduced to obtain semi-insulating buffer materials further complicate correct defect parameter extraction. The polarization field in the AlGaN barrier poses additional problems in the interpretation of measurement data. This is probably the reason why reported parameter values for the same defects show a significant spread throughout the literature.

3.1.2 Interface Defects

Trapping in interface defects are one of the major mechanisms affecting the performance of GaN/AlGaN HEMTs. Unlike in silicon devices, where usually minority carriers controlled by doping form an inversion channel, in HEMTs the defect states at the surface are necessary for the formation of the 2DEG at the channel interface due to the commonly accepted surface donor model as explained in Section 2.3.1. On top of that, the tensile strain formed in the AlGaN layer introduces additional dislocations to the material.

In the case of Schottky HEMTs, electron trapping at the surface is usually reported as current collapse or historically also as “virtual gate” effect. The electrons trapped at the surface of the transistor access regions decrease the surface potential and thus weaken the 2DEG (see also Section 3.2.1). It has later been found that passivating this interface with SiN or other insulating materials helps to decrease the current collapse [36–38].

When creating MIS-HEMTs, the interface between AlGaN and the insulator is not just present at the access regions of the transistor but also underneath the gate. This makes the same surface

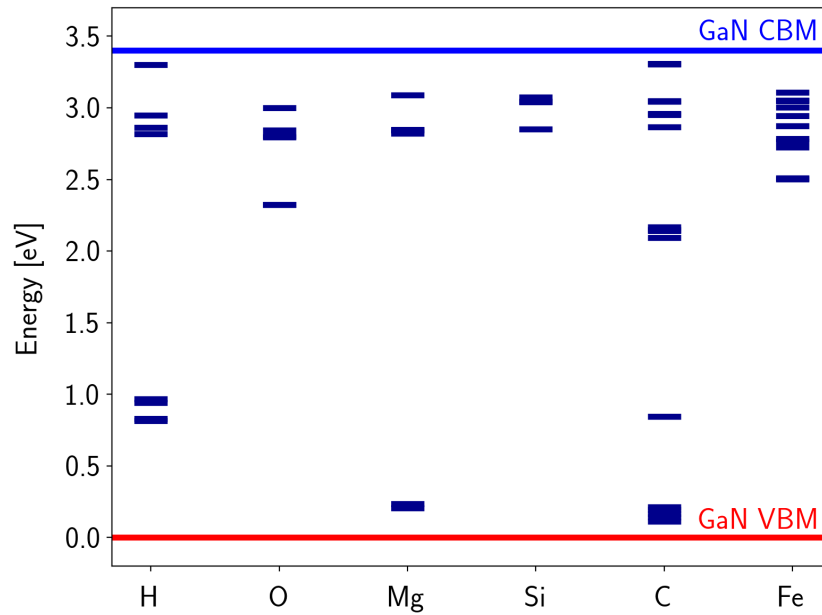


Figure 3.2: Energy levels formed by common impurities and different doping species. The red and blue lines are the valence and conduction band edges in GaN (data taken from [55]).

donor states much more accessible for electrons, especially when applying a positive bias to the gate, see [Section 3.2.2](#).

In general, the structure and the electrical response of the surface donors is still a matter of debate [6, 44, 65]. The impact of different insulator materials on the electrical response of the surface donors is, if at all, only weakly understood up to now either. This is partly because of the lack of a stable native oxide, creating a relatively low-quality oxide semiconductor interface and partly to the sheer number of active defects which has to be in the order of the sheet carrier density of the channel ($\approx 1 \times 10^{13} \text{ cm}^{-2}$) [66].

3.2 Degradation Mechanisms in GaN/AlGaN HEMTs

The long-term stability of GaN/AlGaN HEMTs today is still the main problem prohibits a wider market adoption. Many of the observed degradation phenomena are related to charge trapping and thus especially detrimental to the dynamic response of the devices [67–69]. As in every technology, the different degradation mechanisms overlap during regular operation. To identify the responsible mechanism and extract certain defect parameters, special attention has to be paid to the extraction methodology as well-established characterization schemes for silicon often fail for GaN devices.

3.2.1 Current Collapse

One of the first degradation mechanisms investigated in GaN devices was the transient drain current dispersion after applying a voltage step to the gate or the drain [67, 70]. An example of the observed degradation can be seen in [Figure 3.3](#). To study the nature and the location of the defects responsible for this type of degradation, the drain current is monitored using single or double pulsed drain current measurements to obtain the transient response after a certain stress

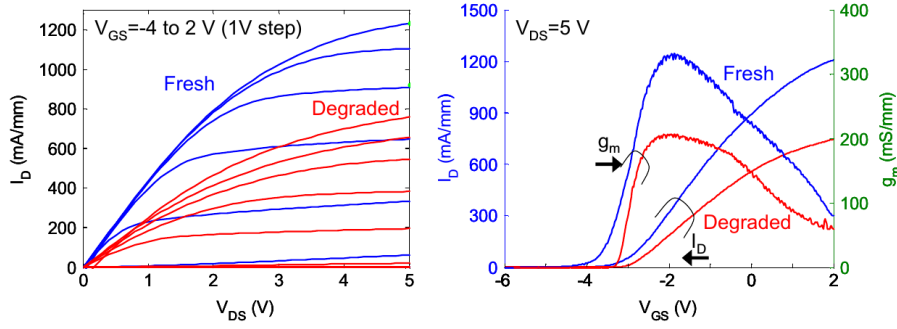


Figure 3.3: Measured degradation on a GaN HEMT after a step-stress experiment with a gate bias of -10 V to -50 V in steps of 1 V for 1 minute each. The degradation can be seen in both, the output (left) and the transfer characteristics (right) of the device (from [67]).

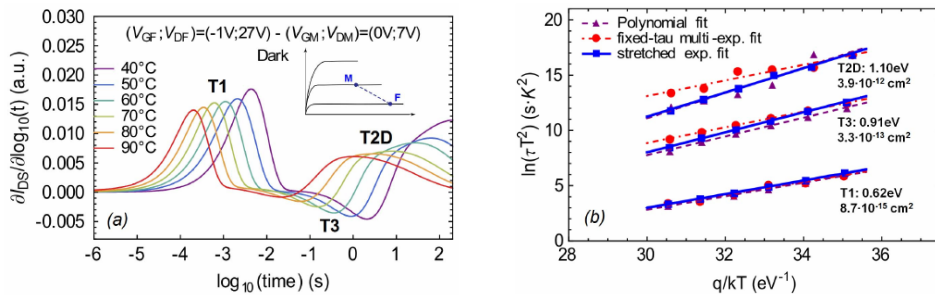


Figure 3.4: Three different defects extracted from measurements of the current collapse. The transient drain current response at different temperatures (left) is used to extract the capture cross-sections and apparent activation energies (right) (from [73]).

condition was applied [71, 72]. This method is often also referred to as drain current transient spectroscopy.

Current collapse is usually investigated at different temperatures with the devices being in the off-state and with high biases applied to the drain. Under these bias conditions, bulk and surface defects residing at the drain sided access regions can be investigated [68, 69, 73]. By changing the bias conditions, defects residing in different areas of the device can be explored. If, for example, all terminals are grounded and a negative substrate bias is applied, defects in the whole buffer region contribute and a mixture between V_{th} drift and current collapse will be observed [68]. The degradation can be evaluated equivalently either in terms of transient drain current, R_{on} or g_m behavior. From this data, defect properties like the apparent activation energy and capture cross-sections can be extracted. A proper selection of temperature, bias conditions and filling pulses thus allows conducting a full defect spectroscopy [73]. An example of the cross-sections and apparent activation energies of three bulk defect levels extracted by this method is given in Figure 3.4.

3.2.2 Threshold Voltage Drift

Threshold voltage drift can in general be caused by all kinds of defects residing underneath the gate. A negative V_{th} shift can be caused for example by positively charged defects in the GaN buffer when applying negative biases at the gate or the substrate. This mechanism is usually attributed to a local modulation of the buffer potential by acceptor defects used for buffer compensation [68, 73]. Other works attribute NBTI in MIS-HEMTs mainly to electron emission from the partially charged surface donors at the gate [74]. Because of the absence of the electron

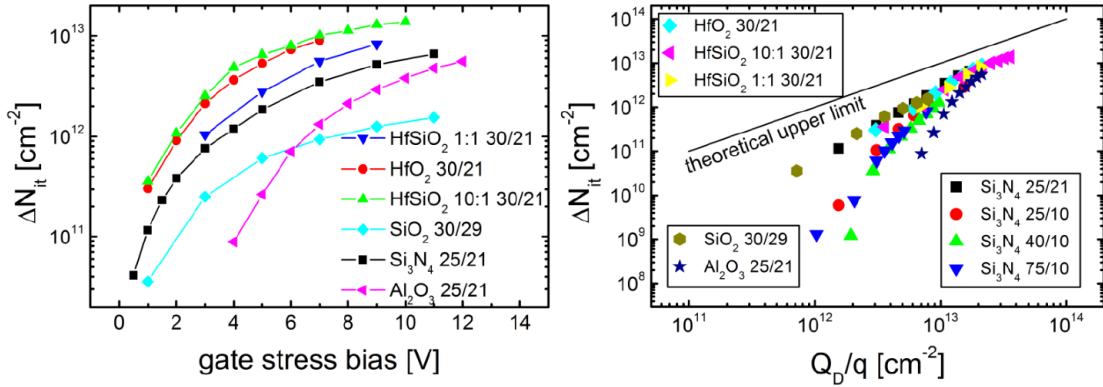


Figure 3.5: Left: The density of charged interface traps during stress increases with the gate stress bias and the specific capacitance of the dielectric. **Right:** The number of trapped charges always approaches the theoretical maximum of $\Delta Q_{d,max} = V_g C_d$. This suggests that the interface trap density is higher than $\Delta Q_{d,max}$ for all of the investigated cases (from [66]).

channel shielding the buffer defects from the defects at the barrier and the interface, the dominant of the two mechanisms can only be determined by the extracted defect properties.

On the other hand, PBTI in normally-on devices is only observed for positive gate bias conditions, ruling out the influence of buffer traps because of the strong shielding due to the electron channel [65, 74, 75]. It was shown that the amount of charge being captured at the dielectric interface under PBTI stress is only limited by the theoretical value calculated from the dielectric capacitance rather than by the interface defect density [66]. When assuming that all charges are stored at the dielectric interface, the maximum amount of charge which can be stored in the dielectric capacitor C_d per unit area is $\Delta Q_{d,max} = V_g C_d$. The amount of trapped charges at that interface ΔN_{it} is

$$\Delta N_{it} = \Delta V_{th} \frac{C_d}{q}. \quad (3.1)$$

Figure 3.5 shows that the amount of V_{th} drift and thus the trapped charges almost reaches the theoretical limit for many devices with different dielectric materials and heights. This suggests that the density of accessible interface states is always higher than the amount of charges stored at the dielectric interface, which is consistent with the predictions coming from the surface donor model discussed in Section 2.3.1.

Studies on the stress and recovery dynamics of interface defects in GaN MIS-HEMTs show that there is almost a logarithmic relationship of the PBTI drift over time for both, stress and recovery [65, 74, 75]. This indicates a very broad distribution of capture and emission time constants of the surface defects as shown in Figure 3.6. There is still a fair amount of uncertainty on whether the observed time constants stem from intrinsic defect properties or if other effects like the transport through the barrier or charge feedback effects obfuscate the extraction of the real capture and emission times [65, 74, AGJ1, AGC1]. A thorough investigation of charge feedback effects on the recovery kinetics of PBTI in GaN will be presented in Section 7.1.3 of this work.

3.2.3 Hot Carrier Degradation

Recent publications have revealed another type of degradation of GaN HEMTs which is most prominent in semi-on conditions (i.e. weak channel together with high drain voltages). These conditions are typically considered the worst-case conditions for hot carrier related phenomena

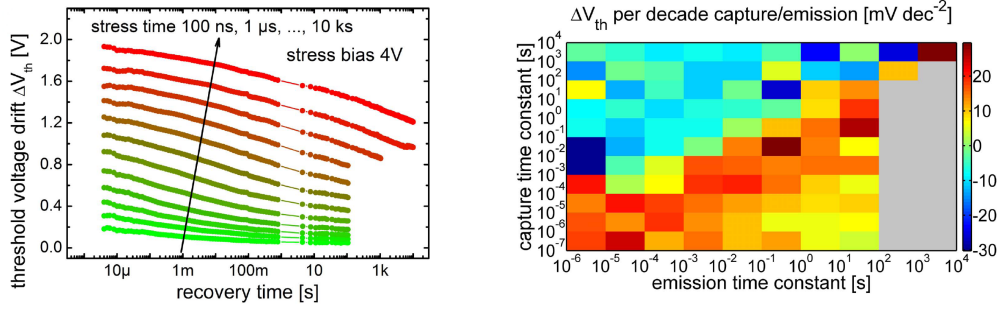


Figure 3.6: Left: Recovery traces after PBTI stress of 4 V and different stress times. The observed logarithmic recovery behavior is typical for GaN MIS-HEMTs. Right: The CET map extracted from the measurements shows a broad distribution of capture and emission times. If this behavior stems from intrinsic defect properties or other effects like charge feedback or the carrier transport through the barrier is still a matter of debate (from [75]).

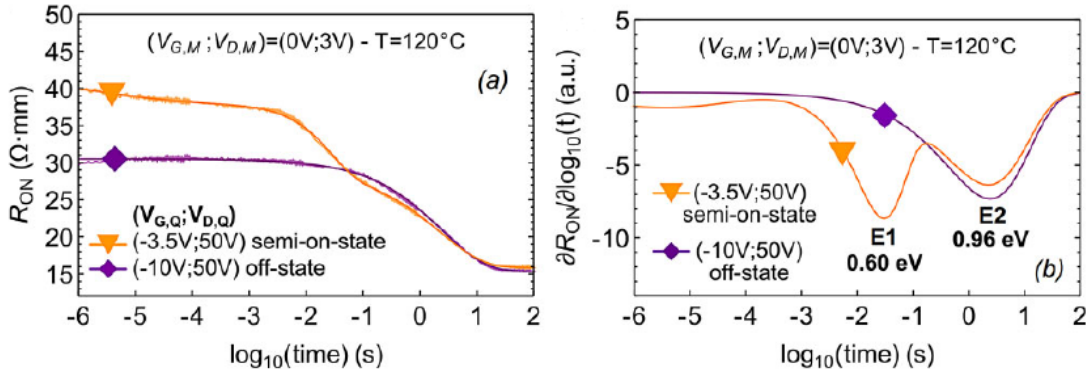


Figure 3.7: Left: The transient R_{on} recovery traces reveal a larger current collapse for semi-on state stress as opposed to off-state stress. Right: Deep level drain current spectroscopy shows an additional defect state which is not present in the off-state degradation which serves as evidence for a hot carrier related mechanism (from [68]).

throughout the literature. Measurements of the R_{on} degradation in off-state conditions compared to semi-on conditions revealed an additional charge trapping mechanism which is usually attributed to bulk defects in the buffer or the barrier (see Figure 3.7). The proposed mechanism is that hot electrons are able to escape from the confinement in the channel and subsequently get captured in pre-existing bulk defects [68, 76, 77]. A recent study delivers solid experimental evidence for channel hot electrons and impact ionization present in N-face GaN HEMTs using electroluminescence in conjunction with the gate currents [78]. Unfortunately this study does not deliver any other information on the observed device degradation except of the elevated gate currents.

As opposed to silicon technology [AGC2, AGJ2], experimental studies on long-term damage caused by channel hot carriers are still missing for GaN technology. One degradation mechanism potentially creating permanent damage at high fields was proposed by del Alamo [67]. It is related to the piezoelectric properties of GaN, where the material experiences a certain amount of mechanical stress when exposed to high electric fields. This could eventually lead to crack formation and the creation of electrically active defects.

Chapter 4

Experimental Considerations

This chapter discusses the most important experimental aspects to obtain the macroscopically observed V_{th} drift of large-area devices under BTI stress as well as measurement techniques for single-defect characterization commonly observed on nano-scale devices.

Experimental characterization of devices requires both, knowledge about the structure and characteristics of the device under test, and the advantages and disadvantages inherent to the chosen measurement technique. [Section 4.1](#) thus contains a discussion of four important measurement techniques for PBTI drift with respect to their applicability to GaN MIS-HEMTs and their potential to extract intrinsic defect parameters from the obtained results.

To be able to extrapolate lifetimes based on voltage or temperature accelerated measurements, or simply to optimize recipes to improve a certain technology, it is mandatory to obtain a deeper physical understanding of the involved mechanisms. Thus another important requirement is the ability to extract intrinsic defect parameters from the obtained data.

In a best case scenario, specific defect candidates can then be identified by comparing the results to first principle simulations [[56](#), [57](#)], microscopic methods like transmission electron microscopy (TEM) [[79](#)], scanning tunneling microscopy (STM) [[80](#)] or atomic force microscopy (AFM) [[81](#)] and physical methods like electrically detected magnetic resonance (EDMR) [[82](#), [83](#)] or electron spin resonance (ESR) [[84](#), [85](#)]. However, many of those methods cannot be applied to fully processed devices or require special treatment of the sample, which makes a straight-forward comparison between the measurements complicated.

Because of the high density of surface defects in GaN, large devices will always contain the simultaneous response of a huge number of defects. For example, in a GaN device with a surface defect density of about 10^{13} cm^{-2} and an active area of $10 \mu\text{m}$, already 10^6 defects will be active. A promising method to identify the microscopic nature of the defects responsible for the V_{th} drift is single defect characterization which has already been demonstrated for silicon technology [[86](#)]. Therefore in [Section 4.2](#), methods to enable single defect measurements in GaN devices are outlined.

4.1 BTI Measurement Techniques

In this section, a brief introduction for the measurement of V_{th} drifts using four different well-established methods is given. The focus thereby lies on the discussion of these methods in the context of PBTI in GaN MIS-HEMTs [[65](#)]. Although the pros and cons of these methods are in principle valid for other technologies too, their impact on the actual results can be significantly different depending on the specific context. A more general discussion on measurement methods for BTI can be found in [[87–89](#)].

4.1.1 Hysteresis Measurements

The most straightforward way to measure the transient V_{th} drift is to record two or more successive $I_d(V_g)$ or CV characteristics. The resulting hysteresis is then directly proportional to the number of trapped charges [74, 90, 91]. This method has several severe drawbacks for the extraction of the electrical response of the defects. The first one is that the hysteresis will be a function of the sweep rate, the minimum and maximum bias and the initial sweep direction (up-sweep first vs. down-sweep first) [92, 93]. In addition, hysteresis measurements mix up the effects of bias acceleration with the influence of stress and recovery times.

Even though $I_d(V_g)$ characteristics can be recorded down to the nanosecond regime with specialized equipment, the minimum sweep times available for standard parameter analyzers are limited to milliseconds. For PBTI in GaN, this poses an additional complication because of the very broad range of capture and emission times and the large density of defects. Defects being much faster than the sweep rate will cause an apparent dispersion of the $I_d(V_g)$ characteristics. This can easily lead to misinterpretations of the obtained results, for example as R_{on} degradation. The separation of all these influences can be challenging, so the hysteresis method must be used with utmost caution for the extraction of intrinsic BTI defect parameters in GaN [94].

In the case of photo-assisted capacitance methods, for a wavelength above the AlGaIn bandgap, the trap occupancy depends on the non-equilibrium concentrations of electrons or holes generated [95]. The defect response in this case is obfuscated because of the absence of holes in regular operating conditions and the unknown distribution of electrons and holes at the interface. For the defect properties obtained by photo-ionization experiments with below bandgap light, additional care has to be taken because of the Franck-Condon shift between the optical and thermal energies [96].

4.1.2 Conductance Frequency Method

The conductance-frequency ($G - \omega$) method has been used extensively to characterize interface defects in Si/SiO₂ structures [97]. There are, however, two major problems when this method should be applied to GaN/AlGaIn stacks. The first one is that like all other capacitance based methods, it relies on an evaluation of a small signal excitation around a quasi-constant operating point. As all PBTI experiments show, this constant operating point cannot be established at forward bias conditions because of the ongoing V_{th} drift. The second main difference is the response of the barrier. The conductivity of the barrier is a function of the applied forward bias, thus the measured changes in the conductivity always represent a superposition of the response of the barrier and the defects [93]. A detailed discussion on the limitations of the $G - \omega$ method in GaN MIS-HEMT structures can be found in [44].

Although the $G - \omega$ method is not very well suited for MIS-HEMTs, it can still be applied to regular MOSFET devices where the electron channel is in contact with the III-N interface [93, 98]. Still, problems in the large-signal stability of the devices during characterization can lead to sweep-rate dependent conductance values [44].

4.1.3 Measure-Stress-Measure Method

A well-established method to obtain the threshold voltage shift following BTI stress is the measure-stress-measure (MSM) method. The idea is to measure the initial threshold voltage $V_{th,0}$ of a device, apply stress to the device at elevated bias conditions and temperatures for some time

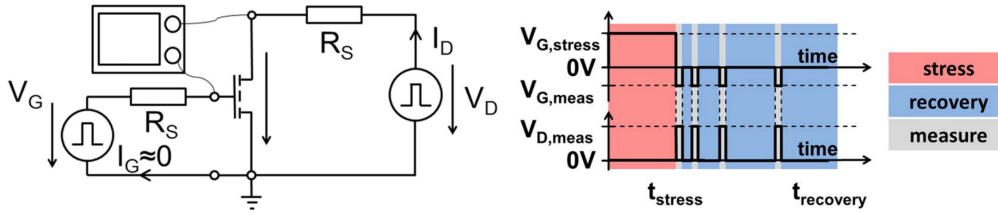


Figure 4.1: Left: The measurement setup used for the eMSM characterization of the devices in Section 7.1.2. The internal resistance $R_S = 50 \Omega$ of the pulse generator was used to record the transient drain currents (from [65]).

Right: The bias conditions during the different phases of the eMSM measurements. After a well-defined stress pulse (red), a short measurement pulse (grey) is used to record degradation. The drain current measurements are repeated in regular intervals during the recovery phase plotted in blue (from [65]).

and after that measure V_{th} again. The threshold voltage shift can then be simply calculated as $\Delta V_{th} = V_{th} - V_{th,0}$.

There are, however, two shortcomings to this approach. The first one, shared with all other methods which require a mapping from the drain current to ΔV_{th} , is the required $V_{th,0}$ of the “fresh” device, which in general depends on the V_{th} extraction method used on the transfer characteristics of the device [88]. Additionally, the initial $I_d(V_g)$ characterization can already impose significant BTI stress if the response time of defects around V_{th} is sufficiently small, as is the case for the devices studied here.

The second problem is the measurement delay t_d between the stress phases. As pointed out by Ershov in 2003, even for the comparably stable silicon technology there is a significant gap between the number of trapped charges between two subsequent stress cycles due to the delay [99]. This clearly demonstrates that the assumption of having a constant trap occupancy (i.e. no recovery) during the measurement of V_{th} is not fulfilled.

Extended Measure-Stress-Measure Method

In the traditional MSM schemes, the stress phases are only interrupted as briefly as possible to record the degraded values of V_{th} . The recorded stress data can therefore only contain information about charge capture events. Naturally, information about the recovery of defects cannot be obtained with these measurements. Instead of the short interruptions for V_{th} measurements, in extended measure-stress-measure (eMSM) measurements a defined recovery phase spanning up to several decades in time is added between subsequent stress phases [100]. If the stress times between two phases are chosen to increase exponentially, the influence of the recovery phase typically can be neglected if the total degradation is sufficiently small [101].

A typical eMSM setup which has been used to record the V_{th} drift in Section 7.1.2 can be seen in Figure 4.1. There are, however, some fundamental drawbacks of this method. Similar to the traditional MSM method, problems arise with the determination of $V_{th,0}$, the delay times between the different phases, and the amount of degradation imposed by the initial characterization. The other main drawback of this method is the insertion of the short measurement pulses during stress or recovery which potentially leads to a variety of unwanted side-effects like accelerated recovery or additional responses from other defects due to the characterization. This error builds up over time and systematically underestimates the measured degradation.

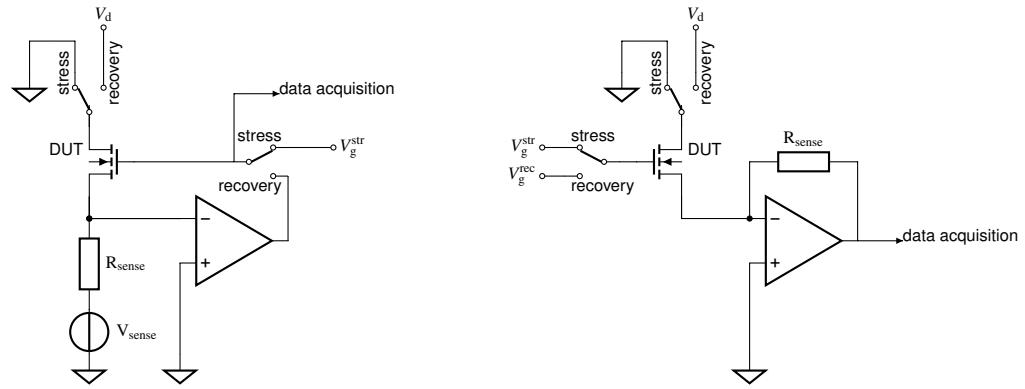


Figure 4.2: Alternative measurement principles to the setup presented in [Figure 4.1](#).

Left: A constant current setup for continuous measurement of the BTI degradation. The analog feedback loop makes the measurement resolution independent from the drain current. The main drawbacks of this principle are that short-cuts or contact interruptions can cause additional stress for the device and the variable oxide field during recovery, which then changes proportionally to ΔV_{th} (from [101]).

Right: The operating principle of the TDDDS measurement instrument (TMI) uses a current-voltage converter to monitor V_{th} degradation. The switches can be replaced by programmable voltage sources, offering the possibility to also conduct on-the-fly (OTF) measurements. The full-scale current and thus the measurement resolution can be set by using different values for R_{sense} .

The distinct measurement pulses (see grey areas in [Figure 4.1](#) (left)) are usually a consequence of the available measurement resolution of the analog-digital converter (ADC). This is because the available voltage resolution is determined by the current resolution of the setup, the noise level and the transconductance of the device at the readout voltage. The trade-off between these three parameters therefore sometimes requires to choose different readout voltages for the V_{th} measurements and the biases applied for recovery.

The steep subthreshold slope together with the very large PBTI drifts observed in GaN/AlGaN MIS-HEMTs can lead to situations where the amount of degradation exceeds the dynamic range spanned between the readout voltage and complete turn-off of the device. This either requires an adaption of the measurement bias between the stress pulses or another type of setup using constant current regulation as shown in the left picture of [Figure 4.2](#) [101]. The benefit of such a solution is that measurement resolution is only limited by the resolution of the data acquisition device for the gate voltage range of the device. The downside is that a constant drain current instead of a constant gate bias is set during recovery. The consequence is that the applied gate bias changes with recovery which complicates the analysis of voltage acceleration of BTI. Furthermore, since the characterization is only a single-point measurement, dispersion effects in the $I_d(V_g)$ characteristics due to hot-carrier degradation (HCD) or current collapse will falsely be accounted as BTI drift. Another problem if the measurements are conducted in wafer probers is that a loss of contact or short-cuts between source and drain of the device will cause a full-scale gate bias as the desired drain current cannot be regulated any more. This potentially induces additional unwanted BTI stress to the gate or may even damage the device.

The right picture in [Figure 4.2](#) shows the operation principle of the TMI, which is an instrument specifically designed by the Institute for Microelectronics to conduct reliability measurements [102, AGJ3, AGJ4]. The current-voltage converter ensures that the source potential on the device is pinned to zero independently of the drain current. The measurement resolution can be chosen by selecting a full-scale drain current $I_{d,max}$. Together with the full scale voltage V_{max} of the data acquisition unit, the required measurement resistor is calculated as $R_{sense} = V_{max}/I_{d,max}$.

Note that the voltage switches for stress and recovery are actually programmable voltage sources allowing a much faster switching between the bias levels compared to physical switches. The main benefits of the TMI are the constant bias conditions during stress and recovery, very fast and low-noise operation and the possibility to conduct OTF measurements.

4.1.4 On-The-Fly Method

One method which was introduced to overcome the limitations of the MSM techniques due to the measurement delay after stress is the OTF measurement [103, 104]. In these measurements, the drain current or the gate capacitance is monitored continuously across stress and recovery conditions. Therefore this method is particularly well suited for measurements where the device can be stressed and monitored within similar bias ranges. If the device is operated in the linear regime, the drain current I_d can be mapped to V_{th} using the initial threshold voltage $V_{th,0}$ and the corresponding drain current $I_{d,0}$ [105].

$$\Delta V_{th} \approx \frac{I_d - I_{d,0}}{I_{d,0}} (V_g - V_{th,0}) \quad (4.1)$$

A more elaborate technique uses a small-signal modulation of the gate voltage around the current value to monitor the transconductance. These values are subsequently used calculate the BTI drift [103]. Although there is no measurement delay between stress and recovery, the intrinsic delay of the setup limits the minimum recordable stress time. Another intrinsic error stems from the fact that a characterization of the fresh device also needs to be conducted for stress conditions. This induces a certain amount of degradation to the device which eventually leads to an apparent dispersion and an overestimation of the stress during subsequent measurements.

4.2 Single Defect Characterization

In BTI measurements on large-area devices, only the response of a large ensemble of defects can be identified because the electrostatic impact of a single defect cannot be resolved by the measurements. In a simple charge sheet approximation, this can be explained by the charges stored in a capacitor being directly proportional to the capacitance at a fixed gate bias. Thus the relative impact of a single charge grows for small-area devices. The ongoing downscaling of devices allowed to identify charge transition events of single defects as step functions, first observed as RTN in the 1980s [106–108]. The link between RTN and the $1/f$ noise in large-area devices was found in 1984 by Uren *et al.* [107].

4.2.1 Random Telegraph Noise

Since the first observations of single defect transitions in the 1980s, several different types of complex RTN signals have been identified and successfully linked to single defects with multiple states. The term *anomalous RTN* was introduced to describe regular two-level or three-level RTN interrupted by inactive phases [108]. In this work, the authors already put forward configuration coordinate (CC) diagrams of three-state defects including a metastable state which later became the basis for using NMP models with metastable states to describe BTI degradation [109, 110]. A closely related phenomenon is *temporary RTN*, where in contrast to anomalous RTN the signal does not reappear again [110]. An example of two defect configurations producing anomalous and temporary RTN can be seen in Figure 4.3.

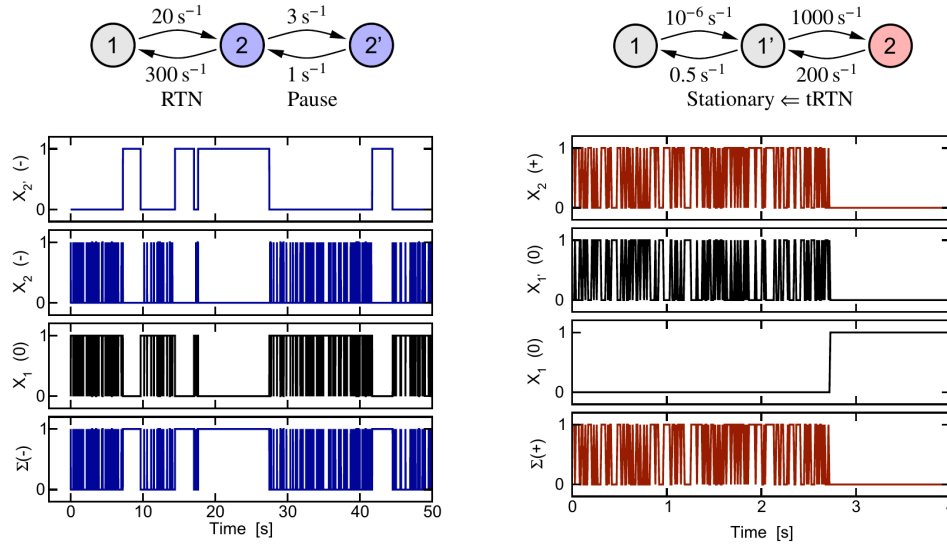


Figure 4.3: Simulated traces of defects producing *anomalous RTN* (left) and *temporary RTN* (right). Both can be represented with a three-state defect with one metastable state (top). The two configurations differ only in their charge states which can either be neutral (grey), negative (blue) or positive (red). The traces show the individual occupancies of each state as well as their resulting RTN signal (bottom row) (from [110]).

More recent studies have also revealed evidence for a link between stress induced leakage current (SILC) and RTN by measuring correlated emissions in the gate and drain currents of a device [111] and linked even more complex capture and emission patterns to four-state defects with two metastable states [112, 113].

The measurement of RTN signals is simply done by applying different biases to the gate in the linear regions of the device and by continuously recording the resulting drain current. If desired, the resulting drain current fluctuations can then be mapped to ΔV_{th} . Additional complications arise if the same measurements are performed on GaN/AlGaIn heterostructures. The first one is a technological problem as the fabrication processes for MIS-HEMTs usually do not allow to produce devices down to the required size. The presence of the barrier layer additionally reduces the expected step-sizes of the individual defects which deteriorates the signal to noise ratio (SNR) of the measurements. Also, the usable voltage range is limited as long-term BTI drift on top of the RTN dramatically complicates the defect parameter extraction.

For GaN technology, it is thus advisable to cool down the devices as far as possible to a) reduce long-term drift by bringing as many defects out of the measurement window as possible and b) increase the SNR by reducing the thermal noise level. This is in contrast to silicon technologies, where temperatures and voltages are usually ramped up to trigger more defect responses. For normally-on devices, the voltage range is also limited to negative voltages because of their large instability in the forward-bias range.

4.2.2 Time Dependent Defect Spectroscopy

The time-dependent defect spectroscopy (TDDS) was derived from the deep level transient spectroscopy, which has been extended for single defects in 1988 [114]. Single defect DLTS relies on two major assumptions, namely an exponential distribution of the emission times and different step heights for each defect. In their original study, Karawath and Schulz used a statistical analysis of the emissions observed after several stress and recovery cycles at different temperatures to

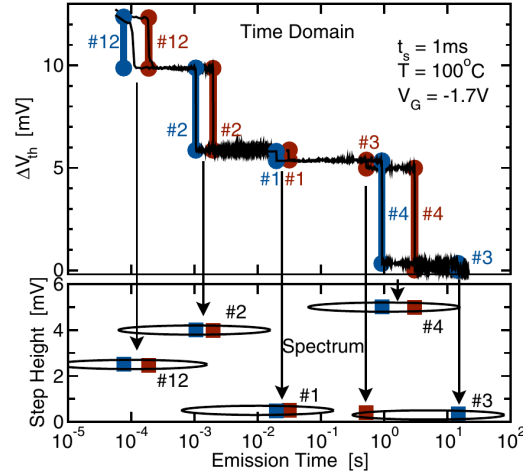


Figure 4.4: Two recovery traces recorded from the same device repeatedly stressed under the same conditions (top). The spectral maps to extract the capture and emission times are constructed by plotting the step-heights versus the observed emission times (bottom) (from [116]).

extract the emission time constants of the defects and calculate their activation energies [114]. It was originally also assumed that the defects are always charged after applying a certain stress to the device, thus the influence of stress time on defects with larger time constants was neglected.

The influence on the stress time on the average defect occupancies was finally accounted for with the introduction of the TDDS [115, 116]. It takes advantage of the fact that the average occupancy f of a defect at a certain stress bias and temperature is a function of the applied stress time t_s [110]:

$$f(t_s) = f_H + (f_L - f_H)e^{-t_s/\tau_c} \quad (4.2)$$

The symbols f_H and f_L are the equilibrium probabilities of the defect being charged at the stress and recovery bias respectively. In TDDS measurements this relation can be used to calculate the capture time of the defect by performing a series of eMSM measurements with different stress times and count the number of emission events of a certain defect. Given a series of N measurements with N_e emission events of a defect at a certain stress time, an estimator of the average occupancy is given by

$$f(t_s) \approx \frac{N_e(t_s)}{N}. \quad (4.3)$$

The confidence limits of the average occupancy above highly depend on the values of N and N_e [117]. The emission time constant can easily be calculated by taking the mean value of all emission times of a specific defect after the recovery voltage has been applied. The defects are usually visualized by two-dimensional histograms called spectral maps, which contain the step-heights and the emission times extracted from the series of recovery traces for a certain stress time, temperature, and bias condition (see Figure 4.4). The observed changes in the spectral maps then can be leveraged to extract a variety of parameters by assuming a certain defect model (usually NMP transitions, see Section 5.2).

When compared to RTN measurements, TDDS in general can be used to characterize a larger number of defects simultaneously. In addition, they can be used to extract the time constants for a much broader voltage range because the capture time constants are derived indirectly from the emission events. This is particularly useful because it allows the sampling rate to be lower than in RTN measurements. A further optimization of the measurement window can be done by using logarithmic sampling which allows obtaining a larger maximum recovery time for a given memory depth of the data acquisition equipment.

Chapter 5

Defect Modeling

In this chapter, the essential models which will be used to describe BTI phenomena throughout this work are presented. After a brief explanation of capture emission time (CET) maps being a clever way to visualize experimental eMSM data in [Section 5.1](#), the theoretical foundation of the NMP theory to describe charge trapping is discussed in [Section 5.2.1](#). Some evidence for the importance of phonon-assisted charge transitions in oxide defects and the framework of the well-known NMP four-state model and the derivation of equations for the time-constants is covered in [Section 5.2.2](#). [Section 5.2.3](#) introduces an extension to the NMP theory for charge trapping in semiconductors by also including the interactions with the local bands.

5.1 Capture-Emission Time Maps

When performing eMSM measurements, the recorded stress and recovery traces contain detailed information about the active defects at a given temperature and bias condition [[100](#)]. One way to visualize the capture and emission times of the active defects and their impact on V_{th} is using CET maps [[115](#), [118](#), [119](#)]. In the case of large-area devices, the response of the individual active defects per area $\Delta\tau^2$ is grouped together into a density g_{ij} defined as [[110](#)]

$$g_{ij} = g(\tau_{c,i}, \tau_{e,j}) = \sum_k \frac{\eta_k a_k}{\Delta\tau^2} \text{rect}\left(\frac{\tau_{c,k} - \tau_{c,i}}{\Delta\tau^2}\right) \text{rect}\left(\frac{\tau_{e,k} - \tau_{e,i}}{\Delta\tau^2}\right), \quad (5.1)$$

with η_k being the individual step height, a_k the maximum occupancy change of the defect and the rectangle functions

$$\text{rect}\left(\frac{\tau}{\Delta\tau}\right) = \begin{cases} 1 & |\tau - \Delta\tau| \leq 1/2 \\ 0 & \text{otherwise} \end{cases}. \quad (5.2)$$

With this definition, (5.1) simply collects all defects within a certain area of τ_c and τ_e into g_{ij} (see [Figure 5.1](#)). The change in the occupancy after stressing for t_s seconds and t_r seconds of recovery, normalized by the maximum change in occupancy is given by [[110](#)]

$$h(t_s, t_r; \tau_{c,i}, \tau_{e,j}) = \frac{\Delta f(t_s, t_r; \tau_c, \tau_e)}{a_k} = (1 - e^{-t_s/\tau_{c,i}})e^{-t_r/\tau_{e,j}}. \quad (5.3)$$

The total degradation can be calculated by summing over all capture and emission times.

$$\Delta V_{th}(t_s, t_r) \approx \sum_i \sum_j g(\tau_{c,i}, \tau_{e,j}) h(t_s, t_r; \tau_{c,i}, \tau_{e,j}) \quad (5.4)$$

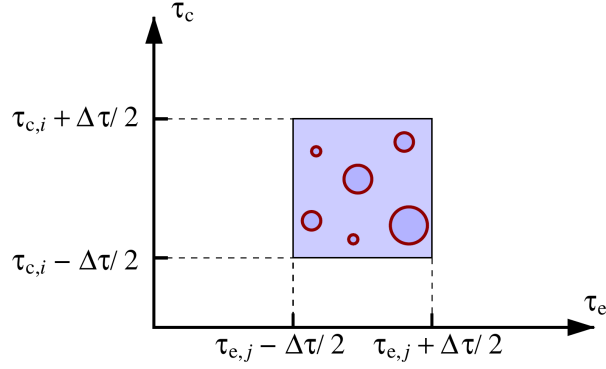


Figure 5.1: Instead of considering a large number of defects, similar defects around around $(\tau_{c,i}, \tau_{e,j})$ are grouped into $g_{i,j}$ defined in (5.1). The size of the circles represents the impact of the defect on ΔV_{th} , which is given by its step height η_k times the maximum occupancy change a_k (from [110]).

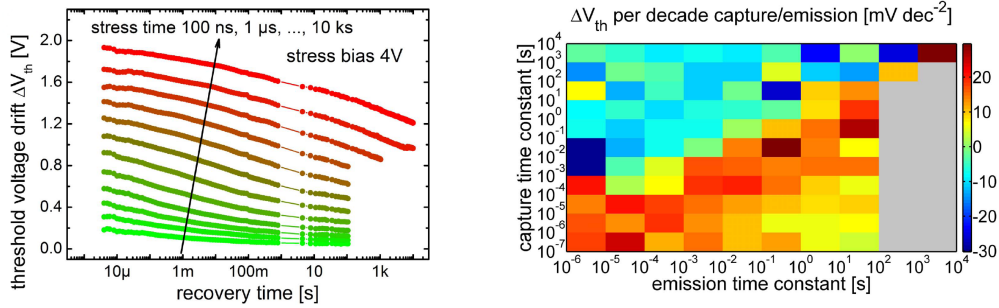


Figure 5.2: The recovery data recorded in eMSM measurements (left) can be directly used to calculate the CET map (right). With equation (5.5), the density plot in the CET map can be used to calculate the V_{th} drift for an arbitrary stress and recovery time (from [75]).

In the limit $\Delta\tau \rightarrow 0$, the rectangle functions in (5.1) can be replaced by Dirac functions. If the exponential terms in (5.3) are replaced by two unit step functions around τ_c and τ_e , one obtains a simple relationship between ΔV_{th} and g [110]:

$$\Delta V_{th}(t_s, t_r) \approx \int_0^{t_s} d\tau_c \int_{t_r}^{\infty} d\tau_e g(\tau_c, \tau_e) \quad (5.5)$$

In other words, the degradation is obtained by summing up all the defects being charged until t_s , but not yet discharged after t_r . This means that the g can easily be obtained directly from the measurements by calculating the mixed partial derivative of the measured stress and recovery traces [110] as

$$g(\tau_c, \tau_e) \approx -\frac{\partial^2 \Delta V_{th}(\tau_c, \tau_e)}{\partial \tau_c \partial \tau_e} \quad (5.6)$$

Thus, the density g in the CET maps stands for the capture and emission times of the defects weighted by their individual impact on V_{th} . It should be noted that the data in a CET map in general is only valid for a particular set of stress and recovery biases at a specific temperature. In large-area devices where a lot of defects contribute to ΔV_{th} , CET maps are usually used to visualize the average response of a large number of defects for a certain technology at a certain stress condition. An example for a CET map obtained from eMSM measurements on GaN MIS-HEMTs is given in Figure 5.2.

5.2 Non-Radiative Multi-Phonon Model

Charges trapped at distinct microscopic defect sites naturally affect their local surroundings. As such, the process of charge exchange with some reservoir always causes the defect site to be deformed due to the change in the local potential surface and a subsequent relaxation towards the new thermal equilibrium [120]. Early applications of the NMP theory were dedicated to the modeling of deep levels in semiconductors [121–124]. The modeling of transient V_{th} drift with NMP transitions to oxide defects was first done by Tewksbury [125, 126].

This section briefly only summarizes the most important concepts of the NMP theory and its applications for describing the transient V_{th} drift of devices. A more elaborate discussion on the NMP theory including the discussion of different BTI models can be found in [88]. Details on the derivation of the NMP model in the context of charge transitions to oxide defects are given in [86].

5.2.1 Theoretical Background

Charge transitions between two states (i.e., the defect state and some reservoir) can in general either happen directly (radiative) or phonon-assisted (non-radiative). The physics of these transitions is governed by electron-phonon coupling which describes the coupling between the wave functions for the system of electrons and phonons. The solution of the two coupled Schrödinger equations is, however, unfeasible for all practical purposes. This problem can be circumvented by taking advantage of the fact that the motions of electrons is usually much faster than the motion of phonons. The so-called Born-Oppenheimer approximation [127] allows solving the two systems separately from each other.

The electron wave functions can thus be obtained for the Coulomb energies of the fixed nuclei positions leading to a $3N$ -dimensional potential energy surface for a system of N atoms. By taking into account that transitions between two states of the potential energy surface are most likely to happen across the minimum-energy path, this multidimensional surface can be reduced to a so-called CC diagram describing the dominant transition path between two states (see Figure 5.3). Even though the shape of the energy surface in the CC diagram can theoretically be derived from first-principle simulations [128] for certain defect candidates, the solution of the resulting non-analytic quantum systems is still cumbersome.

Only the approximation of the actual potential energy surface in the CC diagram by quantum harmonic oscillators allows for an analytic treatment of the problem. The total energy of a system being in state i with the minimum energy $E_{i,\min}$ at the equilibrium position q_i and the curvature c_i reads

$$E_i = E_{i,\min} + c_i(q - q_i)^2. \quad (5.7)$$

The transition rate in the harmonic approximation can be derived from the carrier distribution function f and the density of states D in the reservoir, the electronic matrix element A_{ij} and the lineshape function f_{ij} [125]:

$$k_{ij}(E) = \int_{-\infty}^{\infty} D(E)f(E)A_{ij}(E)f_{ij}(E)dE \quad (5.8)$$

The electronic matrix element A_{ij} thereby accounts for elastic tunneling of electrons from the reservoir to the defect site. The lineshape function accounts for the overlaps between the initial and final vibrational states of the system and can be calculated from the Franck-Condon factors [86, 129, 130].

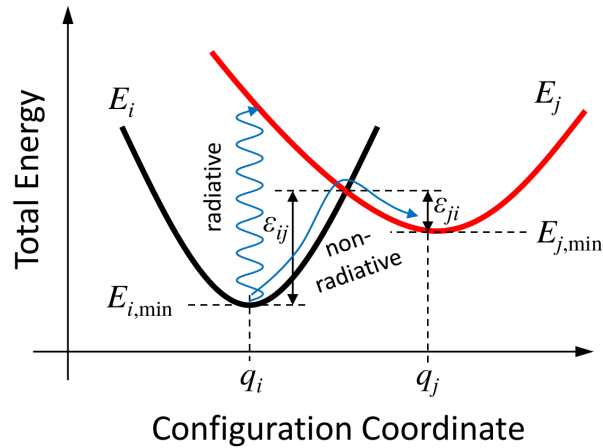


Figure 5.3: A schematic CC diagram of a system with a neutral charge state (black) and a positive charge state (red). Transitions can either be radiative (due to photon excitation) or non-radiative due to phonon interaction (from [88]).

5.2.2 Oxide Defects

There is convincing evidence that phonon-assisted transitions play an important role in charge trapping events in oxide defects. Since thermal emission into the conduction or valence bands at the defect site is typically unfeasible, tunneling of carriers from the channel (or the gate) to the defect site is indispensable. If elastic tunneling was the dominant process, the measured time constants should possess certain properties, all of which are contradicted by existing measurement data [88]:

- The measured temperature activation of the time constants is measured to be Arrhenius-like whereas for pure elastic tunneling the temperature dependence is distinctively weak [131].
- The time constants of devices with thin oxides would have to be tunnel-limited which is in contradiction to measurements done for thin oxides [132].
- Defects closer to the channel would have smaller time constants and thus would be charged first. This mechanism implies a correlation between the trap positions extracted from single-defect measurements, usually obtained from the intersection between τ_c and τ_e , and the capture times which could not be found experimentally [133].
- The time constants should have a nearly linear bias dependence; single defect studies however revealed a more complex behavior [116].

The examples above thus show that phonon interactions have to be an integral part of a physics-based model for BTI.

The NMP Four-State Model

The NMP four-state model emerged from one of the first models which used NMP transitions for the description of charge trapping in oxide defects called the two-stage model [134]. The two-stage model already allowed to model defects producing anomalous RTN by using a three-state NMP model with one meta-stable state. The development of the four-state model was driven by single defect TDDS measurements in ultra-scaled silicon p-metal-oxide-semiconductor (MOS) devices which revealed two distinct types of oxide defects which were named fixed oxide

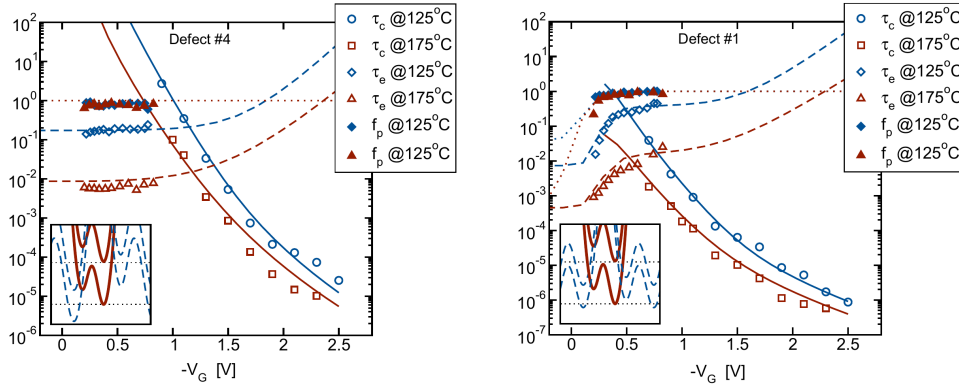


Figure 5.4: Singe defect measurements revealed two different type of defects, the fixed oxide defects (left) and the switching oxide defects (right). The fixed oxide defects have an almost constant bias dependence of the emission times. On the other hand, switching oxide defects show a distinct bias dependency of the emission times (from [110]).

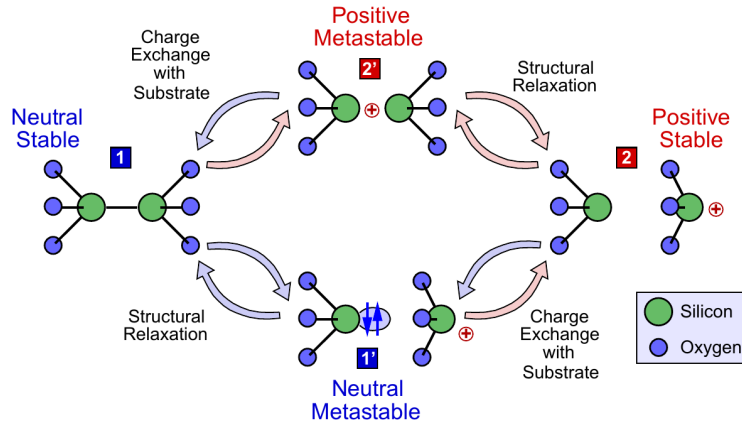


Figure 5.5: The E' center as an example for a four-state defect with its two stable states 1 and 2 and its two metastable states 1' and 2'. The four states offer two distinct pathways for fixed oxide (over 2') and switching oxide (over 1') defects (from [110]).

defects and switching oxide defects. While fixed oxide traps possess a nearly constant bias dependence, switching oxide traps show a distinct bias dependency of the emission times for lower gate voltages (see Figure 5.4). These switching type of defects were also consistent with other single-defect observations like anomalous or temporary RTN.

The NMP four-state model has already proven to successfully describe charge trapping in advanced silicon technology [88, 135, 136, AGC3], but also other advanced technologies like silicon germanium (SiGe) [AGJ3, AGJ4, AGC4], GaN [AGC1], and various two-dimensional materials [137, AGJ5, 138, 139]

By combining first principle simulations with the defect parameters extracted from different measurements, several different defect candidates having the proper structure to fulfill all the requirements of the four-state NMP model in silicon dioxide (SiO_2) could be identified [86, 128]. Each of the four states thereby links to a specific atomic configuration and charge state. As an example, Figure 5.5 shows one of these defect candidates, the E' center. The bias-dependent switching oxide transition path goes from state 2 to 1' to 1, whereas the fixed oxide defect path would be from 2 to 2' to 1.

The schematic CC diagram for a defect similar to the one shown in Figure 5.5 can be found in Figure 5.6. The NMP transition barriers ε_{ij} in the classical limit can be calculated directly from

the intersection point of the two parabolas defined according to (5.7).

$$\varepsilon_{ij} = \frac{c_i(q_j - q_i)^2}{\left(\frac{c_i}{c_j} - 1\right)^2} \left(1 \pm \sqrt{\frac{c_i}{c_j} + \frac{(E_{j,\min} - E_{i,\min})(\frac{c_i}{c_j} - 1)}{c_j(q_j - q_i)^2}} \right)^2. \quad (5.9)$$

Because of the transitions over lower energy barriers being dominant, the positive sign of the square root in (5.9) can be neglected. The equation can be rewritten using common definitions for the energy differences ΔE_{ij} , relaxation energy S_{ij} and the curvature ratios R_{ij} ,

$$\Delta E_{ij} = E_{j,\min} - E_{i,\min}, \quad (5.10)$$

$$S_{ij} = c_i(q_j - q_i)^2, \quad (5.11)$$

$$R_{ij} = \sqrt{\frac{c_i}{c_j}}, \quad (5.12)$$

which for $R_{ij} \neq 1$ leads to

$$\varepsilon_{ij} = \frac{S_{ij}}{(R_{ij}^2 - 1)^2} \left(1 - R_{ij} \sqrt{\frac{S_{ij} + \Delta E_{ji}(R_{ij}^2 - 1)}{S_{ij}}} \right)^2. \quad (5.13)$$

The singularity for $R_{ij} = 1$ can be removed and the transition barrier then reads

$$\varepsilon_{ij} = \frac{(S_{ij} + \Delta E_{ji})^2}{4S_{ij}}. \quad (5.14)$$

Note that the trap levels 1 and 1' are field dependent with the energy shift being proportional to the trap position z in a constant field approximation.

$$\Delta E_{ij}(F) = \Delta E_{ij} - qzF. \quad (5.15)$$

In the classical limit, the energy ε_{ij} has to be overcome for a NMP transition to happen. Boltzmann statistics gives the probability of the vibrational system to be excited by this energy, and thus the transition rate and thus (5.8) can be calculated to

$$k_{ij}(E) = \int_{-\infty}^{\infty} D(E)f(E)A_{ij}(E)e^{-\varepsilon_{ij}(E)/(k_B T)} dE. \quad (5.16)$$

The thermal transition rates to the metastable states without charge transfer can simply be modeled by an Arrhenius law with the activation energy ε_{ij} and the attempt frequency ν as

$$k_{ij} = \nu e^{-\varepsilon_{ij}/(k_B T)}. \quad (5.17)$$

Further simplifications can be done if instead of the NMP transitions happening to a band of states as depicted in (5.8), only the band edges of the conduction and valence band of the reservoir are considered. This makes the transition barrier ε_{ij} and the electronic matrix element A_{ij} independent of energy and thus they can be moved out of the integral. Furthermore, the distribution function in equilibrium conditions can be identified as the Fermi-Dirac distribution and with

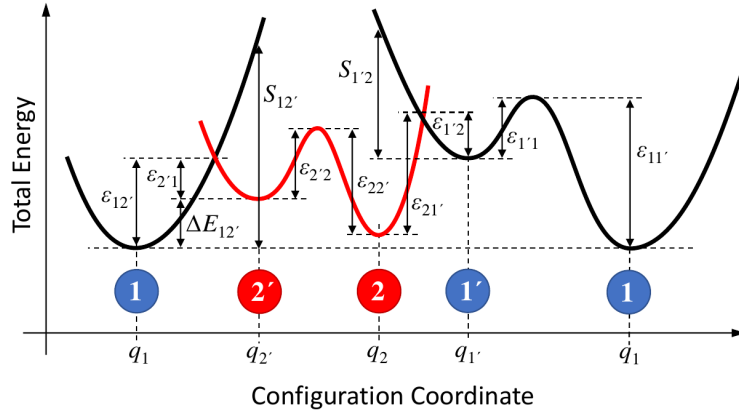


Figure 5.6: The schematic CC diagram of a four-state defect similar to the one shown in [Figure 5.5](#). Note that the trap level E_1 depends on the oxide field and the distance of the defect from the interface z via (5.15) (from [88]).

$\beta = (k_B T)^{-1}$, the reverse rate is linked to the forward rate by:

$$f_n(E) = 1 - f_p(E) = 1 - \frac{1}{1 + e^{\beta(E_f - E)}} = f_p e^{-\beta(E_f - E)} \quad (5.18)$$

$$f_p(E) = 1 - f_n(E) = 1 - \frac{1}{1 + e^{\beta(E - E_f)}} = f_n e^{-\beta(E - E_f)} \quad (5.19)$$

If the integral in (5.16) and the NMP barriers are split up for the conduction and valence band, in the hole picture of [Figure 5.6](#) one obtains

$$k_{ij,\text{CB}}(E) \approx A_{ij}(E_c, z_t) e^{-\beta \epsilon_{ij}} e^{-\beta(E_f - E_c)} \int_{E_c}^{\infty} D_n(E) f_n(E) dE \quad (5.20)$$

$$k_{ji,\text{CB}}(E) \approx A_{ji}(E_c, z_t) e^{-\beta \epsilon_{ji}} \int_{E_c}^{\infty} D_n(E) f_n(E) dE \quad (5.21)$$

$$k_{ij,\text{VB}}(E) \approx A_{ij}(E_v, z_t) e^{-\beta \epsilon_{ij}} \int_{-\infty}^{E_v} D_p(E) f_p(E) dE \quad (5.22)$$

$$k_{ji,\text{VB}}(E) \approx A_{ji}(E_v, z_t) e^{-\beta \epsilon_{ji}} e^{-\beta(E_v - E_f)} \int_{-\infty}^{E_v} D_p(E) f_p(E) dE. \quad (5.23)$$

The remaining integrals can easily be identified as the electron and hole concentrations respectively. As mentioned earlier, the electronic matrix element A_{ij} accounts for elastic tunneling of the carriers. The electronic matrix element A_{ij} can be calculated by multiplying the tunneling probability λ with a capture cross-section σ_0 and a thermal velocity v_{th} of the carriers. Furthermore, the process is assumed to be symmetrical so that $A_{ij} = A_{ji}$.

$$A_{ij}(E, z_t) \approx \sigma_0 v_{th} \lambda(E, z_t) \quad (5.24)$$

The tunneling probabilities could be theoretically derived from the Schrödinger equation. However, the potential well around the defect site and thus the electronic wave function of the defect can hardly be determined. Therefore, the Wentzel–Kramers–Brillouin (WKB) approximation [140] is often used to calculate λ analytically for different shapes of the energy barriers.

The most important energy barriers for BTI are the trapezoidal and the triangular barrier depicted in [Figure 5.7](#). With the carrier energy E , the elementary charge q_0 and the effective tunnel mass

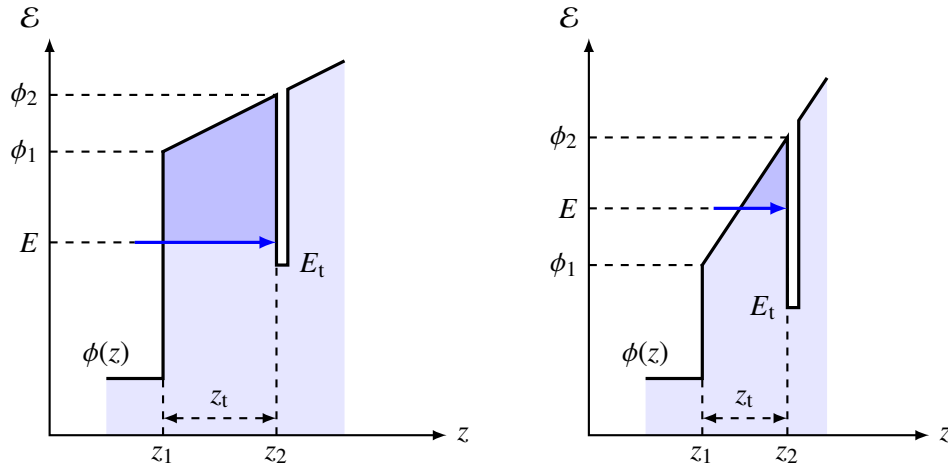


Figure 5.7: Schematic representation of trapezoidal (left) and triangular (right) tunneling barriers used to calculate the WKB factors in (5.25) and (5.26) (adapted from [142]).

m_t , the equation for a trapezoidal barrier reads [141, 142]:

$$\lambda(E, z_t) \approx \exp\left(-\frac{4}{3\hbar q_0} \cdot \frac{z_t}{\phi_2 - \phi_1} \cdot \sqrt{2m_t} \left((\phi_2 - E)^{3/2} - (\phi_1 - E)^{3/2}\right)\right) \quad (5.25)$$

For the special case of a triangular barrier, the equation simplifies to:

$$\lambda(E, z_t) \approx \exp\left(-\frac{4}{3\hbar q_0} \cdot \frac{z_t}{\phi_2 - \phi_1} \cdot \sqrt{2m_t} \cdot (\phi_2 - E)^{3/2}\right) \quad (5.26)$$

When putting all of the above together, the expressions for the NMP rates can finally be approximated with:

$$k_{ij, \text{CB}}(E) = \sigma_0 v_{th} \lambda(E_c, z_t) e^{-\beta \varepsilon_{ij}} e^{-\beta(E_f - E_c)} n \quad (5.27)$$

$$k_{ji, \text{CB}}(E) = \sigma_0 v_{th} \lambda(E_c, z_t) e^{-\beta \varepsilon_{ji}} n \quad (5.28)$$

$$k_{ij, \text{VB}}(E) = \sigma_0 v_{th} \lambda(E_v, z_t) e^{-\beta \varepsilon_{ij}} p \quad (5.29)$$

$$k_{ji, \text{VB}}(E) = \sigma_0 v_{th} \lambda(E_v, z_t) e^{-\beta \varepsilon_{ji}} e^{-\beta(E_v - E_f)} p \quad (5.30)$$

Note that in Figure 5.6, the (gate voltage dependent) trap level corresponds to E_1 and E_2 is the conduction or valence band edge of the reservoir and the calculation of the NMP barriers in (5.13) needs to be adapted accordingly. Additionally, the NMP barriers for the forward and backward transitions are not independent from each other and can be calculated from $\varepsilon_{ij} = \Delta E_{ij} + \varepsilon_{ji}$.

The rates derived in (5.27)-(5.30) can be used to calculate the NMP transition rates for all kinds of defects, as long as their positions and CC diagrams are known. In large-area devices with many different pre-existing defects, the total BTI degradation can be modeled in a Monte Carlo fashion, which means distributing the main defect parameters like ΔE_{ij} , S_{ij} , and R_{ij} around a their mean values using probability distributions.

In real devices different types of defects exist alongside each other, each one having a certain structure and number of states. On top of that, the amorphous nature of oxides will inevitably lead to random fluctuations of the NMP parameters for each of those defect types. The resulting set of CC diagrams and eventual correlations between different NMP parameters thus can hardly be determined at all. Because of that, the Monte Carlo sampling of the four-state model is usually done on a set of independent and normally distributed parameters [135].

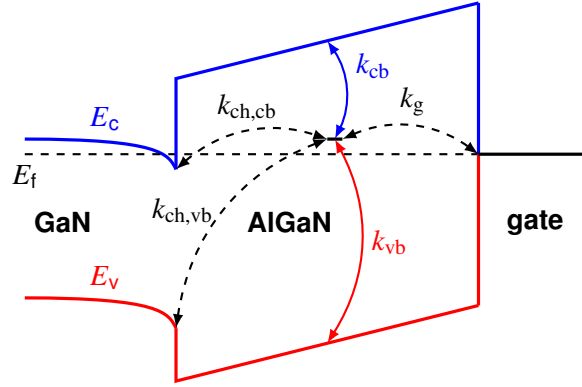


Figure 5.8: All possible charge exchange paths for a defect in the barrier of a GaN/AIGaN HEMT. The dashed rates are calculated in a similar manner as the NMP rates for an oxide defect, see (5.27)-(5.30). For defects in a semiconductor material, additional rates are needed to account for the charge exchange with the local conduction and valence bands. (adapted from [AGT1])

The average number of defects N can easily be calculated from the assumed defect concentration within the oxide. Their random fluctuations are commonly thought to follow a Poisson distribution given by [143, 144]

$$P_N(k) = \frac{N^k}{k!} e^{-N}. \quad (5.31)$$

The electrostatic impact of each defect based on its position within the oxide can be estimated using a simple one-dimensional charge sheet approximation. Alternatively, the defect charges can simply be entered to the right-hand side of the Poisson equation in electrical device simulators.

Unlike the NMP two-stage model, the four-state model does not include any mechanism for the creation and passivation of defects. Thus it is limited to pre-existing defects in the oxide commonly attributed to the recoverable part of BTI. A number of studies however propose a coupled mechanism of creation and annihilation of defect sites driven by the relocation of hydrogen [145, 146]. This triggered another evolution of the NMP four-state model called the gate-sided hydrogen release model [147]. The mechanisms covered by this model are very specifically designed for the Si/SiO₂ interface and therefore beyond the scope of this work.

Within the framework of the NMP theory, defects can interact not only with a single but with an arbitrary number of charge reservoirs (usually the transistor channel and the gate). Thus, the NMP model can also be extended to investigate trap-assisted tunneling (TAT) and SILC in MOSFETs and flash memory devices [148–151, AGT1].

5.2.3 Local Band Interactions

For most BTI defects, the charge exchange is governed by elastic tunneling from the charge reservoir (gate or channel) followed by a phonon-assisted transition, see (5.8). This, however, omits two additional exchange paths, namely the conduction and the valence bands at the site of the defect. Neglecting these rates is a valid approach for insulators as the ionization energies from the trap level to the local band edges are usually very large and the density of electrons and holes are rather low. For shallow defects in semiconductors or high fields, the local band rates can be of relevance, however. As an example, in Figure 5.8 all possible charge exchange paths of a defect in the barrier of a GaN/AIGaN HEMT with a Schottky gate are shown.

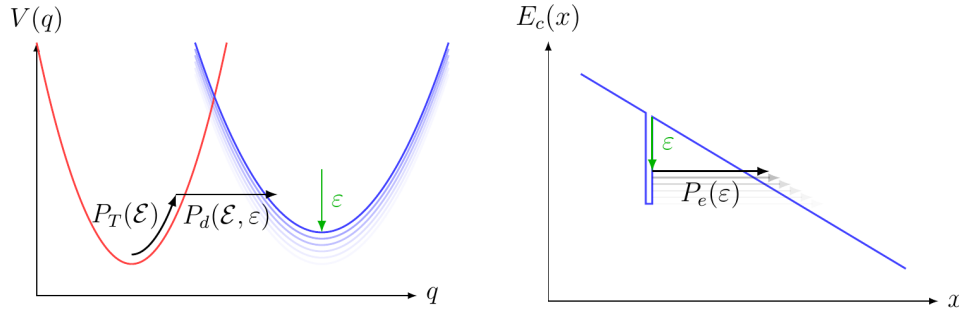


Figure 5.9: The two concurring processes for NMP transitions into the local bands (from [AGT1]).

For emissions to the local bands at the site of the trap, no electron tunneling is required. Therefore the electronic matrix element A_{ij} in (5.24) is evaluated for $\lambda = 0$ and thus a constant. Because of the transition taking place locally, the implicit field dependence of the NMP barriers vanishes so that the transition barriers are constant and independent of the field. At higher fields, non-local transitions with tunneling to neighboring bands can become more favorable compared to pure thermal excitations.

The charge transition mechanism can again be split into three parts, the thermal excitation probability $P_T(\mathcal{E})$, the NMP transition probability $P_d(\mathcal{E}, \varepsilon)$ and the tunneling probability $P_e(\varepsilon)$, see Figure 5.9. In that case, the rates can be split into a field-dependent and a field-independent part:

$$k_{ij} = k'_{ij} + k''_{ij} \quad (5.32)$$

The field-independent part of the rates can be calculated by (5.27)-(5.30) with $z_t = 0$. Note that for this transition, the NMP barrier is also independent of the applied field. When band-bending in the semiconductor can be neglected (i.e., the electric field across the semiconductor can be assumed constant), the field-dependent part is calculated by taking into account elastic tunneling through a triangular barrier. For each infinitesimal barrier lowering, an additional amount of states can be reached. The product of effective density of states, lineshape function and electronic matrix element must then be summed over the barrier lowering ε . With the lowered barrier $E'_c = E_{c,0} + \varepsilon$, the field dependent rates for holes are given by [AGT1]:

$$k''_{ij,\text{CB}} = - \int_{\varepsilon} \frac{dn(\varepsilon)}{d\varepsilon} f_{ij}(E'_c - E_t) A_{ij}(E'_c, F) e^{-\beta(E_f - E'_c)} d\varepsilon \quad (5.33)$$

$$k''_{ji,\text{CB}} = - \int_{\varepsilon} \frac{dn(\varepsilon)}{d\varepsilon} f_{ji}(E_t - E'_c) A_{ji}(E'_c, F) d\varepsilon \quad (5.34)$$

$$k''_{ij,\text{VB}} = - \int_{\varepsilon} \frac{dp(\varepsilon)}{d\varepsilon} f_{ij}(E'_v - E_t) A_{ij}(E'_v, F) d\varepsilon \quad (5.35)$$

$$k''_{ji,\text{VB}} = - \int_{\varepsilon} \frac{dp(\varepsilon)}{d\varepsilon} f_{ji}(E_t - E'_v) A_{ji}(E'_v, F) e^{-\beta(E'_v - E_f)} d\varepsilon \quad (5.36)$$

The differentials for electrons and holes in (5.33)-(5.36) stand for the additional number of states that can be reached due to an infinitesimal barrier-lowering. By assuming that the lowest point of the integration is still well above the Fermi level (i.e., by assuming Boltzmann distributions) and with $\beta = (k_B T)^{-1}$ the additional amount of electrons and holes can be calculated as

$$\begin{aligned} \frac{dn(\varepsilon)}{d\varepsilon} &\approx -\beta n(\varepsilon), \\ \frac{dp(\varepsilon)}{d\varepsilon} &\approx +\beta p(\varepsilon). \end{aligned} \quad (5.37)$$

Inserting the expressions for the number of electrons and holes yields [AGT1]

$$\frac{dn(\varepsilon)}{d\varepsilon} = -\beta \int_{E_{c,0}}^{\infty} D_n(E - E_{c,0}) e^{-\beta(E+\varepsilon-E_f)} dE, \quad (5.38)$$

$$\frac{dp(\varepsilon)}{d\varepsilon} = +\beta \int_{-\infty}^{E_{v,0}} D_p(E_{v,0} - E) e^{-\beta(E_f-E-\varepsilon)} dE. \quad (5.39)$$

With similar simplifications for the lineshape function and the electronic matrix element as explained in Section 5.2.2, the final rates for the interaction with the local bands can be expressed as [AGT1]:

$$k_{ij,\text{CB}}(F) = \sigma_0 v_{th} \left(n(0) e^{-\beta\varepsilon_{ij}} e^{-\beta(E_f-E_c)} + \beta \int_{\varepsilon} n(\varepsilon) \lambda(\varepsilon, F) e^{-\beta\varepsilon'_{ij}} e^{-\beta(E_f-E'_c)} d\varepsilon \right) \quad (5.40)$$

$$k_{ji,\text{CB}}(F) = \sigma_0 v_{th} \left(n(0) e^{-\beta\varepsilon_{ji}} + \beta \int_{\varepsilon} n(\varepsilon) \lambda(\varepsilon, F) e^{-\beta\varepsilon'_{ji}} d\varepsilon \right) \quad (5.41)$$

$$k_{ij,\text{VB}}(F) = \sigma_0 v_{th} \left(p(0) e^{-\beta\varepsilon_{ij}} + \beta \int_{\varepsilon} p(\varepsilon) \lambda(\varepsilon, F) e^{-\beta\varepsilon'_{ij}} d\varepsilon \right) \quad (5.42)$$

$$k_{ji,\text{VB}}(F) = \sigma_0 v_{th} \left(p(0) e^{-\beta\varepsilon_{ji}} e^{-\beta(E_v-E_f)} + \beta \int_{\varepsilon} p(\varepsilon) \lambda(\varepsilon, F) e^{-\beta\varepsilon'_{ji}} e^{-\beta(E'_v-E_f)} d\varepsilon \right) \quad (5.43)$$

There, the NMP barriers ε'_{ij} and ε'_{ji} are the field-dependent transition energies calculated from E_t and E'_c or E'_v respectively. In general, transitions to the local bands only contribute significantly to the overall NMP rates if the trap level is shallow or if the fields are very high. This model can be used as a more physical alternative to other TAT models like for example the Frenkel-Poole model [152] used to calculate barrier leakage currents in GaN/AlGaN HEMTs. Another benefit of this model is its seamless transition from a regime dominated by Fowler-Nordheim tunneling [141] into a regime where the defect properties dominate the tunneling currents [41, 42].

Chapter 6

Extraction of the Characteristic Time-Constants

This chapter deals with the extraction of the characteristic time-constants from the stochastic capture and emission events of RTN signals. [Section 6.1](#) will be dedicated to a short description of the basics of HMMs and their relevance for RTN producing single-charge defects. The following sections introduce the most basic case, a simple two-state defect ([Section 6.2](#)) followed by more complex cases, namely defects with multiple states ([Section 6.3](#)) and systems composed of several arbitrarily shaped defects ([Section 6.4](#)). In [Section 6.6](#) different histogram-based methods will be discussed, which allow extracting the time constants of certain defects from their stochastic capture and emission events. In the last part of this chapter, a method to extract the time constants of multiple defects with an arbitrary number of states, an algorithm to train a certain HMM to a set of observations, the Baum-Welch Algorithm will be introduced. After discussing the basics of the Baum-Welch algorithm, an implementation of a HMM library (see [Appendix A](#)) will be tested for its robustness against data sampled from a known system of defects.

6.1 Markov Processes and the Hidden Markov Model

Markov processes are widely used to describe stochastic transitions between two or more abstract states across many fields of science (physics, chemistry, speech recognition, robotics, etc.) [[110](#), [153](#), [154](#)].

In the real world, statistical processes produce observable signals which can be measured by some kind of device. In the case of charge transfer reactions in MIS-HEMTs, the charge cannot be measured directly, but only indirectly due to its electrostatic influence on the V_{th} . This fact potentially introduces noise in the measurements, which depends on the device itself, the measurement equipment and other systematic errors like the mapping from the drain or gate current to V_{th} [[89](#)].

Throughout the next sections, the following notation is used:

T ... the length of the observation sequence

N ... the number of states in the model

M ... the number of observable symbols

$Q = \{q_0, q_1, \dots, q_{N-1}\}$... the set of possible states of the Markov Model

$X = \{x_0, x_1, \dots, x_{T-1}\}$... the sequence of states from Q

$o = \{0, 1, \dots, M-1\}$... the set of possible observations

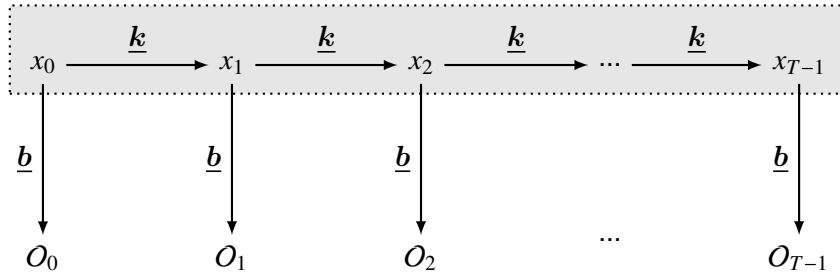


Figure 6.1: A schematic of the Hidden Markov Model. The hidden state sequence x_i is unknown and should be reconstructed from the corresponding observations O_i .

$O = \{O_0, O_1, \dots, O_{T-1}\}$... the sequence of observations from o

\underline{k} ... the state transition probability matrix

\underline{b} ... the observation probability matrix

π ... the initial state probabilities

The working principle can be seen in [Figure 6.1](#). The grey region denotes the inner state sequence X of the Markov Model, which can be one of the states x_i for each item in the observed sequence O . Note that the each inner state x_i can only be identified by its corresponding observation O_i .

As a subset of all stochastic processes, Markov processes can be described as a series of stochastic events, where each event x_t from a discrete state space \mathcal{Q} occurs at a certain time t . In general, the set of events X is described by:

$$X(t) = \{x_t\}, \quad t = 0 \dots T - 1 \quad (6.1)$$

Each of the events $X(t)$ is determined by its own cumulative distribution function (CDF),

$$F(x_t, t) = F(x_t) = P\{X(t) = x_t\}. \quad (6.2)$$

The CDF of the whole set X can be found by writing down the joint CDF for all T events:

$$F(x_0, x_1, \dots, x_{T-1}; t_0, t_1, \dots, t_{T-1}) = P\{X(t_0) = x_0, X(t_1) = x_1, \dots, X(t_{T-1}) = x_{T-1}\} \quad (6.3)$$

To actually construct the CDF for a given series of events, conditional probabilities are used to express the probability of the next observation, given a certain history of observations. In general, this can be a very complex task, since the conditional probability depends on all past observations. At that point, the so called Markov property helps to simplify the problem. It states, that for a Markov process, the conditional probability to enter the next state only depends the current state [[153](#), [154](#)]. With other words, Markov processes have no memory and thus the probability to reach a certain state at time $t + 1$ only depends on the current state:

$$P\{X(t + 1) = x_{t+1} | X(t) = x_t\} \quad (6.4)$$

In the context of defect capture and emission events we look at continuous-time discrete-space Markov processes, also called Markov chains, which will be used in the following section to calculate the probability density function (PDF) of a simple two-state defect.

The state transition probability matrix \underline{k} is of size $N \times N$. It contains the conditional probabilities to go from state i to state j and each of the rows sum to one because the probability of being in one of the states is one (i.e., \underline{k} is row stochastic). Note that in this case the braces only mark the

instants in time of the state sequence X as the transition probabilities of Markov chains are time independent.

$$k_{ij} = P\{X(t+1) = q_j | X(t) = q_i\} \quad (6.5)$$

The observation probability matrix \underline{b} is also row stochastic and time independent. It holds the probabilities to observe the symbol o_i given a certain state q_i . The size is $N \times M$, as the number of possible observations not necessarily reflects the number of inner states. One example would be thermal transitions of a defect without charge transfer.

$$b_i = P\{o_i(t) | q_i(t)\} \quad (6.6)$$

The HMM is fully defined by \underline{k} , \underline{b} and π , and is denoted by $\lambda(\underline{k}, \underline{b}, \pi)$.

6.2 Two State Defects

The most simple case of a RTN producing defect follows a two-state Markov chain with the defect being neutral in state 1 and negatively charged in state 2, see [Figure 6.2](#). The simulations were done with the Hidden Markov library presented in [Section 6.7](#), with a small amount of Gaussian noise added to the emissions. The Master equation of such a defect can be constructed from equation (6.3). For that, the chain is assumed to be in state 1 at time t . The conditional probability to go from state 1 to 2 is

$$P\{X(t+dt) = q_2 | X(t) = q_1\} = k_{12} dt. \quad (6.7)$$

On the other hand, the probability of staying in state 2 is

$$P\{X(t+dt) = q_2 | X(t) = q_2\} = 1 - k_{21} dt. \quad (6.8)$$

The probabilities $p_1(t)$ and $p_2(t)$ to be in state 1 and 2 at time t together with equations (6.7) and (6.8) give the probability to be in state 2 at time $t+dt$:

$$p_2(t+dt) = (k_{12} dt)p_1(t) + (1 - k_{21} dt)p_2(t) \quad (6.9)$$

After rearranging one obtains

$$\frac{p_2(t+dt) - p_2(t)}{dt} = \frac{dp_2(t)}{dt} = k_{12}p_1(t) - k_{21}p_2(t) \quad (6.10)$$

At any time, the charge needs to be in one of the states, giving $p_1(t) + p_2(t) = 1$. With the same approach for $p_1(t)$, the Master equation of the process can be written as [110]:

$$\frac{dp_1(t)}{dt} = k_{21}(1 - p_1(t)) - k_{12}p_1(t) \quad (6.11)$$

$$\frac{dp_2(t)}{dt} = k_{12}(1 - p_2(t)) - k_{21}p_2(t) \quad (6.12)$$

The solution of the Master equation describes the probability over time for the defect to be in a certain state. With the initial probabilities $p_1(0)$ and $p_2(0)$ and the characteristic time constant

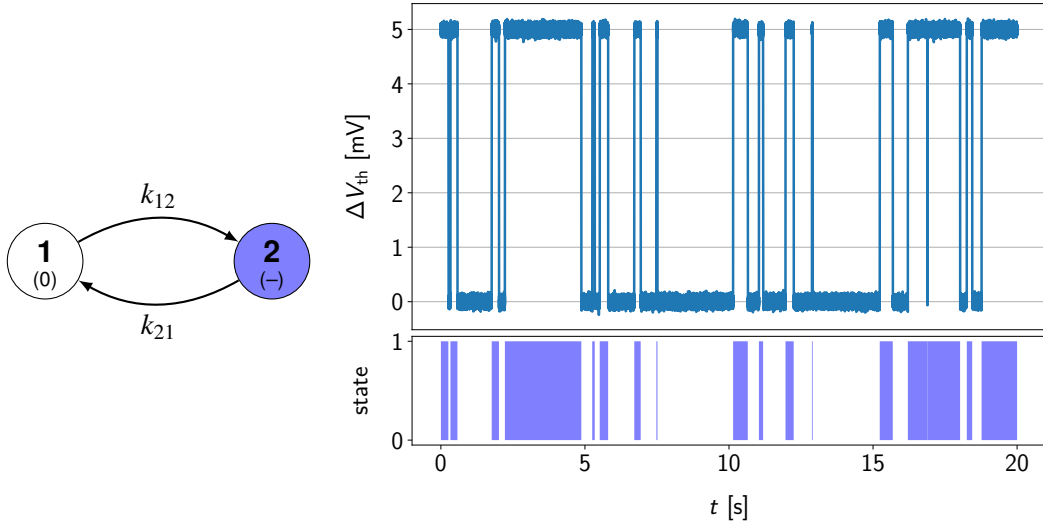


Figure 6.2: **Left:** The Markov chain of a simple two-state defect. The charge state is neutral if the defect is in state 1 and negative if in state 2. **Right:** Simulated emissions of the two-state defect for $k_{12} = k_{21} = 2 \text{ s}^{-1}$. The apparent differences in capture and emission times are only due to the stochastic nature of the process. The simulations were done with the Hidden Markov library presented in [Section 6.7](#), with a small amount of Gaussian noise added to the emissions.

$\tau = (k_{12} + k_{21})^{-1}$ of the defect it reads:

$$p_1(t) = \frac{k_{21}}{k_{12} + k_{21}} + \left(p_1(0) - \frac{k_{21}}{k_{12} + k_{21}} \right) \exp\left(-\frac{t}{\tau}\right) \quad (6.13)$$

$$p_2(t) = \frac{k_{12}}{k_{12} + k_{21}} + \left(p_2(0) - \frac{k_{12}}{k_{12} + k_{21}} \right) \exp\left(-\frac{t}{\tau}\right) \quad (6.14)$$

In the limit $t \rightarrow \infty$, the well-known results for the equilibrium probabilities (i.e. occupancies) of the two states are retrieved:

$$p_1 = \frac{k_{21}}{k_{12} + k_{21}} \quad (6.15)$$

$$p_2 = \frac{k_{12}}{k_{12} + k_{21}} \quad (6.16)$$

The Master equation can easily be expanded to describe more complex defects, as will be shown in [Section 6.3](#). However, in most cases one is not interested in the temporal evolution of the probabilities for being in a certain state, but rather in the PDF of the time it takes to change from one state to another. These are known as the first passage times and can also be obtained from the Master equation by setting the respective reverse rate to zero and the initial probability to be in the starting state to one. The first passage time to capture an electron is directly calculated from the Master equation:

$$\frac{dp_1(t)}{dt} = -k_{21}p_1(t) \quad \Rightarrow \quad p_1(t) = \exp(-k_{21}t) \quad (6.17)$$

Equation (6.17) can be used to calculate the random variable τ_{12} , which is the point in time where the transition actually takes place. The probability of the defect already being in state 2 at a certain time is given by $p_2(t) = 1 - p_1(t)$. Therefore τ_{12} must also be smaller than t . The

probability (and thus the CDF) of being in state 2 is then given by [110]

$$F(\tau_{12}) = P\{\tau_c \leq t\} = p_2(\tau_{12}) = 1 - \exp(-k_{21}\tau_{12}) \quad (6.18)$$

Exactly the same method can be applied to obtain the CDF of the emission time. Once the CDFs of the randomly distributed variables for capture and emission is known, the PDF can be obtained by taking the derive of the CDF. For the two-state defect, the PDF of the random variables for capture and emission τ_{12} and τ_{21} are given by [110]:

$$g(\tau_{12}) = k_{12} \exp(-k_{12}\tau_{12}) \quad (6.19)$$

$$g(\tau_{21}) = k_{21} \exp(-k_{21}\tau_{21}) \quad (6.20)$$

The random variables for the charge capture (τ_{12}) and emission (τ_{21}) times are thus exponentially distributed. The mean values of the observed times are given by the expectation values of the exponential distributions, which are the inverse rate constants of the process:

$$\bar{\tau}_{12} = \int_0^{\infty} \tau_{12} g(\tau_{12}) d\tau_{12} = \frac{1}{k_{12}} \quad (6.21)$$

$$\bar{\tau}_{21} = \int_0^{\infty} \tau_{21} g(\tau_{21}) d\tau_{21} = \frac{1}{k_{21}} \quad (6.22)$$

An estimator for the mean values can be obtained directly from measurements and are commonly used to specify the characteristic time constants of defects. This is done simply by averaging over the observed capture and emission times for a single defect. Naturally, the variance and the confidence interval of this estimator become smaller the more observations are available for averaging.

6.3 Multi State Defects

The procedure presented in the previous section can easily be generalized to more than two states. The Master equation for state i is given by the probabilities to go from state i to state j together with the probability to stay in state i [153]:

$$P\{X(t+dt) = q_j | X(t) = q_i\} = k_{ij} dt \quad (6.23)$$

$$P\{X(t+dt) = q_i | X(t) = q_i\} = 1 - \sum_{i \neq j} k_{ij} dt \quad (6.24)$$

From those probabilities, the Master equation is again obtained analogous to equation (6.10) using $dt \rightarrow 0$.

$$\frac{dp_i(t)}{dt} = -p_i(t) \sum_{i \neq j} k_{ij} + \sum_{i \neq j} k_{ji} p_j(t) \quad (6.25)$$

Since the charge has to be in one of the states, from N equations, only $N-1$ are independent from each other. The PDF of the first passage times for multi state defects can again be derived from the solution of the Master equation. For a three-state defect this is done in [110], while [153] gives more general approaches to derive analytic expressions for that problem in Chapter 6. In principle, the same procedure as presented in Section 6.2 can be used to derive the equilibrium first passage times for neighbouring states, if the correct occupancies are inserted. For non-neighbouring states, the PDFs are the normalized differences of exponential distributions, where the faster defect truncates the distribution of the slower defect [110]. Since both of the extraction

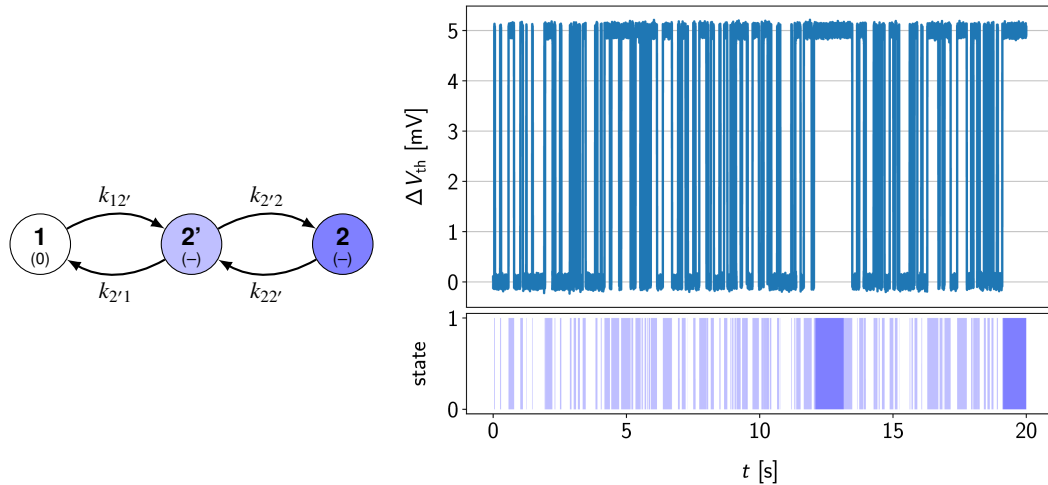


Figure 6.3: **Left:** The Markov chain of a three-state defect with a thermal state, producing anomalous RTN. The charge state is neutral if the defect is in state 1 and negative if in state 2' or 2. **Right:** Simulated emissions of the three-state defect for $k_{12'} = k_{2'1} = 10 \text{ s}^{-1}$ and $k_{2'2} = k_{22'} = 0.5 \text{ s}^{-1}$. Note that there is no straight-forward way to determine the thermal state directly from measurements. The simulations were done with the Hidden Markov library presented in [Section 6.7](#), with a small amount of Gaussian noise added to the emissions.

methods presented in [Sections 6.6](#) and [6.7](#) do not rely on the knowledge of the PDFs, the analytic expressions are omitted at this point.

Typical candidates for multistate defects are three-state defects producing two or three level anomalous RTN [[108](#), [112](#), [113](#)], see also [Section 4.2.1](#). In the case of two-level anomalous RTN as seen in [Figure 6.3](#), the third state possesses the same charge state as the second one. Such states are attributed to a structural relaxation of the defect in the NMP four-state model and will be referred to as thermal states throughout the remainder of this work. There is no straight-forward way to determine a thermal state from the measurements with the histogram methods described in [Section 6.6](#). With respect to Markov models they are also called tied states because they share the same emissions (and thus the same PDF for the emissions) with the state they are tied to.

An example for an even more complex RTN signal commonly called three-level anomalous RTN, is given in [Figure 6.4](#). Here the Markov state 2 is charged negatively, while the states 3' and 3 are charged double-negatively. Again, the states 3' and 3 cannot be separated easily, because they share the same charge state.

6.4 Multiple Defects

If several different defects are observed in measurements, additional steps are needed to construct the resulting Markov chain. The transition matrix of the system is then given by the Kronecker product of the individual defects (see [[155](#)]). For N independent defects, the resulting transition matrix is:

$$\underline{k} = \underline{k}_N \otimes \underline{k}_{N-1} \otimes \cdots \otimes \underline{k}_1 \quad (6.26)$$

Note, that the size of the system matrix equals the product of the matrix sizes of the individual defects. This can be a serious problem in terms of computation time, because the number of different states quickly grows to impractical sizes (for four three-state defects, the resulting system

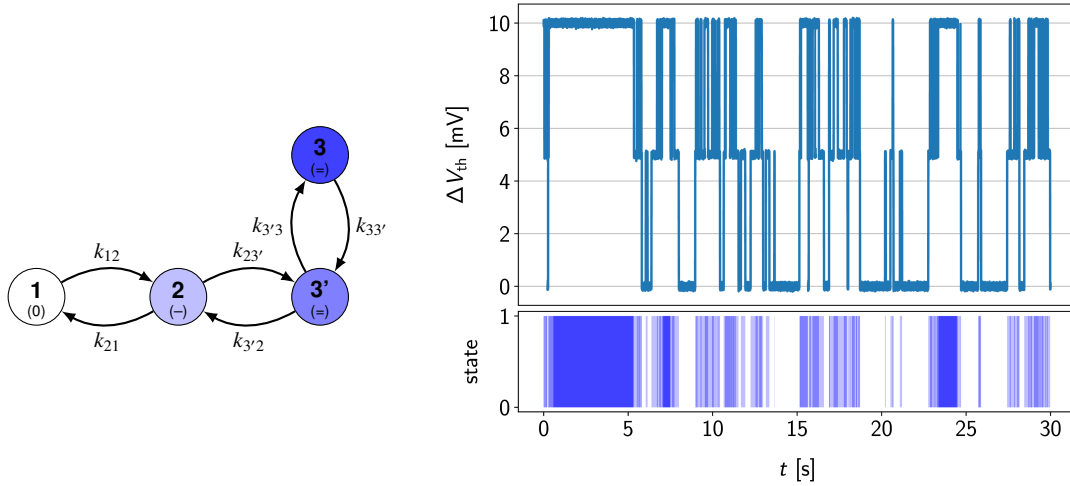


Figure 6.4: **Left:** The Markov chain of a three-level defect with a thermal state. The charge state is neutral if the defect is in state 1, negative if in state 2 and double negative in the states 3 and 3'. **Right:** Simulated emissions of the three-level defect for $k_{12} = k_{21} = 2 \text{ s}^{-1}$, $k_{23'} = k_{3'2} = 10 \text{ s}^{-1}$ and $k_{3'3} = k_{33'} = 0.5 \text{ s}^{-1}$. The simulations were done with the Hidden Markov library presented in Section 6.7, with a small amount of Gaussian noise added to the emissions.

would already have $N = 3^4 = 81$ different states). The number of different levels of each defect j when changing its charge only depends on its electrostatic influence on the device (usually given in terms of ΔI_D or ΔV_{th}) and the number of states with a charge transfer.

$$\underline{V}_{th}^j = \{q_i \cdot \Delta V_{th}\}, \quad i = 1 \dots (n - n_{th}) \quad (6.27)$$

Here, \underline{V}_{th}^j is the vector of different charge levels for defect j , ΔV_{th} is the electrostatic influence of one charge, n is the number of states of the defect, n_{th} is the number of thermal states, and q_i is the absolute charge state of the defect in state i . Note that the levels for the thermal states correspond to the states they are tied to. That means that for each thermal state, a duplicate entry has to be created at the correct position in (6.27) so that the size of the set is again n .

As an example, for a three-state defect without thermal state, the levels for $\Delta V_{th} = 3 \text{ mV}$ would be $\underline{V}_{th} = \{0 \text{ mV}, 3 \text{ mV}, 6 \text{ mV}\}$. On the other hand, the defect depicted in Figure 6.4 would possess the ΔV_{th} vector $\underline{V}_{th} = \{0 \text{ mV}, 3 \text{ mV}, 6 \text{ mV}, 6 \text{ mV}\}$.

To determine the corresponding levels for a system of N defects according to equation (6.26), the Cartesian product of the sets defined in (6.27) can be used.

$$\underline{V} = \underline{V}_{th}^1 \times \underline{V}_{th}^2 \times \dots \times \underline{V}_{th}^N \quad (6.28)$$

Since the resulting subsets in \underline{V} contain the corresponding levels of all defects for every state in (6.26), the final step is to sum over all the entries in each subset to obtain the vector \underline{V}_{th} with the actual voltage levels for all the states in \underline{k} (the indices here mark the N^{th} subset).

$$\underline{V}_{th} = \left\{ \sum V_1, \sum V_2, \dots, \sum V_N \right\} \quad (6.29)$$

To clarify the construction of the ΔV_{th} levels in (6.27) – (6.29), an example of a system consisting of a two-state defect with $\Delta V_{th} = 5 \text{ mV}$ and the three-state defect from Figure 6.4 with $\Delta V_{th} =$

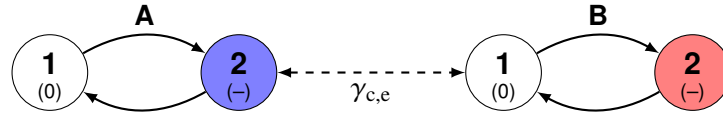


Figure 6.5: The Markov chains of a system of coupled two-state defects. The capture and emission rates of defect A are multiplied by γ_c and γ_e if defect B captured a charge.

5 mV should be given. Inserting \underline{V}_{th} of the defect into (6.28) gives

$$\begin{aligned} \underline{V} &= \{0 \text{ mV}, 5 \text{ mV}\} \times \{0 \text{ mV}, 3 \text{ mV}, 6 \text{ mV}, 6 \text{ mV}\} \\ &= \{[0 \text{ mV}, 0 \text{ mV}], [0 \text{ mV}, 3 \text{ mV}], [0 \text{ mV}, 6 \text{ mV}], [0 \text{ mV}, 6 \text{ mV}], \\ &\quad [5 \text{ mV}, 0 \text{ mV}], [5 \text{ mV}, 3 \text{ mV}], [5 \text{ mV}, 6 \text{ mV}], [5 \text{ mV}, 6 \text{ mV}]\}. \end{aligned} \quad (6.30)$$

Performing the summation for every sub-vector in (6.30) delivers the final defect levels of the combined system:

$$\underline{V}_{th} = \{0 \text{ mV}, 3 \text{ mV}, 6 \text{ mV}, 6 \text{ mV}, 5 \text{ mV}, 8 \text{ mV}, 11 \text{ mV}, 11 \text{ mV}\} \quad (6.31)$$

In principle, the combined transition matrix and the combined levels derived above are sufficient to train a HMM to find the most likely combined system (see Section 6.7). However, the entries in the combined transition matrix of the individual defects are not independent from each other. That means that two independent defects would converge towards one single defect with the combined number of states, which is of course unphysical (see Section 6.7.3). Another constraint is given by the ΔV_{th} levels of the defects depicted in (6.27) as they can only be multiples of the elementary charge for each state. As with the transition matrix for the combined system, the entries of \underline{V}_{th} in (6.29) are also not independent of each other.

These two constraints require modifications for the parameter update in the HMM training which will be explained in Section 6.7.3.

6.5 Coupled Defects

This section investigates the modeling of correlated RTN from two defects which are coupled to each other. While the physical origin of the coupling mechanism is of minor importance at that point, previous studies identified possible mechanisms as the Coulomb interaction between the defects or percolation path effects due to random dopants [112, 113, 156, 157, AGC5]. The RTN traces of such systems can help to determine empirical coupling factors, which then can be validated using TCAD simulations.

To determine the behaviour of coupled defects, the most basic case of two two-state defects A and B as seen in Figure 6.5 is chosen, but can be used for any type of Markov chain. The dashed line denotes the coupling with the factors γ_c for electron capture and γ_e for electron emission. The individual transition matrices of the two defects are:

$$\underline{k}_A = \begin{bmatrix} 1 - k_{12}^A & k_{12}^A \\ k_{21}^A & 1 - k_{21}^A \end{bmatrix}, \quad \underline{k}_B = \begin{bmatrix} 1 - k_{12}^B & k_{12}^B \\ k_{21}^B & 1 - k_{21}^B \end{bmatrix} \quad (6.32)$$

The transition matrix of the system is then given by:

$$\underline{k} = \underline{k}_B \otimes \underline{k}_A = \begin{bmatrix} (1 - k_{12}^B) \cdot \underline{k}_A & k_{12}^B \cdot \underline{k}_A \\ k_{21}^B \cdot \underline{k}_A & (1 - k_{21}^B) \cdot \underline{k}_A \end{bmatrix} \quad (6.33)$$

Let's assume that if for example defect B captures an electron, the capture and emission rates of defect A change. The modified transition matrix of defect A with the coupling factors γ_c and γ_e then is:

$$\underline{k}'_A = \begin{bmatrix} 1 - \gamma_c k_{12}^A & \gamma_c k_{12}^A \\ \gamma_e k_{21}^A & 1 - \gamma_e k_{21}^A \end{bmatrix} \quad (6.34)$$

Finally, the modified transition matrix \underline{k}'_A has to be inserted into (6.33) for the states where B has captured an electron. The transition matrix of the coupled system then becomes:

$$\underline{k} = \begin{bmatrix} (1 - k_{12}^B) \cdot \underline{k}_A & k_{12}^B \cdot \underline{k}_A \\ k_{21}^B \cdot \underline{k}'_A & (1 - k_{21}^B) \cdot \underline{k}'_A \end{bmatrix} \quad (6.35)$$

Figure 6.6 illustrates the differences between an uncoupled system and systems with various different coupling factors for electron capture and emission rates.

This method will be used in Section 7.2.4 to estimate the electrostatic coupling factors for correlated RTN signals recorded on GaN/AlGaIn fin-MIS-HEMTs.

6.6 Histogram-Based Methods

In single defect measurements, defects are commonly identified by clustering the observed step heights into spectral maps, see Section 4.2 and [110, 115–117, 158]. The most important difference in obtaining the stochastic times between TDDS and RTN measurements is that for TDDS measurements, emission times of a defect are equal to the points in time of the steps in the recovery traces. That means the clusters can be obtained in a rather straight-forward manner by printing the step-height distribution versus the observed emission times to a two-dimensional histogram (see Figure 4.4). For RTN signals, either the probability distribution of the signal amplitude when being in a certain state or the time differences between adjacent capture and emission events have to be calculated. The extraction of those can be rather challenging, especially for more complex RTN signals. Such signals can either be composed of defects with multiple emissions or emissions from different defects possessing similar step-heights [AGC5]. The most common conventional methods to extract time constants directly from measurements are done from histograms of the signal amplitudes or so-called lag plots, which are correlation plots between neighboring measurement samples. The main problem of those two methods is that they are quite prone to measurement noise and tend to miss smaller, non-dominant defects (i.e. the minor modes of the multimodal distributions) due to noise. They also rely on the absolute values of the measurement signal, which makes them sensitive to long-term drift on the RTN signal [133, 159, 160].

After a brief introduction to the histogram method in Section 6.6.1 and the lag plot method in Section 6.6.2, Section 6.6.3 presents a modified version of the histogram method for the extraction of time constants in TDDS measurements [116].

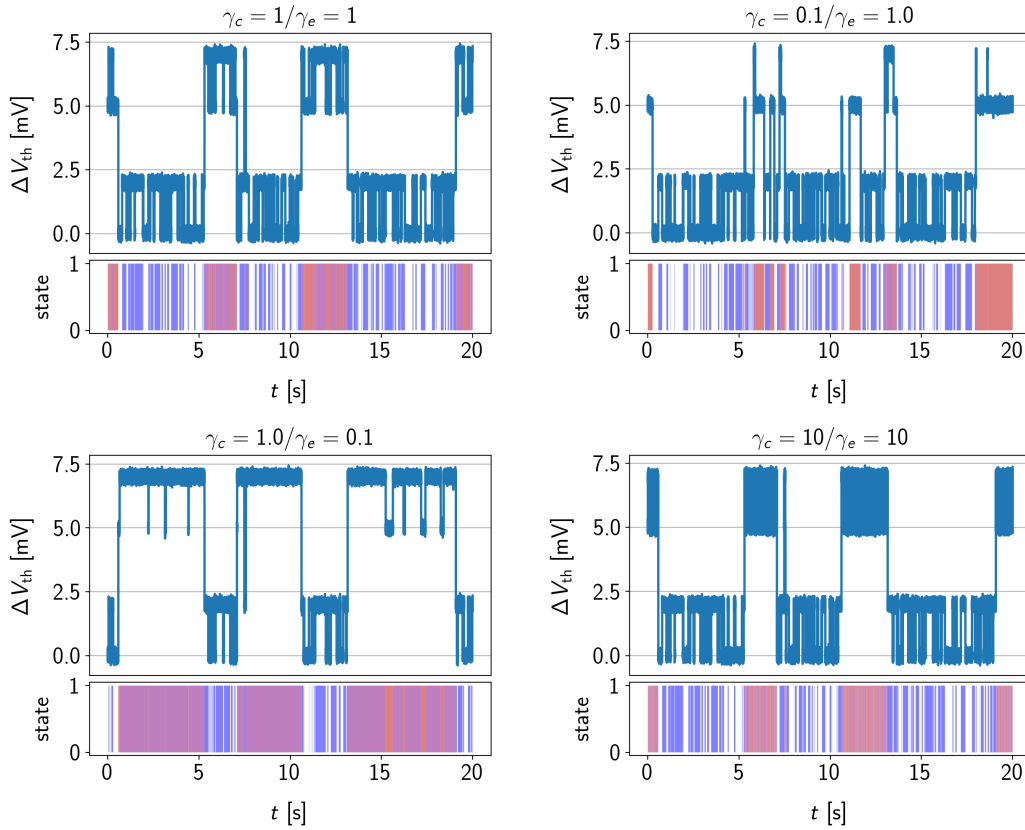


Figure 6.6: The RTN signals of a system with two coupled two-state defects according to Figure 6.5. The two defects A and B have step heights of 2 mV and 5 mV respectively. When the slow defect B captures an electron, the emissions rates of defect A are changed by a factor of $\gamma_{c,e}$. If the coupling factors γ_c for capture and γ_e are different from each other, the average occupancy between the coupled and uncoupled emissions changes. On the other hand, if $\gamma_c = \gamma_e$, the average occupancy is unchanged but the capture and emission times will change accordingly.

6.6.1 Amplitude Histograms

A method which makes use of the signal amplitude histograms to derive step heights and time constants of RTN producing defects was first introduced by Yuzhelevski and co-authors in 2000 [160]. In order to calculate the defect properties, the measurement values are directly binned into a histogram. The step-height(s) ΔV_{th} can in principle be derived directly from the observed amplitude distributions. If the number of samples is sufficiently large and the peaks of the Gaussian distributions are well separated, the capture and emission times of the (presumably) two-state defects can be calculated with the histogram areas A_L and A_H , the bin size ΔU of the Gaussians as well as the number of transitions k and the sampling time Δt [160] using

$$\tau_c = \frac{2}{k} \frac{A_L}{\Delta U} \Delta t, \quad (6.36)$$

$$\tau_e = \frac{2}{k} \frac{A_H}{\Delta U} \Delta t. \quad (6.37)$$

An example of a simulated RTN trace together with the resulting amplitude histogram are given in Figure 6.7. In general, for each ΔV_{th} level of the combined system, one Gaussian distribution will be present in the histogram. In order to extract the defect properties, multimodal Gaussians have to be fitted to the histogram in a maximum likelihood manner. The amplitude

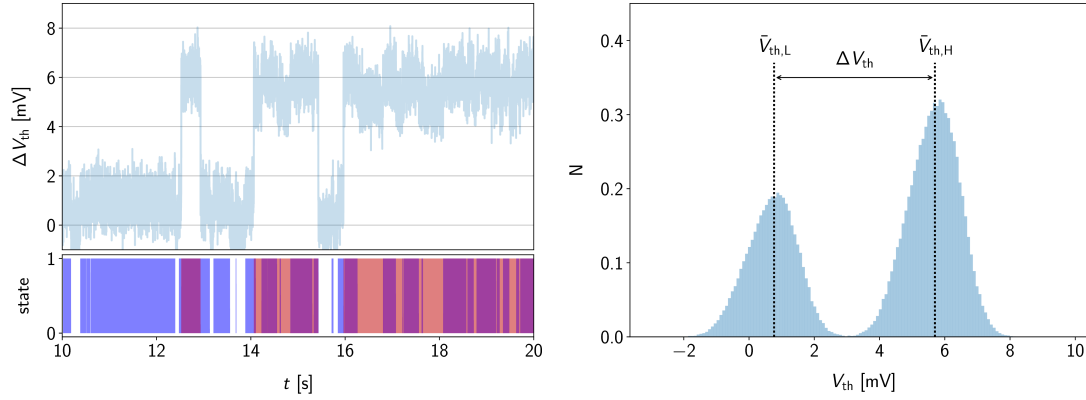


Figure 6.7: The simulation of two independent two-state defects with different amplitudes. **Left:** A sample from the time series shows the dominant defect with a step height of 5 mV as well as a small defect with a step height of 1 mV. The upper pictures show part of the time signal, while the lower pictures depict the color-coded state of the respective defect. The SNR of the small defect is around 3. **Right:** The histogram of the whole data set apparently consists of only two Gaussian distributions because the two maxima are only weakly separated.

signal was generated by two independent two-state defects with step heights of 5 mV and 1 mV. When the histogram is plotted, the separation between the distributions of the dominant and the non-dominant defect almost appear as a single Gaussian distribution, which can easily be misinterpreted as a single defect with levels around 6 mV and 1 mV (instead of four levels at 0 mV, 1 mV, 5 mV and 6 mV). Thus the proper identification of the trap levels in noisy signals is one of the major shortcomings of this method.

Another major problem is long-term drift in the signal as it will disperse the resulting histogram even more and would lead to misclassifications of defect states if the drift exceeds the step-heights of the defects. It can only be solved by preprocessing the data with some kind of baseline estimators. This problem is inherent to all methods using absolute signal values such as the lag plot method in [Section 6.6.2](#) and the HMM in [Section 6.7](#), where several efficient baseline estimation algorithms will be investigated and compared to each other.

6.6.2 Lag Plots

The lag plot method works by evaluating a correlation plot between neighboring measurement samples instead of binning them into a histogram (see [Figure 6.8](#) and [133, 159]). The amplitude levels will then show up as clusters, one for each transition in the combined Markov chain of the system. The main benefit of this method is that also the transitions from one state to another can be observed as off-diagonal elements.

The time constants can be estimated by calculating the average duration time of being in a certain state (i.e. via the number of points in the clusters along the main diagonal). The calculation can be carried out in the same way as (6.37). On this occasion, it should be noted that the fraction of area A and bin size ΔU in (6.37) can be identified by the number of points attributed to the respective distribution. For the lag plot that means that when there is no overlap between the clusters, it is sufficient to simply count the points inside a certain cluster. The step heights can again be derived directly from the lag plot.

If the clusters are not separated well enough, clustering algorithms like the K-means algorithm and others can help to assign the individual points to a state. However, the benefits and drawbacks

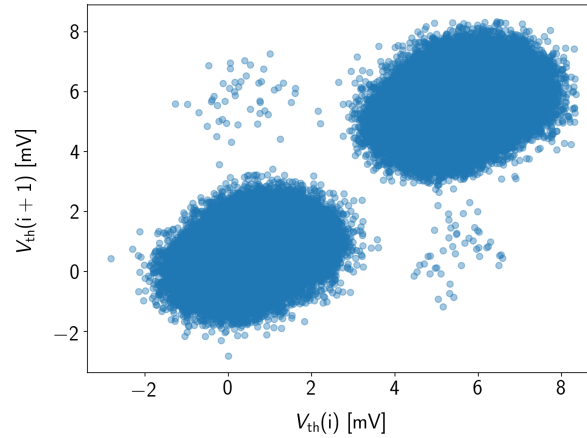


Figure 6.8: The lag plot of the RTN sequence from Figure 6.7 consisting of two independent two-state defects. Again, the clusters are not separated sufficiently to reliably determine all four states. The clusters along the main diagonal represent the probability of staying in a certain state while the off-diagonal elements are the actual transitions from one state to another.

of certain algorithms and their application to RTN signals are beyond the scope of this work and can be found in [158, 161].

Although variations of this method were successfully used to extract the properties of multiple defects [162, 163], the main drawbacks of the lag plot method are its lacking robustness against measurement noise and its inability to detect thermal states.

6.6.3 Spectral Maps

The method presented here is heavily based on the spectral maps generated for the TDDS presented in Section 4.2.2 and [110, 116, 117, 164]. Its main benefit is that it uses step heights instead of absolute levels. This helps to mitigate both of the major problems of the methods explained in the previous sections. The spectral maps method has already been used successfully to obtain single defect properties from RTN signals in GaN/AlGaIn fin-MIS-HEMTs [AGC5] and molybdenum disulfide (MoS₂) FETs [AGJ6].

To construct the spectral maps, first of all a step detection algorithm which delivers the step times, the step heights and the sign of the steps is employed. A universal step detection algorithm should not be not influenced by long-term drift of the RTN signal. Linear filtering of the extracted step heights often also helps to filter some of the measurement noise from the extracted step heights.

Step Detection and Clustering

One of the most popular algorithms for edge detection, the Canny edge detection algorithm, is able to fulfill both of these requirements with a relatively low computational effort [165].

The variant of the algorithm used for step detection throughout this work is based on the convolution of the original signal $\Delta V_{th}(t)$ with the derivative of a Gaussian distribution $g(t, \sigma_0)$ of width σ_0 .

$$f(t) = \Delta V_{th}(t) * \frac{d}{dt}g(t, \sigma_0) \quad (6.38)$$

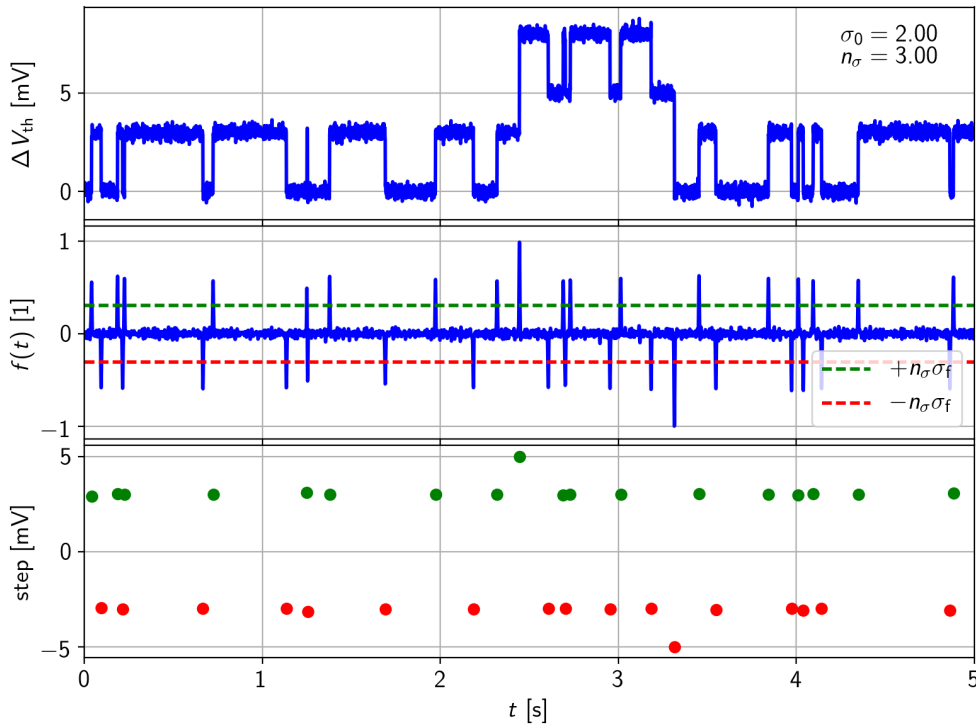


Figure 6.9: The Canny algorithm used for step detection throughout this work. In the upper picture a sample signal of two independent two-state defects is shown. After applying the linear Gauss filter, the long-term drift is removed and the steps are marked by the peaks of the resulting signal (middle picture). The corresponding times are found by a peak-detection algorithm and the step heights are found by taking the values of the original signal before and after the step (lower picture).

The resulting signal $f(t)$ is then free of long-term drift, with peaks for each sudden change in the signal (see middle picture in [Figure 6.9](#)). Further filtering is done by defining a threshold value for the signal $f(t)$ being $\pm n_\sigma \sigma_f$ with the standard deviation of the filtered signal σ_f . A simple peak finding algorithm then selects the instance in time for all peaks above the threshold values. Thus the set of step times is

$$t_s = \text{local max.}\{|f(t)| > n_\sigma \sigma_f\}. \quad (6.39)$$

Note that due to the linear Gauss filter, the step times will be shifted to the right by σ_0 . This can either be corrected manually, or the algorithm is applied for a second time with the time-inversed signal and the resulting step times are averaged. The step sizes for low-noise signals can simply be calculated by taking the values of the original signal before and after each step $t_{s,i}$.

$$\Delta V_{\text{th},i} = \Delta V_{\text{th}}(t_{s,i} + \Delta t) - \Delta V_{\text{th}}(t_{s,i} - \Delta t) \quad (6.40)$$

For this naive approach to obtain ΔV_{th} , the noise in the extracted step heights is the same as the measurement noise. Thus, additional filtering of the step heights (i.e. mean or median filters, splines) potentially can be of advantage for very noisy signals. The results of the step-detection algorithm can be seen in [Figure 6.9](#). In the upper picture, the original RTN signal is shown. In the middle picture of [Figure 6.9](#), the resulting signal from (6.38) to extract the step times together with the applied threshold values can be seen. Finally, the lower picture gives the extracted steps after inserting the values from (6.39) into (6.40).

The free parameters σ_0 and n_σ of the algorithm can be used to adjust the step detection to different signals:

- A larger value of σ_0 decreases the sensitivity to short-term peaks in the signal and thus helps to suppress measurement noise. At the same time the sensitivity for fast defects decreases.
- A larger value of n_0 also helps to suppress measurement noise but decreases the sensitivity for defects with small values of ΔV_{th} .

As the corresponding steps sizes of capture and emission events have opposite signs, the data from the detected steps then have to be clustered by the absolute value of the step sizes. Subsequently, the time differences between adjacent capture and emission events of each cluster are calculated. Any algorithm for the extraction of the times is highly dependent on the assumed Markov chain of the defect (more generally, the individual Markov chains of the defects for a system with multiple defects) and the accuracy of the step detection. Usually, some kind of state-machine is used to relate a certain capture event to a certain emission event and vice versa.

In this case, the steps are clustered with a K-means algorithm, developed independently by four different research groups [166–169]. The step data is divided into positive (electron capture) and negative (electron emission) events. Initial conditions for the algorithm are the assumed number of defects and the initial values of the assumed step heights, which then delivers the time sorted data to a simple state machine. Afterwards, the state machine calculates the time differences between two subsequent steps within one cluster if the two adjacent step directions possess different signs. Otherwise, if for example two positive steps occur after each other, the second one of the double-steps will be dropped. As this simple state machine does not allow two subsequent capture or emission events of the same defect, it is thus only suitable for extracting two-state defects.

Time Constants Extraction

In [Figure 6.10](#), the construction of the spectral maps and the subsequent extraction of the time constants is shown using the same two defects as in the previous sections. The upper picture shows parts of the RTN signal together with the results from the state machine. The discontinuities in the extracted signal are due to dropped double capture or emission events which emerge from inaccuracies in the step detection. The lower left picture gives the calculated time differences plotted into a spectral map. To obtain the characteristic times, an exponential distribution as given in (6.20) is fitted to each cluster. In the lower right picture, the extracted values, the histograms and the fitted distributions can be seen. The extracted time constants for the fast defect ($\Delta V_{th} = 1$ mV, $\tau_c = 0.1$ s, $\tau_e = 0.2$ s) and the slow defect ($\Delta V_{th} = 5$ mV, $\tau_c = 2$ s, $\tau_e = 3$ s) in [Figure 6.10](#) are both within 20% error. This is actually a quite remarkable accuracy since both histograms suffer from opposite problems:

- The fast defect has many emissions and thus a good statistical foundation. The step detection is however very noisy due to the low value of ΔV_{th} , producing false-positives and missed steps in the clustering algorithm.
- The histograms of slow defect, on the other hand, contain quite few samples but very accurate clusters for capture and emission times.

In other words, errors due to inaccuracies in the step detection (and thus in the clustering) can be made up by a larger sample size while errors stemming from a small sample size can be made up by an accurate step detection.

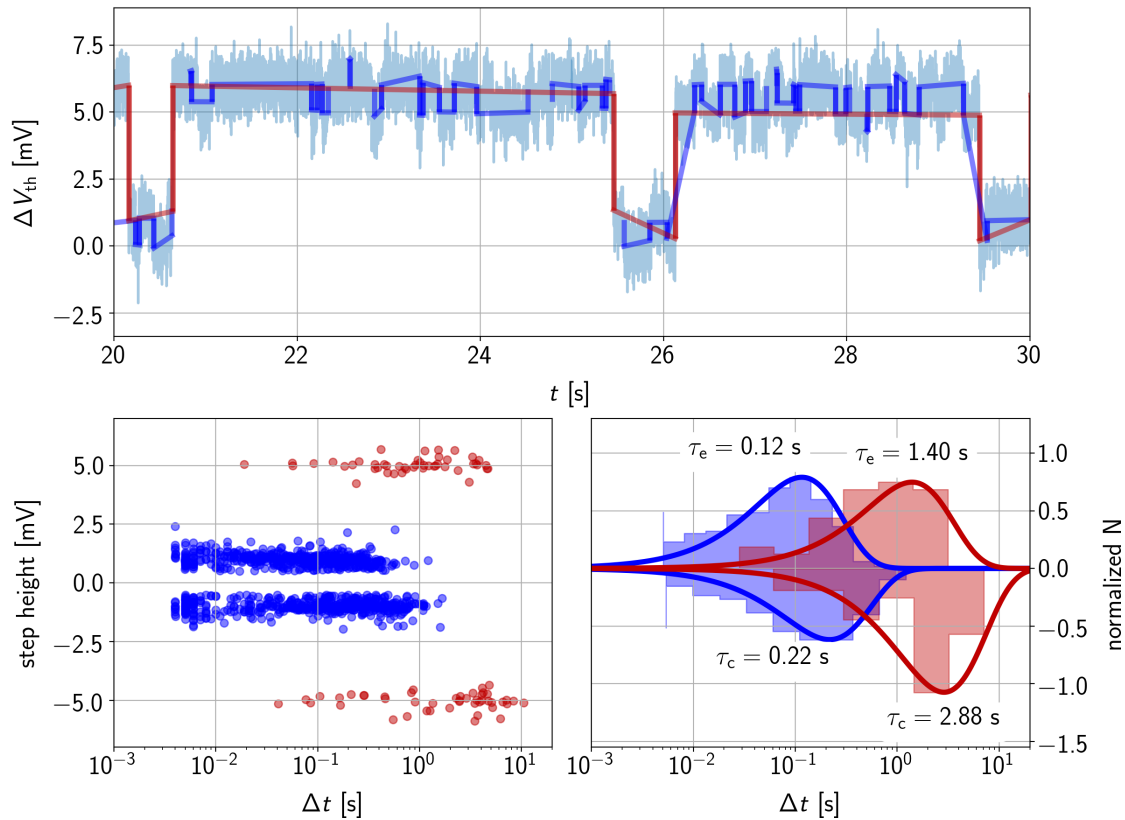


Figure 6.10: The upper picture shows part of a RTN signal generated by the same two defects used in Sections 6.6.1 and 6.6.2. After step detection, a simple state machine calculates the time differences between adjacent capture and emission events. These are shown as spectral maps in the lower left picture. To obtain the mean capture and emission times, exponential distributions are fitted to the clusters in the spectral maps (lower right).

Problems

At a first glance, creating such state machines may look trivial, but for measurement data with a low SNR, multi-state defects as presented in Figure 6.4, or systems containing multiple defects with weakly separated step heights it can be prone to errors. As an example, we look at a simulation of two independent two-state defects with different levels of noise (see Figure 6.11). The step heights were set to 5 mV and 3 mV, the capture and emission times were 2 s/3 s for the 5 mV defect and 0.1 s/0.2 s for the 3 mV defect. The sampling rate was set to 1 kHz. After step detection, if the signal-noise ratio is sufficiently high, this allows to reliably extract the capture and emission times. On the other hand, if the signal quality is poor, errors in the step detection (i.e. missed steps or noisy step heights) will quickly degrade the accuracy of the extracted delta times.

Another problem with all of the methods presented in Sections 6.6.1 to 6.6.3 is that defects with thermal states (i.e. states with no charge transition) cannot be identified. A minor shortcoming of the spectral method is also that it is prone to errors if the bias regions with RTN activity from different defects with similar step heights overlap. This can be seen at the border regions of the two defect pairs (compare voltage regions in Figure 7.15). An example is shown in Figure 6.12, where signals of the overlapping regions from the defects investigated in Section 7.2 are plotted. A state machine for the extraction of the delta times as explained before will probably fail to map the observed emissions to the defects correctly. The impact of these falsely assigned emissions

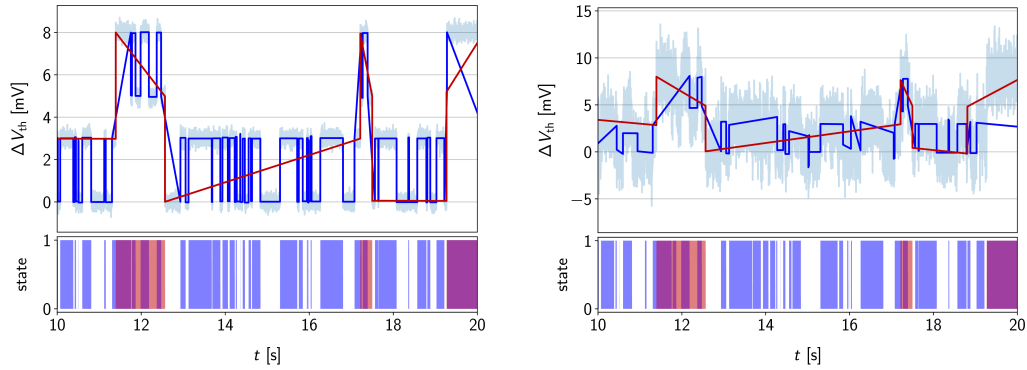


Figure 6.11: Simulation of two independent two-state defects with the same capture and emission times. The upper pictures show a part of the trace with the reconstructed RTN signal, while the lower pictures show the state of the respective defect. **Left:** The signal to noise ratio is sufficiently high to accurately extract the delta times. **Right:** For noisy signals, the step detection will either miss steps or the extracted step-heights will be affected too much by the noise. The apparent double capture or emission events show up as holes in the reconstructed RTN signals. This will put random errors on the extracted delta times, which will in turn degrade the quality of the extracted characteristic time constants.

on the overall result highly depends on the cumulated number of capture and emission events. Thus defects with relatively slow time constants will be affected most.

Finally, the maximum time constants which can be extracted is practically limited by half of the recorded time of one measurement because at least one full transition on average is needed to calculate the time differences.

To be able to reliably quantify thermal states, multi-state defects and RTN produced by several different defects, a more robust method for the extraction of the characteristic time constants is needed. One method able to overcome most of the aforementioned shortcomings of the spectral method has been known for a long time in the context of machine learning and speech recognition. It is called the Baum-Welch algorithm which is able to train a HMM to a set of observations in a maximum-likelihood manner and will be explained in the next section.

6.7 Hidden Markov Model Training

In order to find the maximum-likelihood solution for a given Markov chain of some defect (or any other Markov chain), three problems need to be solved:

- Given a HMM $\lambda(\underline{k}, \underline{b}, \pi)$ and a sequence of observations \mathcal{O} , the likelihood of the observed sequence $P(\mathcal{O}|\lambda)$ should be determined. This problem is usually referred to as the forward algorithm, see [Section 6.7.1](#).
- Given a HMM $\lambda(\underline{k}, \underline{b}, \pi)$ and a sequence of observations \mathcal{O} , the optimal state sequence X of the underlying Markov model should be found. This problem is usually referred to as the backward algorithm, see [Section 6.7.2](#).
- Given a sequence of observations \mathcal{O} , the HMM $\lambda(\underline{k}, \underline{b}, \pi)$ which maximizes the probability $P(\mathcal{O}|\lambda)$ should be found. This problem is usually referred to as the parameter update, see [Section 6.7.3](#).

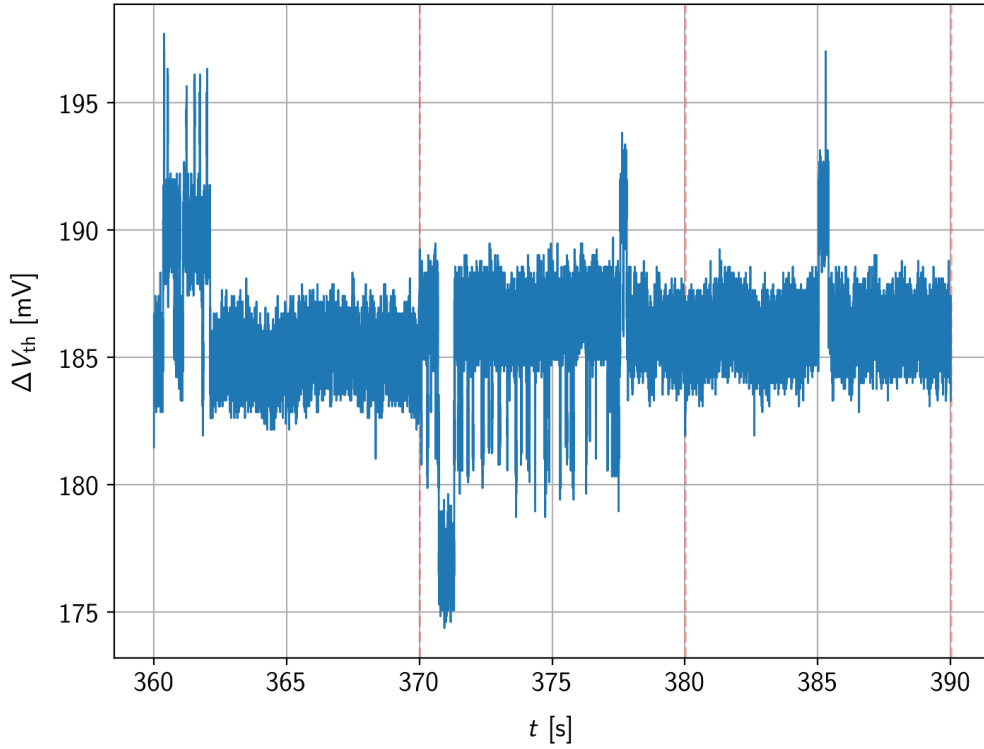


Figure 6.12: Traces with overlapping RTN activity coming from different defects. The overlapping regions can easily be identified by the different duty cycle of the observed signals. The step-height based time extraction algorithm will probably assign both of the signals to a single defect.

In the literature, the iterative solution to these three distinct problems is often called the Baum-Welch algorithm which can be seen as an expectation maximization (EM) algorithm for this particular problem [170, 171]. The following description closely follows the excellent tutorials given in [172, 173].

6.7.1 Forward Algorithm

The problem related to the forward algorithm is to determine the likelihood $P(O|\lambda)$ of observing the sequence O given the model $\lambda(\underline{k}, \underline{b}, \pi)$.

For a certain sequence X of statistically independent inner states, by the definition of \underline{b} the probability to observe O given a sequence and a model is given by:

$$P(O|X, \lambda) = b_{x_0}(O_0)b_{x_1}(O_1) \dots b_{x_{T-1}}(O_{T-1}) \quad (6.41)$$

On the other hand, from the definition of \underline{k} and π it follows that the probability of having sequence X is

$$P(X|\lambda) = \pi_{x_0}k_{x_0,x_1}k_{x_1,x_2} \dots k_{x_{T-2},x_{T-1}}. \quad (6.42)$$

The joint probability to observe O and X simultaneously is simply the product of (6.41) and (6.42).

$$P(O, X|\lambda) = P(O|X, \lambda)P(X|\lambda) \quad (6.43)$$

To obtain the probability of \mathcal{O} , one simply has to sum over all possible state sequences:

$$\begin{aligned} P(\mathcal{O}|\lambda) &= \sum_X P(\mathcal{O}|X, \lambda)P(X|\lambda) \\ &= \sum_X \pi_{x_0} k_{x_0, x_1} b_{x_0}(\mathcal{O}_0) k_{x_1, x_2} b_{x_1}(\mathcal{O}_1) \dots k_{x_{T-2}, x_{T-1}} b_{x_{T-1}}(\mathcal{O}_{T-1}) \end{aligned} \quad (6.44)$$

The problem of this direct approach is its computational expense since a direct calculation would require about $\mathcal{O}(N^T)$ operations. Fortunately, the problem can also be solved more efficiently by using the so-called forward algorithm, which often also is called α pass. It calculates the probability of a partial observation sequence up to time t , at which the Markov Process is in state q_i :

$$\alpha_t(i) = P(\mathcal{O}_0, \mathcal{O}_1, \dots, \mathcal{O}_t, x_t = q_i | \lambda) \quad (6.45)$$

The main benefit of this definition is that it can be calculated recursively which reduces the number of operations down to $\mathcal{O}(N^2T)$:

1. For all states $i = 0, 1, \dots, N - 1$ calculate $\alpha(i) = \pi_i b_i(\mathcal{O}_0)$.
2. For all states $i = 0, 1, \dots, N - 1$ and times $t = 1, 2, \dots, T - 1$ calculate:

$$\alpha_t(i) = \left[\sum_{j=0}^{N-1} \alpha_{t-1}(j) k_{ji} \right] b_i(\mathcal{O}_t) \quad (6.46)$$

3. From the definition (6.45) it follows that:

$$P(\mathcal{O}|\lambda) = \sum_{i=0}^{N-1} \alpha_{T-1}(i) \quad (6.47)$$

Now that the probability to observe a specific sequence of observations for a certain model is known, the hidden part of the HMM should be uncovered. In general, this step is not always necessary as for the classification of data, for example in speech recognition, one could train different models for different words and find the most likely word by performing the forward algorithm on each of the HMMs. How to train models (i.e find the optimal set of parameters) will be discussed in [Section 6.7.3](#). The difference when looking at defects is that usually the information which defect causes a certain set of observations is of minor value. The major interest is to find out the mean transition rates and the sequence of hidden states of a defect with a certain Markov chain as well as its specific threshold voltage shift. For that, the so-called backward algorithm is introduced.

6.7.2 Backward Algorithm

The backward algorithm is designed to find the most likely state sequence given a certain model and a sequence of observations. The forward algorithm defined the probability to observe a sequence up to time t . Now the probability to observe the rest of the sequence (i.e. from $t + 1$ to $T - 1$) when being in a specific state at time t is sought. The definition is similar to the forward algorithm with the only difference that it starts with the end of the sequence and progresses backwards in time:

$$\beta_t(i) = P(\mathcal{O}_{t+1}, \mathcal{O}_{t+2}, \dots, \mathcal{O}_{T-1}, x_t = q_i | \lambda) \quad (6.48)$$

Again, the computationally most efficient way is to calculate β recursively as follows:

1. For all states $i = 0, 1, \dots, N - 1$ set $\beta_{T-1}(i) = 1$.
2. For all states $i = 0, 1, \dots, N - 1$ and times $t = T - 2, T - 3, \dots, 0$ calculate:

$$\beta_t(i) = \sum_{j=0}^{N-1} k_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \quad (6.49)$$

It should be noted at that point that there are different possible interpretations of ‘most likely’. One is to find the single most likely state sequence for each of the observed sequences, which is done with the Viterbi algorithm [174]. Another often used criterion is to maximize the expected number of correct states across all sequences. This is not necessarily the same, as the latter criterion could lead to impossible state sequences because it selects the most likely state without taking into regard the probability of the occurrence of the whole state sequence. However, this optimality criterion should be preferred over the Viterbi algorithm, as it delivers the probabilities of being in each of the states for every instant in time. The matrix holding the probabilities to be in each state for every instant in time with size $N \times T$ is usually referred to as the emission matrix and allows to make soft assignments to states. This means that the exact probabilities are stored for each of the states instead of the Viterbi algorithm, which assigns 1 to the most likely state and 0 to all others. This is the reason why this method is usually preferred for **HMM** training.

The probability of being in state q_i at time t within the **HMM** λ can be simply expressed in terms of the forward and backward variables as the forward variable defines the relevant probability up to time t and β the relevant probability after time t :

$$\gamma_t(i) = P(x_t = q_i | O, \lambda) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=0}^{N-1} \alpha_{T-1}(j)} \quad (6.50)$$

If the values of $\gamma_t(i)$ are stored, the forward and backward algorithm together deliver an emission matrix sized $T \times N$ holding the normalized probabilities of being in state q_i . If the actual most likely path is sought, one only needs to take the maximum probability across the states i for every instant in time.

$$\hat{q}(t) = \underset{i}{\operatorname{argmax}}[\gamma_t(i)] \quad (6.51)$$

Up to now, the assumed **HMM** was fixed. In order to be able to train the model, the last and most important step is to find new estimates for the parameters in order to maximize the probability of the observation sequence.

6.7.3 Parameter Update

In general, the parameter update should maximize the probability $P(O|\lambda)$ for an observed sequence given a **HMM** λ . In theory this could be done analytically by maximizing (6.44). For practical applications, this approach is, however, infeasible due to the exponentially growing terms in the sum. Instead of using $P(O|\lambda)$ directly, the parameter updates derived in this section maximize Baum’s auxiliary function given in (6.52) (see [171, 172]).

$$Q(\lambda, \hat{\lambda}) = \sum_{\mathbf{X}} P(\mathbf{X} | O, \lambda) \log[P(O, \mathbf{X} | \hat{\lambda})] \quad (6.52)$$

It has been proven that the maximization of $Q(\lambda, \hat{\lambda})$ also leads to an increased likelihood $P(O|\hat{\lambda})$ [171, 172], i.e.

$$\max_{\hat{\lambda}} [Q(\lambda, \hat{\lambda})] \rightarrow P(O|\hat{\lambda}) \geq P(O|\lambda). \quad (6.53)$$

One of the most appealing properties of HMMs is the fact that the parameters can be efficiently re-estimated by the model itself. As the number of states N and the number of observations M are fixed, only the transition matrix \underline{k} , the emission matrix \underline{b} , and the starting probabilities π need to be updated.

For that, one first needs to find the probabilities of being in state q_i at t and going to q_j at $t + 1$. In terms of the model parameters \underline{k} , \underline{b} , α and β they can be expressed as

$$\begin{aligned} \gamma(i, j) &= P(x_t = q_i, x_{t+1} = q_j | O, \lambda) \\ &= \frac{\alpha_t(i) k_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \end{aligned} \quad (6.54)$$

where $P(O|\lambda)$ can be calculated from (6.47). On a sidenote, (6.54) is related to (6.50) by

$$\gamma_t(i) = \sum_{j=0}^{N-1} \gamma_t(i, j). \quad (6.55)$$

The quantities $\gamma_t(i)$ and $\gamma_t(i, j)$ already allow to update all three model parameters by summation over time. For the update of \underline{k} , the sum over the numerator is then the expected number of transitions from i to j whereas the sum over the denominator holds the expected number of transitions away from state i (including transitions to state i).

$$\hat{k}_{ij} = \frac{\sum_{t=0}^{T-2} \gamma_t(i, j)}{\sum_{t=0}^{T-2} \gamma_t(i)} \quad (6.56)$$

The update of the observation probability matrix \underline{b} to observe the symbol o_i while being in state j is:

$$\hat{b}_j(o_i) = \frac{\sum_{t=0}^{T-1} \gamma_t(j) \delta(O_t = o_i)}{\sum_{t=0}^{T-1} \gamma_t(j)} \quad (6.57)$$

Here, the Kronecker delta $\delta(O_t = o_i)$ is used to mark the states where the HMM is in state j and the symbol o_i is observed. The ratio in (6.57) thus is the expected number of times the model is in state q_j with the observation o_i over the expected number of times the model is in state q_j . For the initial probabilities, simply the values of $\gamma_0(i)$ are used.

$$\hat{\pi}_i = \gamma_0(i) \quad (6.58)$$

It was proven by Baum and his colleagues, that the HMM with the parameters re-estimated in such a way is either more or equally likely compared to the previous model [175]. The presented algorithm, namely the forward-backward algorithm together with the parameter update is often called Baum-Welch algorithm.

On this occasion, it has to be said that this algorithm is sensitive to local minima in the parameter space. Thus a randomization of the initial parameters is often needed to ensure good training

results. Another practical problem are numerical implementation issues, as for example the forward parameters α exponentially approach 0 as T increases. To avoid underflow and at the same time maintaining the validity of the formulas for the parameter update, the variables should be transformed in a certain manner as can be found in [173].

Multiple Defects

In order to train a HMM of a system with multiple defects, two major adjustments to the presented EM algorithm are necessary. One was already mentioned briefly in Section 6.4, namely that the states of the combined transitions matrix of multiple defects are not independent of each other. Since the Baum-Welch algorithm treats all states as independent, direct training would inevitably lead to a drift of the system towards one multi-state defect because of the multiplicative terms from each defect in (6.59).

To mitigate this problem, the algorithm for the parameter update has to be modified. First of all the combined expected transition matrix delivered by (6.56) needs to be split according to the structure of the underlying defects. To illustrate that we look again at the combined transition matrix of two two-state defects as given in (6.33). For clarity, this time the full combined transition matrix is listed. As the transition matrices of the individual defects have to be row stochastic, the relations $k_{11} = (1 - k_{12})$ and $k_{22} = (1 - k_{21})$ hold.

$$\underline{k} = \begin{bmatrix} k_{11}^B k_{11}^A & k_{11}^B k_{12}^A & k_{12}^B k_{11}^A & k_{12}^B k_{12}^A \\ k_{11}^B k_{21}^A & k_{11}^B k_{22}^A & k_{12}^B k_{21}^A & k_{12}^B k_{22}^A \\ k_{21}^B k_{11}^A & k_{21}^B k_{12}^A & k_{22}^B k_{11}^A & k_{22}^B k_{12}^A \\ k_{21}^B k_{21}^A & k_{21}^B k_{22}^A & k_{22}^B k_{21}^A & k_{22}^B k_{22}^A \end{bmatrix} \quad (6.59)$$

To find the expected number of transitions of a certain state from a distinct defect, the sum of all transitions involving this particular state in the combined transition matrix has to be calculated. The rows in (6.59) thereby mark the state from which the transition starts, whereas the columns mark the ending state. The usage of masking matrices provides an efficient way to select the states of interest. For that, when constructing the combined matrix, a mask matrix for each state of each defect is calculated. For defect 'A' those masks would be:

$$\underline{\delta}_1^A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \underline{\delta}_2^A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad (6.60)$$

To find the states in \underline{k} where for example defect A has a transition from i to j , the index-matrix is simply the mask of the starting state times the transposed mask of the ending state. The total number of transitions is obtained by summing over all entries of the selected sub-array.

$$k_{ij}^A = \sum \underline{k} \left[\underline{\delta}_i^A \cdot \underline{\delta}_j^{A^T} \right] \quad (6.61)$$

For the transition rate k_{12}^A , equation (6.61) reads:

$$\begin{aligned}
k_{12}^A &= \sum \underline{k} \left[\underline{\delta}_1^A \cdot \underline{\delta}_2^{A^T} \right] \\
&= \sum \begin{bmatrix} k_{11}^B k_{11}^A & k_{11}^B k_{12}^A & k_{12}^B k_{11}^A & k_{12}^B k_{12}^A \\ k_{11}^B k_{21}^A & k_{11}^B k_{22}^A & k_{12}^B k_{21}^A & k_{12}^B k_{22}^A \\ k_{21}^B k_{11}^A & k_{21}^B k_{12}^A & k_{22}^B k_{11}^A & k_{22}^B k_{12}^A \\ k_{21}^B k_{21}^A & k_{21}^B k_{22}^A & k_{22}^B k_{21}^A & k_{22}^B k_{22}^A \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\
&= \sum \begin{bmatrix} 0 & k_{11}^B k_{12}^A & 0 & k_{12}^B k_{12}^A \\ 0 & 0 & 0 & 0 \\ 0 & k_{21}^B k_{12}^A & 0 & k_{22}^B k_{12}^A \\ 0 & 0 & 0 & 0 \end{bmatrix} = k_{11}^B k_{12}^A + k_{12}^B k_{12}^A + k_{21}^B k_{12}^A + k_{22}^B k_{12}^A
\end{aligned} \tag{6.62}$$

It can be seen that the mask matrices $\underline{\delta}_1^A$ and $\underline{\delta}_2^A$ simply select all the terms of the combined transition matrix, where a transition from state 1 to state 2 of defect A is observed. The sum over those entries thus is the expected number of transitions for a certain transition of a specific defect.

After iterating over all states of the respective defect, the transmission matrix has to be made row stochastic again by dividing each row by the respective row sum. The dependencies of the combined transition matrix are preserved by recalculating the combined matrix from the individual defects.

The second adjustment needs to be done during the update of ΔV_{th} , as the observed levels are also not independent of each other. The forward-backward algorithm already delivers the emission matrix $\underline{\gamma}$ holding the probabilities of being in state i at each instant time (see 6.50). The expected number of emissions (i.e. the number of samples) for each state i is given by

$$\Gamma(i) = \sum_t \gamma_t(i). \tag{6.63}$$

If the measurement sequence is O , the mean values of ΔV_{th}^i of the defect (including the offset) are the weighted average of the observations over time.

$$\hat{\mu}(i) = \frac{\sum_t \gamma_t(i) O_t}{\Gamma(i)} \tag{6.64}$$

From the definition of the variance it follows that:

$$\hat{\sigma}^2(i) = E(O_i^2) - E(O)^2 = \frac{\sum_t \gamma_t(i) O_t^2}{\Gamma(i)} - \hat{\mu}^2(i) \tag{6.65}$$

$$\tag{6.66}$$

Multiple Sequences

If the means and standard deviations of the different states are calculated across multiple observation sequences and the amplitude noise distribution is assumed to be the same across them, sample-based statistics allows to calculate estimates for the aggregated means and variances for

each state. Here, i marks the individual state while the index j stands for the result of j^{th} observation sequence.

The pooled mean $\hat{\mu}_p(i)$ of a state i across all the sequences j can easily be calculated from the weighted average of the means $\hat{\mu}_j(i)$ of the sequences.

$$\hat{\mu}_p(i) = \frac{\sum_j \Gamma_j(i) \hat{\mu}_j(i)}{\sum_j \Gamma_j(i)} \quad (6.67)$$

The pooled variance $\hat{\sigma}_p(i)^2$ of all the sequences for a certain state can be shown to be the mean of the variances plus the variance of the means of the individual sequences. When including Bessel's correction, which gives an unbiased estimator of the of the pooled variance by multiplying the uncorrected sample variance with a factor $n/(n-1)$, the equation reads

$$\hat{\sigma}_p^2(i) = \frac{\sum_j [\Gamma_j(i) - 1] \sigma_j^2(i)}{[\sum_j \Gamma_j(i)] - 1} + \frac{\sum_j \Gamma_j(i) [\hat{\mu}_j(i) - \hat{\mu}_p(i)]^2}{[\sum_j \Gamma_j(i)] - 1}. \quad (6.68)$$

After again using the relation $E[(x - E[x])^2] = E[x^2] - E[x]^2$ and rearranging one obtains

$$\hat{\sigma}_p^2(i) = \frac{\sum_j [(\Gamma_j(i) - 1) \sigma_j^2(i) + \Gamma_j(i) \hat{\mu}_j^2(i)] - [\sum_j \Gamma_j(i)] \hat{\mu}_p^2(i)}{[\sum_j \Gamma_j(i)] - 1}. \quad (6.69)$$

Calculation of the Defect Distributions

Now, one last problem needs to be solved, namely calculating the offset, the variance and the mean values of ΔV_{th} for the individual defects. If the measurement noise is assumed to be independent and normally distributed around ΔV_{th} , the levels of combined system \mathcal{Z} are also normally distributed. The resulting Gaussian can be calculated from the mean values μ_i and the variances σ_i^2 of the distributions of the individual defects:

$$\mathcal{Z} = \mathcal{N}\left(\sum_i \hat{\mu}_i, \sum_i \hat{\sigma}_i^2\right) \quad (6.70)$$

This fact can be used to construct a system of equations to get the distribution parameters μ_i and σ_i^2 of each defect by calculating the Cartesian product of the charge vectors $\underline{\mathcal{Q}}_i$ from (6.27) of the N defects.

$$\underline{\mathcal{Q}} = \{\underline{\mathcal{Q}}_0 \times \underline{\mathcal{Q}}_1 \times \dots \times \underline{\mathcal{Q}}_N\} \quad (6.71)$$

Each tuple then represents one row in the coefficient matrix of the system. The right hand side is given by the values from equation (6.67). For one two-state defect 'A' and one three-state defect 'B' with the charge vectors $\underline{\mathcal{Q}}_0 = [0, 1, 2]$ and $\underline{\mathcal{Q}}_1 = [0, 1]$, the Cartesian product reads

$$\underline{\mathcal{Q}} = \{[0\ 0], [0\ 1], [1\ 0], [1\ 1], [2\ 0], [2\ 1]\}. \quad (6.72)$$

The corresponding system of equations thus would be:

$$\begin{bmatrix} \hat{\mu}_0 \\ \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \\ \hat{\mu}_4 \\ \hat{\mu}_5 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \\ 2 & 0 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix} + \mu_{\text{off}} \quad (6.73)$$

The common offset μ_{off} of the combined system is thereby independent of the charged states of the defects 'A' and 'B'. By definition, the mean values $\hat{\mu}_0 \dots \hat{\mu}_N$ also implicitly contain this constant offset because they have been calculated from the absolute levels of the RTN signal (see (6.64)).

It can easily be seen that for systems containing defects with more than two emitting states, the equation system will always be over-determined. Because of that, in general, it can only be solved in a least-squares sense by minimizing e.g. the Euclidean norm $\|\hat{\mu} - \underline{Q}\mu\|^2$.

The noise in principle can also be split into some background noise and two defect-related noise terms if one of the defects has captured a charge. Although that does not make much sense physically, it can help the iterative process as defects with more uncertainty will be more sluggish when changing their states. The equation system is the same as in (6.73) except that all entries in the coefficient matrix unequal zero have to be changed to ones.

$$\begin{bmatrix} \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \\ \hat{\sigma}_3^2 \\ \hat{\sigma}_4^2 \\ \hat{\sigma}_5^2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \sigma_A^2 \\ \sigma_B^2 \end{bmatrix} + \sigma_{\text{off}}^2 \quad (6.74)$$

By definition, solutions smaller than zero should not be allowed, as the background noise $\hat{\sigma}_0$ should be the lower noise-limit. Thus, negative solutions or large differences in the values of σ_A^2 and σ_B^2 indicate problems with either the baseline estimation explained in the next section or an incomplete HMM (i.e. additional states or defects not covered in the model).

6.7.4 Baseline Estimation

Baseline estimation is one of the most crucial issues when trying to train a HMM to RTN data. This is because the selection of the states happens based on probability distributions around absolute levels which must not change over time. This obviously causes errors in the extracted time-constants especially if the long-term drift or pick-up noise exceeds $\sim \Delta V_{\text{th}}/2$ of a single defect, as the signal-level then crosses the point where a state tends to flip. The noise of the signal will then cause artificial transitions and missed transitions, corrupting the statistics especially for states with longer time-constants.

Figure 6.13 illustrates this problem for a simulation of a three-state defect with added random sinusoidal pick-up noise. At the maximum and minimum peaks, transitions are missed and in the transition region, emissions from the fast state are accounted for the slow state. As a result, the time constants of the fast and slow states are almost identical.

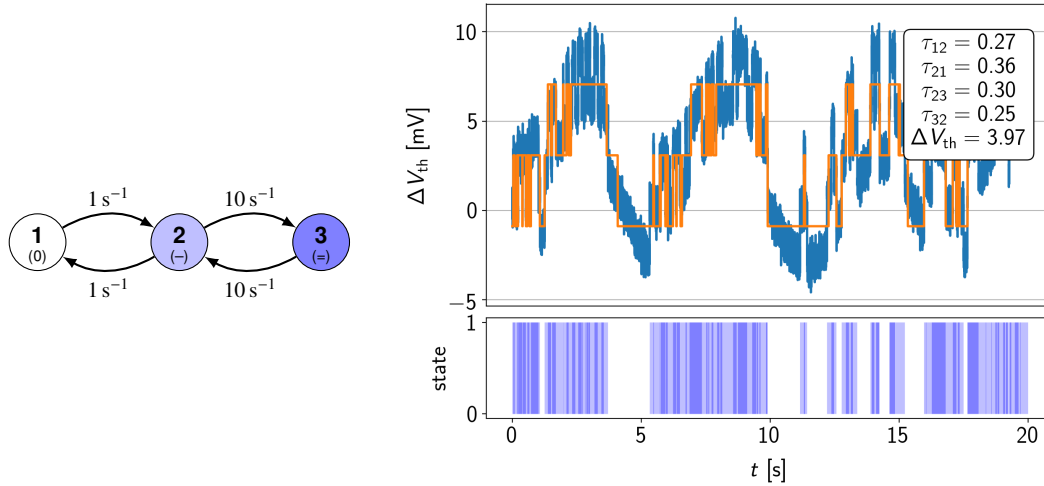


Figure 6.13: Simulated traces (blue) of a three-state defect (left) with random sinusoidal pickup noise. The time-constants are 1 s for the slow state and 0.1 s for the fast state, ΔV_{th} is 3 mV. The orange lines represent the result of the Baum-Welch algorithm without baseline correction. As the Baum-Welch algorithm is sensitive to absolute levels, the noise causes wrong results in terms of the time constants, the offset and ΔV_{th} .

In the following paragraphs, three different methods for data smoothing, namely local regression or scatterplot-smoothing, basis-spline (B-spline) smoothing and asymmetric least-squares smoothing will be discussed. The answer to the question which of the three baseline estimators should be taken for **HMM** training highly depends on the signal quality and the trade-off between quality and computational time. A more detailed discussion about different methods for data smoothing can be found in [176] and [177].

Before going into more detail about the proposed methods, a comparison between the results of the baseline estimation algorithms on real measurement data is given in Figure 6.14. The asymmetrically reweighted penalized least-squares (arPLS) algorithm is specifically designed to find the baseline of the data, whereas the locally weighted scatterplot smoothing (LOWESS) and B-splines algorithm ($m = 3$) delivers least-squares estimates of the local means of the signal. Thus, for the latter two algorithms the **RTN** signal needs to be removed from the data prior to the baseline fitting.

B-splines were the fastest method of the presented ones, however, they should only be used on measurements with low to medium noise, as the inaccuracies introduced at the boundaries and the knot placements highly depend on the signal noise.

On the other hand, **LOWESS** was proven to be the most robust method, delivering decent results for all kind of signals. Despite of being quite slow, just like with the B-spline method, the local mean of the signal is fitted. This makes it necessary to subtract the emissions of the **HMM** from the measurement signal at every iteration to get a raw estimation of the noisy baseline. That can be a problem as the emissions and the baseline estimation become coupled and potentially can lock up in intermediate results. The mechanism is that emissions get fitted into the baseline, which makes them invisible for further **HMM** training.

The **arPLS** method is the only method available to estimate the baseline from the original **RTN** data (i.e., without subtracting the emissions beforehand). This potentially mitigates the lock-up problem between baseline and emissions mentioned before. However, it also imposes a more severe problem: For slow states, the baseline tends to be pulled towards the signal depending on the choice of the smoothing parameter. When used on the raw baseline of the data mentioned

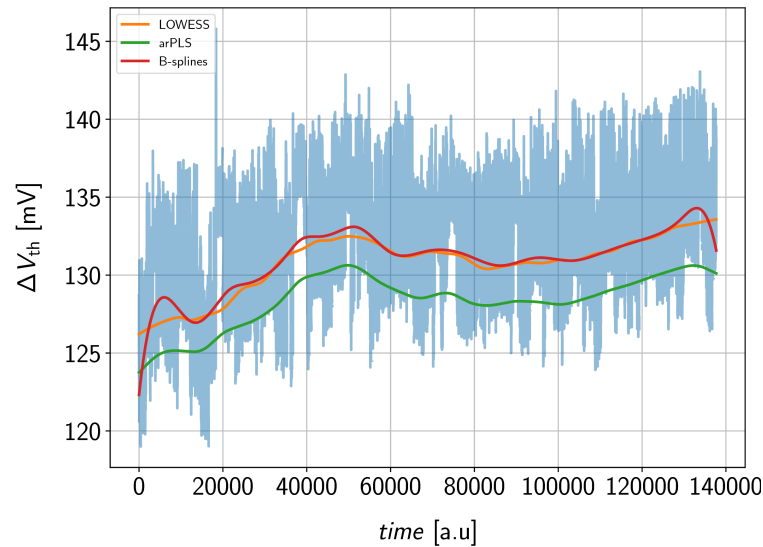


Figure 6.14: A comparison between the discussed baseline estimation algorithms. The arPLS algorithm is specifically designed to find the baseline of the data, whereas the LOWESS and B-splines algorithm ($m = 3$) delivers least-squares estimates of the local means of the signal. Thus, for the latter two algorithms the RTN signal needs to be removed from the data prior to the baseline fitting. The data was taken from real measurements in order to not bias the results by using some analytic drift terms on artificial data.

before, it tends to introduce some numerical drift as it then tries to match the baseline of the residual noise after the removal of the emissions. It was also by far the slowest and most numerically unstable method when used on real data and thus is only of limited value for the examined signals.

Local Regression

A robust algorithm for local regression and scatterplot-smoothing was originally proposed in 1979 by William S. Cleveland and later refined by Cleveland and Devlin [178, 179]. The idea is to fit a low-order polynomial to a point based on a subset of the data at each data point. The polynomial is fitted to the subset by a weighted least-squares algorithm with less weight given to points further away from the fitted data point. Each subset thereby is selected by a nearest-neighbor algorithm. The degree of smoothing is usually selected by a parameter which determines the fraction of data points used to fit the local polynomial.

The degree of the polynomial can be chosen freely, however, high-degree polynomials are numerically unstable, computationally expensive and would tend to overfit the data. For these reasons usually first or second order polynomials are used. For a zero-order polynomial, LOWESS falls back to a running mean. The variant of the LOWESS algorithm used in this work was written by C. Vogel [180] and uses a first-order polynomial. The smoothed value \hat{y}_i for the value y_i is calculated as the sum of weighted projections of all nearest-neighbor points y_j .

$$\hat{y}_i = \sum_j p_{ij} y_j \quad (6.75)$$

Here, p_{ij} is the projection vector calculated from the weighted distances from point i to its neighbors. Note that the distances x_k , x_i and x_j have to be given in units of the window size. The sum

of the weighted distances of all neighboring points k within the window \hat{x} is

$$\hat{x} = \sum_k w_k x_k. \quad (6.76)$$

The projection factors p_{ij} can then be calculated with:

$$p_{ij} = w_j \left[1 + \frac{(x_i - \hat{x})(x_j - \hat{x})}{\sum_k w_k (x_k - \hat{x})^2} \right] \quad (6.77)$$

The weight function w traditionally is the tri-cube weight function:

$$w(x) = \begin{cases} (1 - |x|^3)^3 & \text{for } |x| < 1 \\ 0 & \text{for } |x| \geq 1 \end{cases} \quad (6.78)$$

The biggest advantage of **LOWESS** over many other methods is that it does not require to find a specific analytic function to fit all the data. This makes it a very flexible tool to fit almost all types of curves for which no proper analytic approximations exist. It can also be used to compute the uncertainty of the fit for model prediction and calibration. Another benefit is the simplicity of the model since the user also only has to provide a value for the smoothing parameter and specify the degree of the local polynomial.

One disadvantage of this method is that rather large, densely sampled datasets are needed in order to produce meaningful results. Another drawback is the fact that no analytic regression function is produced so that the results can easily be shared with the community. Maybe the most severe drawback is the computational cost especially for large datasets, as the regression described above in principle needs to be executed for each data point. This problem can be eased by defining a distance for which the previously fitted value is ‘close enough’ and make regressions only for data points lying at least that distance apart. The values in between are then interpolated linearly.

Basis Splines Smoothing

This method works by fitting a sequence of piecewise polynomial functions to a given dataset. The pieces are defined by a sequence of knots, where for a spline of degree m the spline and its first $m - 1$ derivatives need to agree at the knots.

Ordinary splines can be represented by a power series of degree m for k knots [176]

$$S(x) = \sum_{j=0}^m \beta_j x^j + \sum_{j=1}^k \gamma_j (x - \xi_j)_+^m, \quad (6.79)$$

with $(x - \xi_j)_+^m$ being the splines with the coefficients γ_j , $\sum_{j=0}^m \beta_j x^j$ being the values outside the knots and the notation

$$(x - \xi_j)_+ = \begin{cases} (x - \xi_j) & \text{for } x > \xi_j \\ 0 & \text{for } x \leq \xi_j. \end{cases} \quad (6.80)$$

Note that usually additional boundary restrictions have to be introduced, since the power series in (6.79) has $m + 1 + k$ parameters to estimate with only k observations. One possibility would be to fix the values and their derivatives at the boundaries of the dataset (usually to zero).

Although the representation of splines as power series is well suited to understand the fundamentals of spline representation, they are not very well suited for computation [176]. A much

better way is to express them as a linear combination of basis functions called B-splines. Carl de Boor derived a numerically stable algorithm to construct those B-splines in 1976 [181]. The basic principle is that any spline function of the order m on a given sequence of k knots can be represented by a linear combination of a set of B-splines.

$$S_{m,k}(x) = \sum_{j=0}^m \alpha_j B_{j,k}(x) \quad (6.81)$$

If the number of knots equals the number of data points the resulting set of splines can be used for interpolating the data between two points. On the other hand, this method can also be used efficiently to smooth a scatterplot if $k \ll n$. The main problem here is the placement of the knots and the treatment of the boundaries which are often cited as the main drawbacks of spline smoothing. The algorithm used in this work is based on the work of Paul Dierckx [177, 182, 183] and is available as the function *splrep* in the Python library *scipy.interpolate*. It automatically selects the number of knots and their positions based on an empirical smoothing parameter s which is calculated from the weights provided for each data point.

If the sum of the provided weights is a measure for the inverse of the standard deviation of the signal, the smoothing parameter should be in the interval of $s = m \pm \sqrt{2m}$ with m being the number of points. In Figure 6.14 it can be seen that the results of the B-splines closely resemble those of the LOWESS method except on the boundaries of the signal. It has to be noted that this method is very sensitive to the placement of the knots, which makes it less robust compared to LOWESS especially for very noisy signals. On the other hand, it is considerably faster and needs fewer points to deliver decent results.

Asymmetric Least Squares Smoothing

The third and last presented method for baseline estimation is based on a regularized least squares method, where the signal vector y is smoothed by its least squares approximation \hat{y} . The method was specifically developed to estimate the baselines in different kinds of spectroscopy data. The objective function to be minimized is defined as [184]

$$S(\hat{y}) = (y - \hat{y})^T (y - \hat{y}) + \lambda \hat{y}^T D^T D \hat{y} \quad (6.82)$$

with D being the second order difference matrix of the signal.

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \end{bmatrix} \quad (6.83)$$

The first term of (6.82) expresses the fitness to the data, whereas the second term specifies the amount of smoothing. To identify the regions of the peaks, a weight vector is introduced which can be set to zero at the peak regions if they are known beforehand. This however requires peak finding algorithms which can be difficult to implement especially for noisy data. The objective

function with the diagonalized weight vector W then becomes:

$$S(\hat{y}) = (y - \hat{y})^T W (y - \hat{y}) + \lambda \hat{y}^T D^T D \hat{y} \quad (6.84)$$

The solution of the minimization problem in (6.84) is given by [184]:

$$z = (W + \lambda D^T D)^{-1} y \quad (6.85)$$

The advantage of this method is that a proper construction of the weight vector w allows skipping the peak finding algorithm. At first, this was done Eilers and his peers [185, 186], who simply specified an asymmetry parameter p defined as:

$$w_i = \begin{cases} p & \text{for } y > \hat{y} \\ 1 - p & \text{for } y \leq \hat{y} \end{cases} \quad (6.86)$$

Later, Zhang pointed out that the parameters λ and p need optimization to get adequate results and also that the weights in the baseline region are all the same instead of being set according to the difference of signal and baseline. This was the main motivation to set the weights iteratively based on the current estimate [187].

$$w_i = \begin{cases} 0 & \text{for } y \geq \hat{y} \\ \exp\left(\frac{t(y - \hat{y}_i)}{|d|}\right) & \text{for } y < \hat{y} \end{cases} \quad (6.87)$$

The iterative procedure stops if the norm of the vector d containing the negative elements of the residuals $y - \hat{y}$ is smaller than a certain fraction of the norm of y . The problem with these methods is that the values below the baseline in the non-peak regions are considerably stronger weighted compared to those above. This results in an underestimation of the baseline and thus an overestimation of the peaks in the iterative procedure.

To correct for this error, the arPLS method was suggested in [184] which uses a more symmetric way to define the weights using a generalized logistic function defined by the mean value and the standard deviation of d .

$$g(x) = \frac{1}{1 + e^{2(d - (2\sigma_d - \bar{d}))/\sigma_d)}} \quad (6.88)$$

The logistics function gives nearly the same weights for all values below the baseline estimate up to a distance σ_d from the mean. At a distance $4\sigma_d$ from the mean, the weights are already practically zero. Figure 6.15 depicts the logistics function in units of \bar{d} and σ_d . From the three investigated weight functions, on average the logistic function delivered the most reliable results across the measurement data and thus should be used for baseline estimation.

The main benefit of arPLS compared to LOWESS and B-splines is that it can be used on the original data since it is specifically designed to follow the baseline of the signal. The other two methods settle somewhere around the local mean of the signal. Thus they cannot be used on the measurement signal as they are.

The drawbacks of arPLS are the size of the sparse matrix system to solve which is equivalent to the number of data points. Also, the parameter λ for smoothing scales with the signal length and was found to be numerically unstable above a certain threshold. Whether this is due to the accuracy of the used solver or the numeric resolution of the datatypes needs to be investigated. Another drawback is the computation time and the number of iterations needed to reach a stable result. While a single iteration was approximately two times slower than a LOWESS iteration, the algorithm needs at least five iterations to converge, assuming a stable solution. This makes

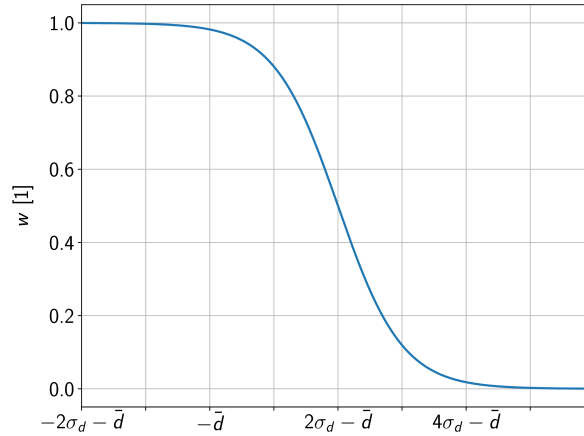


Figure 6.15: The logistic function defined in (6.88). It gives nearly the same weights for all values below the baseline estimate up to a distance σ_d from the mean. At a distance $4\sigma_d$ from the mean, the weights are already practically zero.

it ten times slower compared to LOWESS which only needs one to two iterations if the data is sufficiently dense.

6.7.5 Benchmarks

On the following pages, the implemented HMM library will be tested with sampled data coming from a simulated system of two defects with known properties (for details on the HMM implementation see Appendix A). The simulated data is varied by different means and given to an independent system with the same structure but different initial conditions for training. It is thus possible to check for the robustness of the implementation by comparing the training results to the known system, which would not be possible when using real measurement data.

Note that up to this point, no assumptions regarding the physics or structure of the defects except the Markov property were made. This is quite convenient, since choosing a physical model beforehand could lead to biased results with respect to the time constants.

For the sample system, one four-state defect ‘A’ containing one thermal state is combined with a two-state defect ‘B’. The Markov chain of both defects as well as their RTN signal is given in Figure 6.16.

The first thing to investigate is how many transitions are actually necessary to obtain a reliable result. In order to do that, the confidence intervals of the exponential distribution over the number of transitions are calculated. Given the maximum likelihood estimate of the characteristic time-constant $\hat{\tau}$ and the number of expected transitions N , the $(1 - \alpha)$ confidence intervals are given by [188]:

$$\frac{2N\hat{\tau}}{\chi_{2N}^2(1 - \frac{\alpha}{2})} < \tau < \frac{2N\hat{\tau}}{\chi_{2N}^2(\frac{\alpha}{2})} \quad (6.89)$$

where $\chi_{2N}^2(p)$ is the p -percentile of the chi-squared distribution with $2N$ degrees of freedom. The results in Figure 6.17 show that about 19 transitions are needed within a $1 - \sigma$ confidence interval in order to obtain an maximum error of 30%. For a ninety-percent confidence interval, this value already goes up to 46 (dashed lines). For a maximum error of ten percent, the minimum number of transitions are already 120 for the one-sigma interval and 315 for the ninety-percent interval (solid lines).

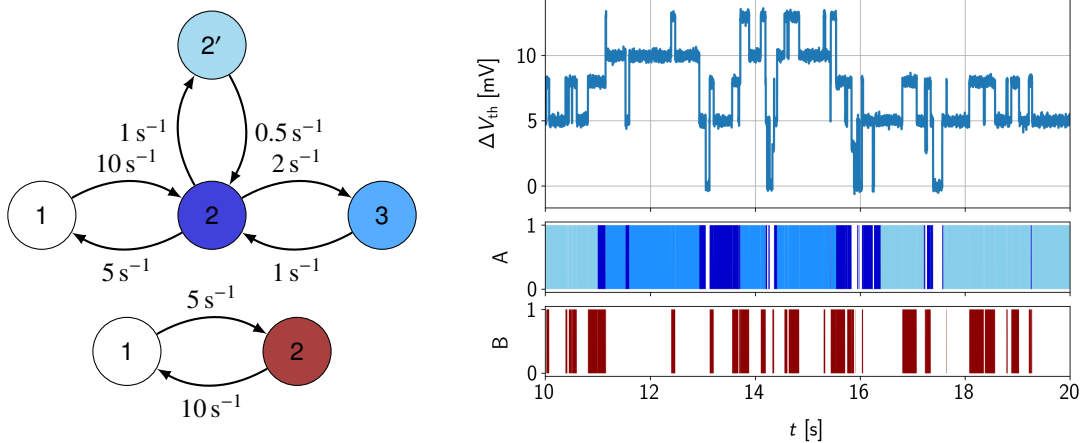


Figure 6.16: Left: The Markov chains of a four-state defect with thermal state and a two-state defect used for the model benchmarks. The charge state is neutral if the defect is in state 1, negative if in state 2' or 2 and double negative in state 3. The step heights are 5 mV and 3 mV.

Right: Simulated emissions of the two defects combined. The RTN signal already looks quite complicated although only two defects contribute. With conventional methods, the thermal state could not be determined.

These numbers can also be used to estimate measurement times for simple RTN signals. Note that for multi-state defects, the number of observed transitions within a time interval scales with the first-passage times of the defect's state (imagine a three-state defect with a slow state on top of a fast one).

The red area in [Figure 6.17](#) shows the Normal approximation of the one-sigma confidence interval of the exponential distribution. It can be seen that in this case, the confidence intervals only diverge significantly from the chi-squared distribution if there are less than five transitions. However, for the 90 % confidence intervals, the number of transitions to obtain a similar well-approximated result is at least one order of magnitude above that value, meaning that the quality of the approximation largely depends on the chosen confidence interval. The HMM library thus routinely calculates the 90 % confidence intervals using equation (6.89).

For the following simulation study, the initial parameters for ΔV_{th} , τ and the simulated measurement signal were chosen randomly with 100 different seeds. The first thing checked is the sensitivity of the HMM library to the initial conditions in terms of the characteristic times and ΔV_{th} . The initial and final time constants of each defect were normalized and binned into a single histogram as shown at the upper part of [Figure 6.18](#). For both defects, the model most of the time converges towards the proper solutions, even with a quite broad distribution of the initial values.

The lower part of [Figure 6.18](#) gives the sensitivity to the step heights of the defects. As with the time constants, the sensitivity to variations of the initial conditions is very low. It should be noted that as mentioned in [Section 6.7](#), the Baum-Welch algorithm converges towards a local maximum. This can be seen especially in the scatterplots for the step heights where distinct secondary peaks in the histograms show up. They are, however, quite small compared to the maximum and thus are not considered as a serious problem.

Another issue to be mentioned is that the initial values for ΔV_{th} were constrained in such a way that the initial value of defect 'A' with a nominal value of 5 mV had to have a larger value than defect 'B' (3 mV) and defect 'B' had to be smaller than the nominal value of defect 'A'. This helps to maintain the structure of the system and helps to find the proper local minimum in the parameter space. On a side note, it has to be mentioned that some of the seeds delivered

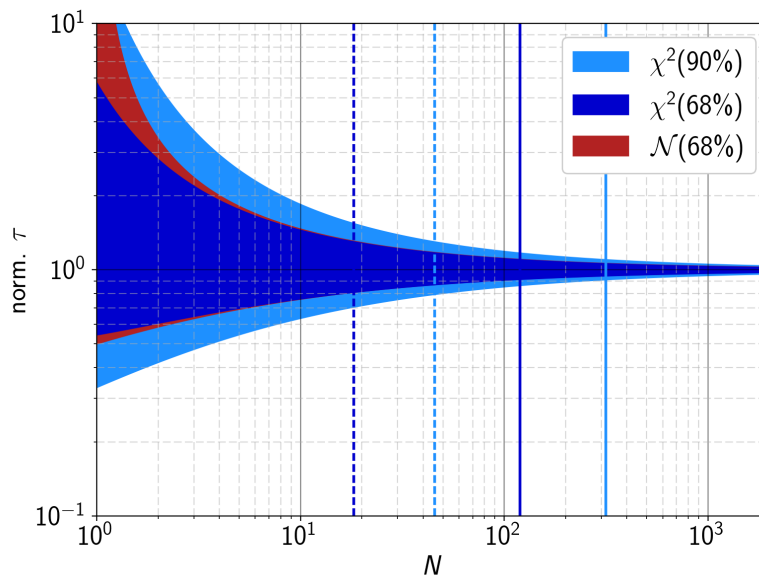


Figure 6.17: The confidence intervals of the exponential distribution over the number of transitions. The dashed lines mark the number of transitions where the error is within 30%, for the solid lines the error is within 10%. The shaded red area is the one-sigma confidence interval in a Normal approximation.

impossible solutions having a log-probability of $P = -\infty$. This is a well-known issue for the maximum a-posteriori (MAP) path decoding used in this work. These solutions were filtered from the results before post-processing of the data.

The next property investigated was the sensitivity of the training results regarding (simulated) external measurement noise. For that purpose, Gaussian noise with different values of σ was added to the measurement signal. Despite some of the seeds converging to different solutions in [Figure 6.19](#) (faint traces), it can be seen that the weighted average (with the weights being) of the time constants are very close to their appropriate values up to about a signal to noise ratio of about 0.75, quickly diverging for values above that.

This can easily be explained by the fact that considering an interval of about $\pm\sigma$, the noise level at that point on average becomes larger than the step height. The point at which the solutions begin to degrade could probably still be pushed a bit further by broadening the statistics (i.e. expanding the cumulated measurement time). Considering that the simulations were sampled at 1 kHz for 1 ks (i.e., 10^6 samples), for a maximum time constant of 2 s (see [Figure 6.16](#)), the effect is thought to be quite small.

The dependence of the solutions over measurement time is given in [Figure 6.20](#) for $\sigma = 0.5$ mV and a sampling frequency of 1 kHz. Except for measurement times below 10 s, the training results are close to their real values on average. The uncertainty of the solutions naturally increases at lower times as there are fewer transitions contributing to the solution within one measurement window. Above ~ 10 s or a few transitions for the slowest state, the statistics for that system seems to be sufficient (see also [Figure 6.17](#)).

It should be said that the intuitive feeling that the minimum number of observed transitions for a given system is the measurement time over the slowest capture or emission time is only valid for systems consisting of two-state defects. This is because for more complicated cases, the first passage times of the defect's Markov chains determine the observed number of transitions. As an example, a three-state defect with a $k_{23} \gg k_{21}$ will hardly be able to reach state 3 at all.

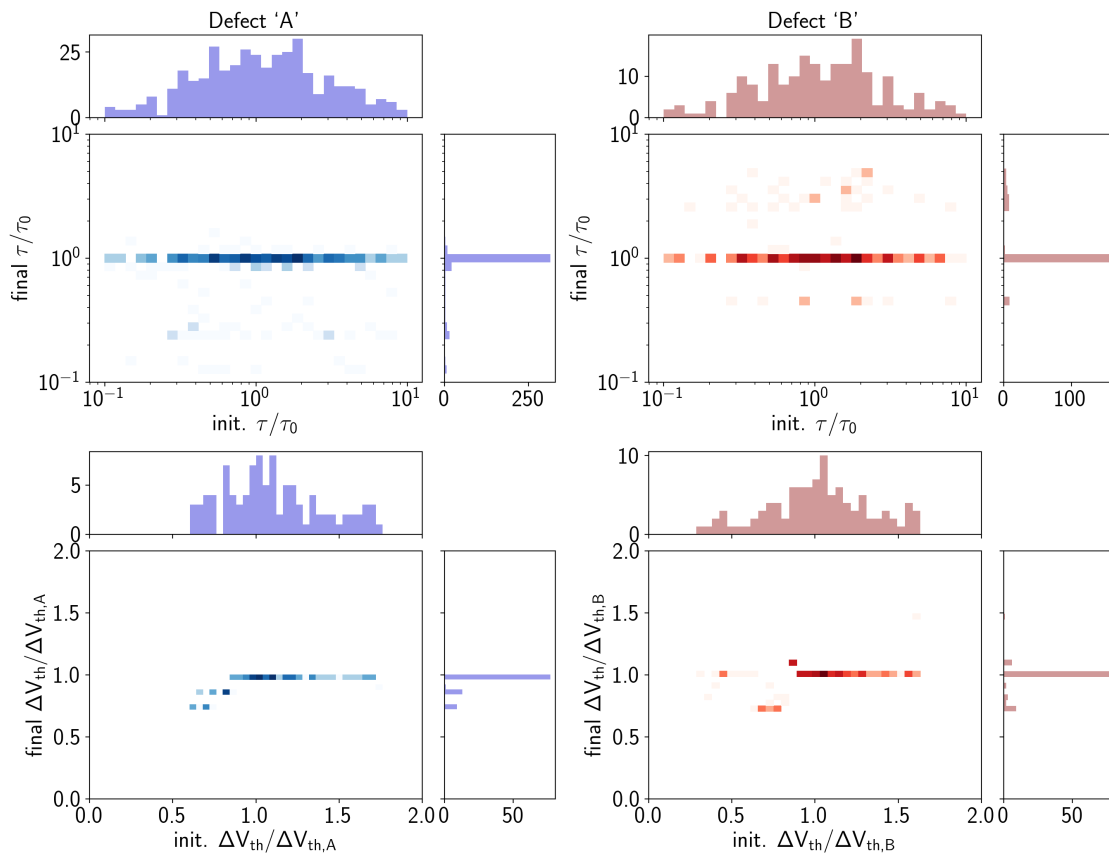


Figure 6.18: The sensitivity of the model regarding the initial conditions. On the top row, the combined histograms for all of the capture and emission times for both defects can be seen. The bottom row shows the sensitivity to variations in the initial ΔV_{th} . Although quite broad normal distributions were chosen in both cases, the solutions converged towards the correct values nearly all of the time.

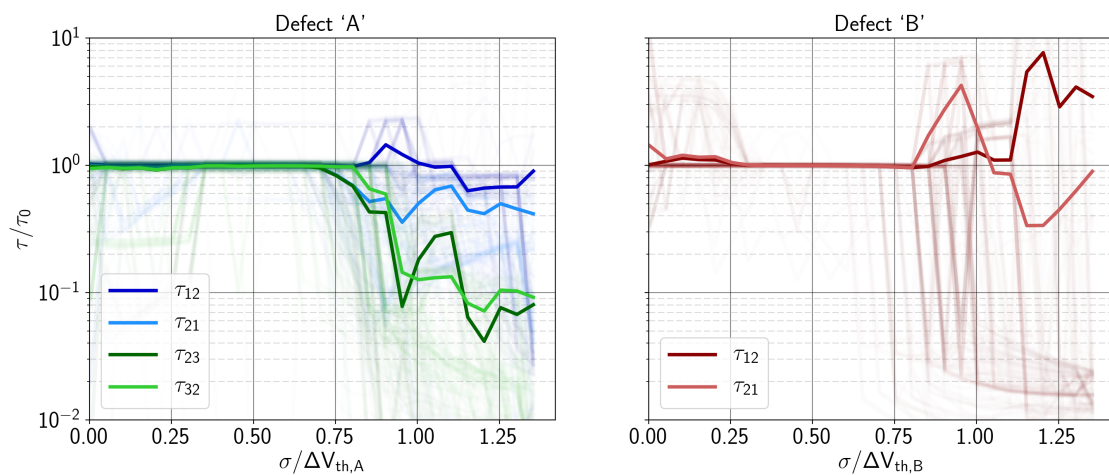


Figure 6.19: The dependence of the normalized time constants on Gaussian noise. The results are close to their appropriate values up to a signal to noise ratio of about 0.75. This can be explained by the fact that around that value, the noise becomes larger than the step height.

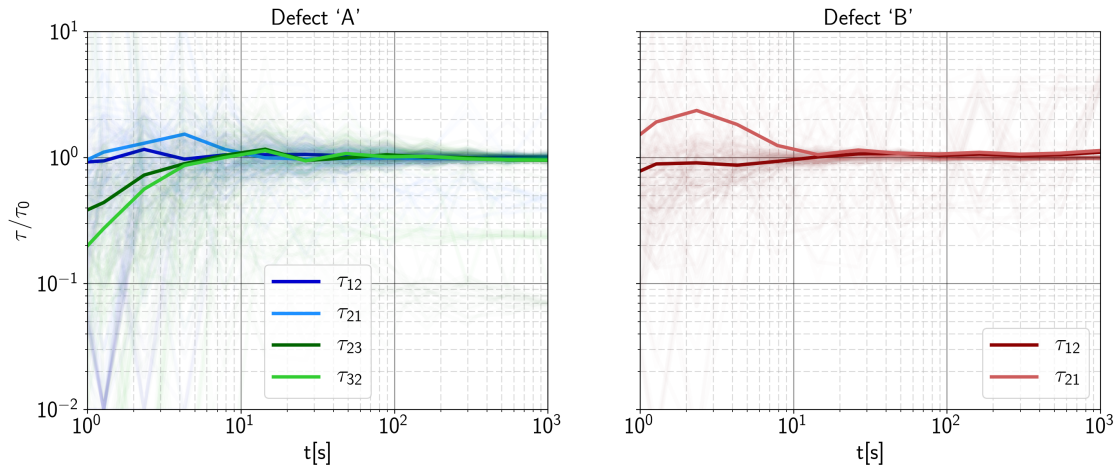


Figure 6.20: The extracted time constants over the measurement time. The uncertainty of the solution increases with decreasing time as less transitions for each state contribute. The minimal time for a reliable result largely depends on the first passage times of the investigated defects. For good results, at least ten to twenty transitions should be observed in the cumulated measurements (see [Figure 6.17](#)).

Also, the number of different simulations help to effectively increase the statistics in a way that the cumulated time increases with every different seed. Although each training was done independently on one simulated trace, the median values of the results naturally follow the cumulated times.

For real devices, another complication arises due to the broad distributions of capture and emission times observed in GaN devices [6, 44]. Faster sampling rates or longer measurement times will always result in additional defect transitions, which had not been within the measurement window before. With other words, the faster the sampling, the more of the fast defects can be seen and the longer the measurement the more of the slow defects can be seen.

The next test was done by comparing the results for different sampling frequencies. The measurement time was chosen to be 1 ks with a Gaussian noise of 0.5 mV. Surprisingly, the results in [Figure 6.21](#) were consistent down to 100 Hz, which already is equivalent to the emission time of the fastest state. This finding can maybe be explained by the fact that emissions faster than the sampling frequency will be partly truncated at the sampling time and partly be sampled out of the signal (depending on the point of sampling). The lower part of an exponential distribution centered around the sampling frequency thus will be seen as an additional peak at the minimum time, not changing the overall expectation value of the distribution very much.

On the other hand, the induced distortions and aliasing effects could potentially obfuscate the real structure of the Markov chain as well as the training results. It is thus recommended to maintain a sufficient Nyquist margin when downsampling data or setting up measurements.

The last benchmark was to test the robustness of the baseline fitting algorithm of the **HMM** library. For that purpose, random walks with different amplitudes per step were generated in order to obtain as many different long-term drift conditions as possible. To further complicate the problem, an additional 50 Hz signal with Gaussian spectrum was added to the drift signal. The Gaussian noise of the samples was again 0.5 mV.

The main motivation of using random walks is to eliminate biased results based on the analytic form of the drift signal (i.e., by using combinations of polynomial or exponential terms for simulating drift). As a measure of severity of the drift the root mean square (RMS) of the baseline

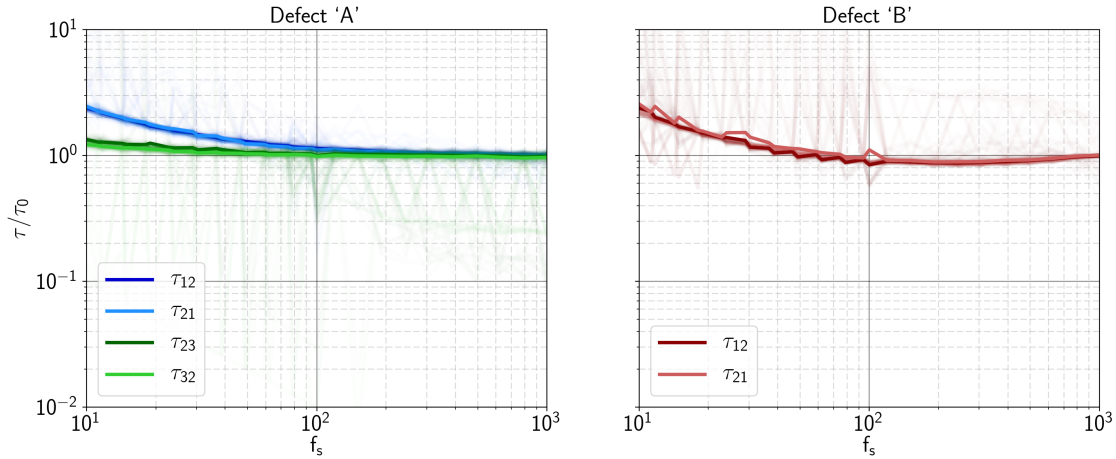


Figure 6.21: The extracted time constants over the sampling frequency. The results are surprisingly consistent even for sampling frequencies equal to the fastest state. This can be explained by the fact that all emissions faster than the sampling frequency will either be sampled out or truncated at the sampling limit randomly and thus not changing the expectation value too much.

was calculated, despite not being ideal as for example also the number of turning points or the maximum gradient of the signal could influence baseline extraction.

In [Figure 6.22](#), the cumulated 2D-histogram of the normalized times versus the RMS values of the drift signal are shown. The solid line was calculated with the LOWESS method and thus gives the center of mass across the density plot. The density plot is printed on a logarithmic color scale in order to show the spreading of the results with increasing drift which was almost not visible on a linear scale.

The center of mass resembles the real values closely up to a value of around $\sigma_{\text{RMS}} = 10 \text{ mV}$. After that, the solutions start to degrade quickly. A partial explanation of that behavior would be that the tails of the histogram of drift values is reached, which leads to more statistical uncertainty due to the small sample size. On the other hand, the initial guess of the forward-backward algorithm (which is done with a constant offset) on average covers fewer emissions. The uncovered emissions are then used wrongly to calculate the baseline signal which in turn forces the HMM into wrong solutions.

A definitive answer to the question whether one of those mechanisms is dominant is hard to estimate. In any case, given a fixed set of measurements, the only way to obtain robust results is to train the model with as many different initial conditions as possible.

6.8 Conclusions

In this chapter, the theoretical foundation to extract the characteristic time constants from stochastic single-defect emissions, especially from RTN signals was laid out. After covering the basics of Markov processes and introducing the Markov chains of different types of defects in [Sections 6.1 to 6.3](#), in [Section 6.4](#) a way to combine different defects into a combined system was developed. [Section 6.5](#) covers the special case of coupled defects which will be used in the next chapter to estimate the electrostatic coupling between neighboring defects.

In [Section 6.6](#), a new way to extract the characteristic time constants based on spectral maps was presented and compared to traditional methods like the histogram method and the lag plot

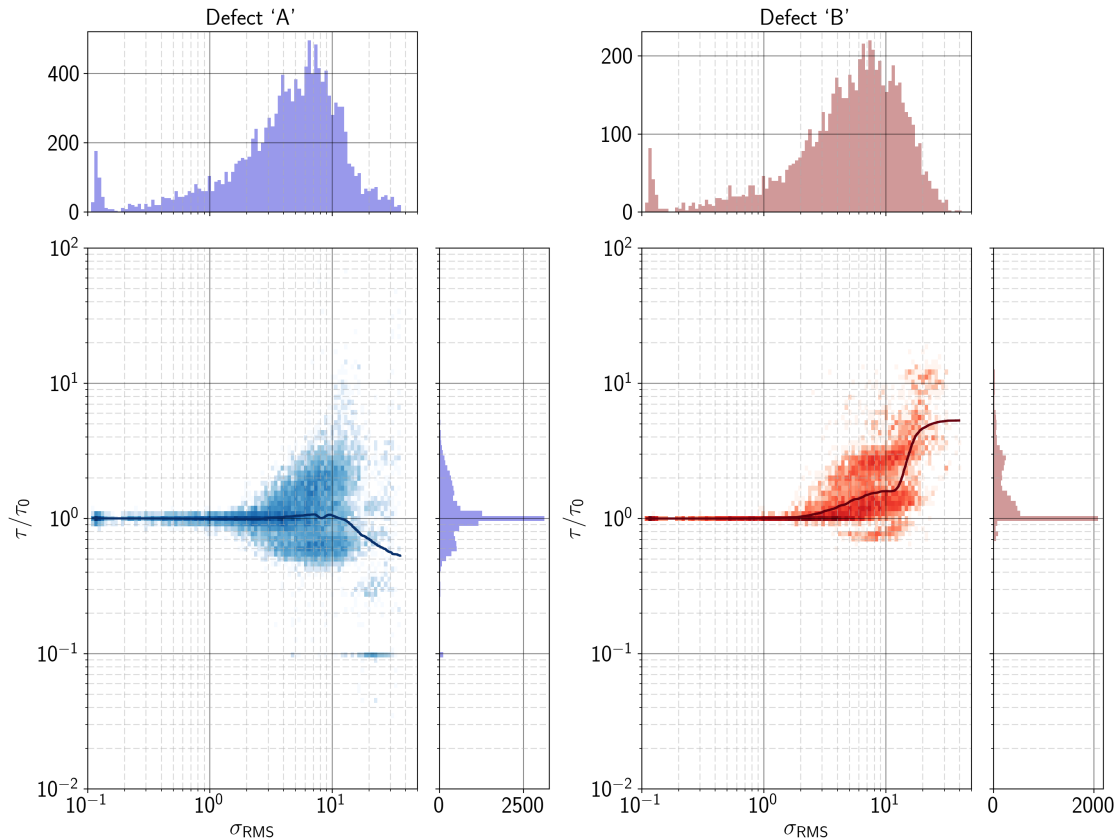


Figure 6.22: The normalized time constants in relation to different baselines. The baselines were generated with random walks of different amplitudes. Up to a RMS value of about 8 mV, the extracted times closely resemble the real values. The colormap used for the density plots is on a logarithmic scale to emphasize the spreading of the results at higher drift values. The accumulation of failed time constants at $\tau = 0.1\tau_0$ can be understood by the limits imposed by the measurement time and the sampling frequency for the different defects.

method. The limitations of these methods, especially their inability to detect thermal states producing anomalous RTN led to the development of a HMM library able to handle systems of multiple, arbitrarily shaped defects (see also Appendix A).

The theoretical foundation for the training of the HMM including the constraints imposed by the special case of systems of defects producing single-charge emissions is given in the preceding Section 6.7. Despite the well-known Baum-Welch algorithm, this section discusses the treatment of multiple defects, multiple sequences and the extraction of ΔV_{th} of the defects and the measurement noise of the combined system. One very important problem which is often ignored for HMM, namely the long-term drift of measurement signals is treated in Section 6.7.4. All of these constraints make the presented HMM library a versatile and robust tool to extract defect properties from complex RTN signals.

The presented HMM library can be used to either simulate RTN signals for different systems or to extract their properties by training the model with a set of measurement sequences. The former case was used on many occasions within this chapter to reveal expected RTN signals produced by certain types of defects or to point out systematic deficiencies of different extraction modes. The latter case was used in the previous section to test the robustness of the library to different influences. In the next chapter the HMM library will be used to investigate the defect structure and to extract the characteristic time constants of defects producing three-level anomalous RTN in nanoscale GaN/AlGaN MIS-HEMTs.

Chapter 7

Charge Trapping Effects in GaN/AlGaN HEMTs

In the first part of this chapter, charge trapping at the surface of GaN/AlGaN HEMTs at different forward bias stress conditions is investigated. It will be shown that in order to correctly model the observed charge trapping, the electrostatic feedback of the charges has to be taken into account by using self-consistent transient simulations. In particular the stress history of the device and local potential changes lead to different dynamics of each individual defect.

In the second part, a single defect characterization on complex three-level RTN measured at cryogenic temperatures on nanoscale GaN/AlGaN fin-MIS-HEMTs will be carried out. First, the energy levels and vertical positions of the defects are extracted using a simple two-state model together with TCAD simulations and the spectral method explained in [Section 6.6.3](#). In the following, the hypothesis of having two coupled two-state defects is questioned by estimating the required coupling factors and comparing them to values acquired by different methods. Based on these findings, two different defect candidates are chosen and investigated using HMM training as put forward in [Section 6.7](#).

7.1 Threshold Voltage Drift in GaN/AlGaN MIS-HEMTs

Threshold voltage drift, especially at forward gate bias, is one of the major reliability issues for GaN/AlGaN MIS-HEMTs [[75](#), [94](#), [189](#), [190](#)]. Measurement studies revealed broad distributions of capture and emission times for the defects present at the interface between AlGaN barrier and insulator. In addition, several studies point out that transport through the barrier layer can have a significant influence on the recovery dynamics of these devices [[AGJ1](#), [93](#)]. Additionally, the huge surface defect densities of about $1 \times 10^{13} \text{ cm}^{-2}$ present at the insulator interface pose a challenge to their experimental characterization [[75](#), [AGJ7](#)]. The observed threshold voltage shifts can thereby even approach the theoretical limit of charge storage in a double capacitor model for different insulating materials [[66](#)]. Another consequence of the large defect densities together with the depleted barrier region at recovery conditions is that the influence of charge feedback on the surface potential is much larger when compared to silicon technologies [[191](#)].

The following sections are dedicated to a simulation study on the forward gate bias V_{th} drift using the NMP charge trapping model with a special focus on the impact of charge feedback on the recovery kinetics of the device [[AGC1](#)].

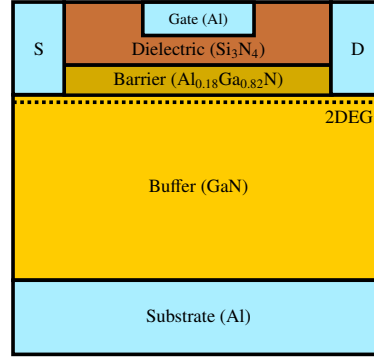


Figure 7.1: The schematic geometry of the investigated devices. A 10 nm thick $\text{Al}_{0.18}\text{Ga}_{0.82}\text{N}$ barrier is placed on top of a $1\ \mu\text{m}$ thick GaN layer. The thickness of the SiN gate dielectric is 25 nm with a gate length of $1\ \mu\text{m}$. To obtain reasonable simulation times, the access regions as well as the source and drain lengths are scaled to 500 nm each. All metal contacts are formed using aluminum (from [AGC1]).

7.1.1 Electrostatic Device Simulation

The investigated devices were formed on a $1\ \mu\text{m}$ GaN buffer grown on a silicon substrate with a 10 nm thick $\text{Al}_{0.18}\text{Ga}_{0.82}\text{N}$ barrier on top. The gate length was $1\ \mu\text{m}$, insulated from the barrier with a 25 nm thick SiN layer. In order to save computation time, the source and drain regions as well as the access regions were scaled to 500 nm. The substrate material was replaced by an aluminum contact because in the investigated bias regions, the bulk region is effectively screened by the electrons in the 2DEG. The simplified device geometry used in the simulations can be seen in Figure 7.1. All simulations were conducted using the device simulator Minimos-NT [21].

Piezoelectric Charges

The net polarization charges at the interfaces are calculated with the data provided by Ambacher *et al.* [192]. The polarization charges at the interface to the substrate are considered to be fully compensated within the highly-defective transition region between the silicon substrate and the GaN buffer. The net charges at the barrier-buffer interface are simply given by the difference between the spontaneous polarization in the buffer and the spontaneous and piezoelectric polarization in the barrier layer. The interface between barrier and dielectric is somewhat more complicated since according to the surface donor model, the negative polarization charges at that interface are compensated partly or fully by positive defects. Additional information on the polarization charges in GaN based devices and the formation of the 2DEG can be found in Sections 2.2.4 and 2.3.1.

According to [192], the polarization charges can be calculated as:

$$\begin{aligned} P_{\text{GaN}}^{\text{SP}} &= -0.034\ \text{C m}^{-2} \\ P_{\text{Al}_x\text{Ga}_{1-x}\text{N}}^{\text{SP}} &= (-0.09x - 0.034(1-x) + 0.021x(1-x))\ \text{C m}^{-2} \\ P_{\text{Al}_x\text{Ga}_{1-x}\text{N}}^{\text{PZ}} &= (-0.0525x + 0.0282x(1-x))\ \text{C m}^{-2} \end{aligned} \quad (7.1)$$

The sheet carrier density at the channel interface is simply the difference between the polarizations between barrier and buffer layer.

$$\sigma_{\text{GaN/Al}_x\text{Ga}_{1-x}\text{N}} = \frac{P_{\text{Al}_x\text{Ga}_{1-x}\text{N}}^{\text{SP}} + P_{\text{Al}_x\text{Ga}_{1-x}\text{N}}^{\text{PZ}} - P_{\text{GaN}}^{\text{SP}}}{q_0} = 1.1 \times 10^{13}\ \text{cm}^{-2} \quad (7.2)$$

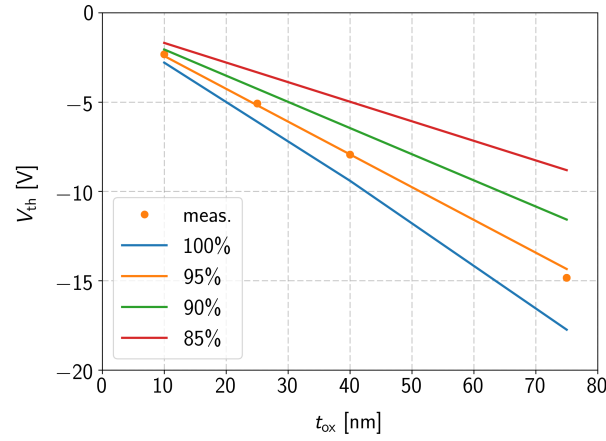


Figure 7.2: The positive charges at the dielectric interface define the surface potential of the device which in turn is responsible for the threshold voltage. The charges can be estimated by comparing the threshold voltages at different dielectric thicknesses. In the picture, V_{th} is shown for different amounts of (negative) polarization charges being compensated by positive surface donors. The best agreement is obtained for 95 % of the piezoelectric charges being compensated (reprinted from [AGC1]).

Up to date there is no solid experimental evidence on the exact defect density of the surface donors [66, AGJ7]. Nonetheless, the sheet carrier density serves as a lower limit to the number of positive charges present in the device due to charge neutrality requirements. Theoretically, all of the charge caused by spontaneous and piezoelectric polarization in the $\text{Al}_{0.18}\text{Ga}_{0.82}\text{N}$ layer needs to be compensated by defects at the surface, which is approximately three times as large as the sheet carrier density in the channel.

$$\sigma_{\text{Al}_x\text{Ga}_{1-x}\text{N}} = \frac{P_{\text{Al}_x\text{Ga}_{1-x}\text{N}}^{\text{sp}} + P_{\text{Al}_x\text{Ga}_{1-x}\text{N}}^{\text{pz}}}{q_0} = -3.22 \times 10^{13} \text{ cm}^{-2} \quad (7.3)$$

This could also serve as a possible explanation for defect densities larger than the sheet carrier density of the device ($> 2 \times 10^{13} \text{ cm}^{-2}$) observed at the insulator interface of different GaN based MIS-HEMTs [66]. Independently of the charge captured by defects at this interface, the number of positive compensation charges and thus the net amount of polarization charges can also be estimated by comparing the threshold voltages of devices from the same technology but with different dielectric thicknesses. In thermal equilibrium, the barrier layer of the MIS-HEMT is depleted by the polarization field and thus effectively acts as a capacitor.

Because of the fact that only the dielectric thickness changes on top of the same layers, the net positive charge at the dielectric interface defines the threshold voltage of the device. Under the assumption of the barrier layer acting as an ideal capacitor, the voltage divider between the gate dielectric and the barrier can be used to calculate these charges (see Figure 7.2). In the picture, V_{th} is shown for different fractions of the theoretical polarization charges being compensated by positive surface donors. It can be seen that the best agreement is obtained for 95 % of the piezoelectric charges being compensated.

Bulk Doping and Barrier Transport

In the simulations, deep donor traps were placed into the buffer and barrier layers at an energy level of about 0.6 eV below the GaN conduction band with concentrations of $1 \times 10^{16} \text{ cm}^{-3}$ and $5 \times 10^{16} \text{ cm}^{-3}$ respectively. The energy levels were chosen according to the most common defect

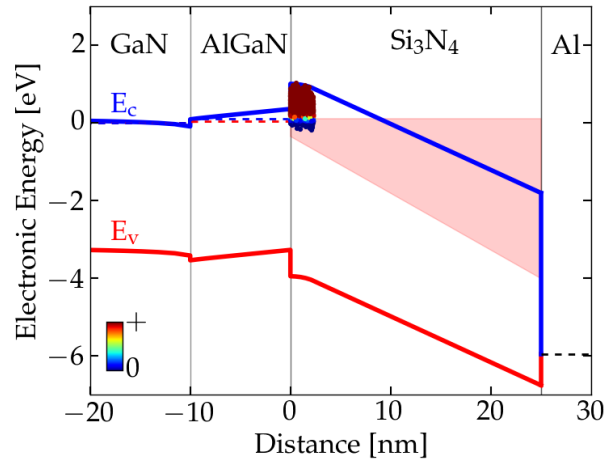


Figure 7.3: The band energy profile along a cut through the gate shows the vertical placement of the defects in the dielectric along with the active energy area (shaded area) for a gate stress bias of 4 V and a recovery bias of 0 V. The active energy area is defined as the region between the Fermi levels of gate and barrier interface for a certain stress and recovery bias. Within this region NMP defects are able to capture and emit electrons during stress and recovery respectively. The charge state of each NMP defect is depicted by its color, red for positively charged defects and blue for neutral ones. The results depict the thermal equilibrium of the device, not taking into account the transient effects discussed in the following sections (from [AGC1]).

levels present in GaN, see Figures 3.1 and 3.2. To describe the transport of carriers from the channel to the barrier, a thermionic field emission model is employed [193]. The band alignments between the buffer, barrier and SiN layers as well as the workfunction difference of $E_w = 1.95$ V for the aluminum gate are set using the data provided by [194].

As the majority of electrically active defects are thought to be at the dielectric interface of the device, the NMP traps are placed evenly within the first 2 nm of the SiN layer. Their concentration is calculated based upon the results from Figure 7.2. Figure 7.3 shows the distribution of defects in thermal equilibrium along a vertical cut through the gate of the device together with the band energy profiles. The shaded area depicts the active energy area between a gate stress bias of 4 V and a gate recovery bias of 0 V. In general, the active energy area is defined as the region between the Fermi levels of gate and barrier interface for a certain stress and recovery bias. Within this region NMP defects are able to capture and emit electrons during stress and recovery respectively. The charge state of each NMP defect is depicted by its color, red for positively charged defects and blue for neutral ones. As the results depict the thermal equilibrium of the device, they do not take into account the transient effects discussed in the following sections.

7.1.2 Forward Bias Threshold Voltage Drift

In order to calibrate the NMP defect model as well as possible, the transient simulations need to follow the experimental procedure provided in [75, 190] as close as possible. The data was recorded by an eMSM setup where the measurements are conducted for a variety of different stress times and forward bias voltages. A discussion of the eMSM measurement procedure together with typical bias conditions can be found in Section 4.1.3.

Since barrier related effects like the transport of carriers to the dielectric interface and its contribution to the measured time constants are still barely understood, the study was focused on stress voltages above the spillover region. In this region, a second electron channel is formed at the barrier/dielectric interface due to the high forward gate bias. This rules out any barrier transport

effects for charge capture because the supply of trapped carriers will preferentially be provided by this second channel.

In the case of recovery, the situation can be different, since the captured electrons need to be transported through the barrier after emission from the defect. Reverse leakage studies on Schottky devices often suggest Fowler-Nordheim tunneling [141] for low temperatures and high reverse bias conditions and Frenkel-Poole like trap-assisted tunneling [152] for high temperatures and low reverse bias conditions respectively [41, 42]. A study on the onset of drift recovery at different voltages and temperatures extracted an apparent activation energy of 0.65 eV for barrier transport, slightly decreasing for lower positive voltages (i.e. high barrier fields)[AGJ1]. That value approximately matches the surface potential of the fully processed SiN devices at thermal equilibrium extracted from the simulations. These findings support the hypothesis of a transport mechanism via the AlGa_N conduction band, most likely a thermal emission or field-enhanced thermal emission of carriers from the defects to the conduction band. Once in the conduction band, the carriers are transported towards the channel through the barrier field. Both effects are covered by the TCAD simulations and therefore should not have a significant influence on the extracted defect parameters.

The observed recovery behavior in these devices can be modeled with two sets of NMP defects with normally-distributed parameters, one for the more permanent part of the degradation and one for the more recoverable part. With the same set of defects, excellent agreement is reached for eight orders of magnitude in stress time and for three different stress voltages. The measurements together with the simulated recovery traces are shown in Figure 7.4. Even for the worst case stress conditions (10 V/10 s), the largest observed degradation only corresponds to a trap occupancy of approximately 25 %. Note that the used defect density in the simulations is still a lower bound to the real situation as only the amount of donor-like traps calculated from Figure 7.2 are considered. Defects with acceptor-like behavior and possible amphoteric states of the donor-like defects within the bandgap are not covered by the simulations either.

Significant deviations from the measurements are only found for large stress times and short recovery times up to about 1 ms. This can most likely be explained by a lack of fast NMP defects being within the active energy area at the specific stress voltage in the simulations. This can be either attributed to a lack of variance in defect distributions or to the fact that the simulations were carried out using the band edge approximation, where the NMP model only considers the conduction band minimum and valence band maximum instead of the continuum of states in the respective bands of the barrier. On top of that, other transport mechanisms through the barrier not covered by the thermionic emission model like trap-assisted tunneling or Frenkel-Poole conduction could potentially lead to faster effective emission times of the bound electrons especially at elevated stress times.

7.1.3 Defect Feedback Mechanisms

Due to the large defect densities observed in all BTI experiments in GaN/AlGa_N MIS-HEMTs, the question arises to what extent the response of these trapped charges has an influence on the recovery behavior of the device. This is of special interest if the nature of the surface donors should be revealed in the future. Various feedback mechanisms like local interactions or temporal changes in the surface potential due to the electrostatic feedback of the defects can potentially obfuscate the extracted defect parameters if not taken properly into account.

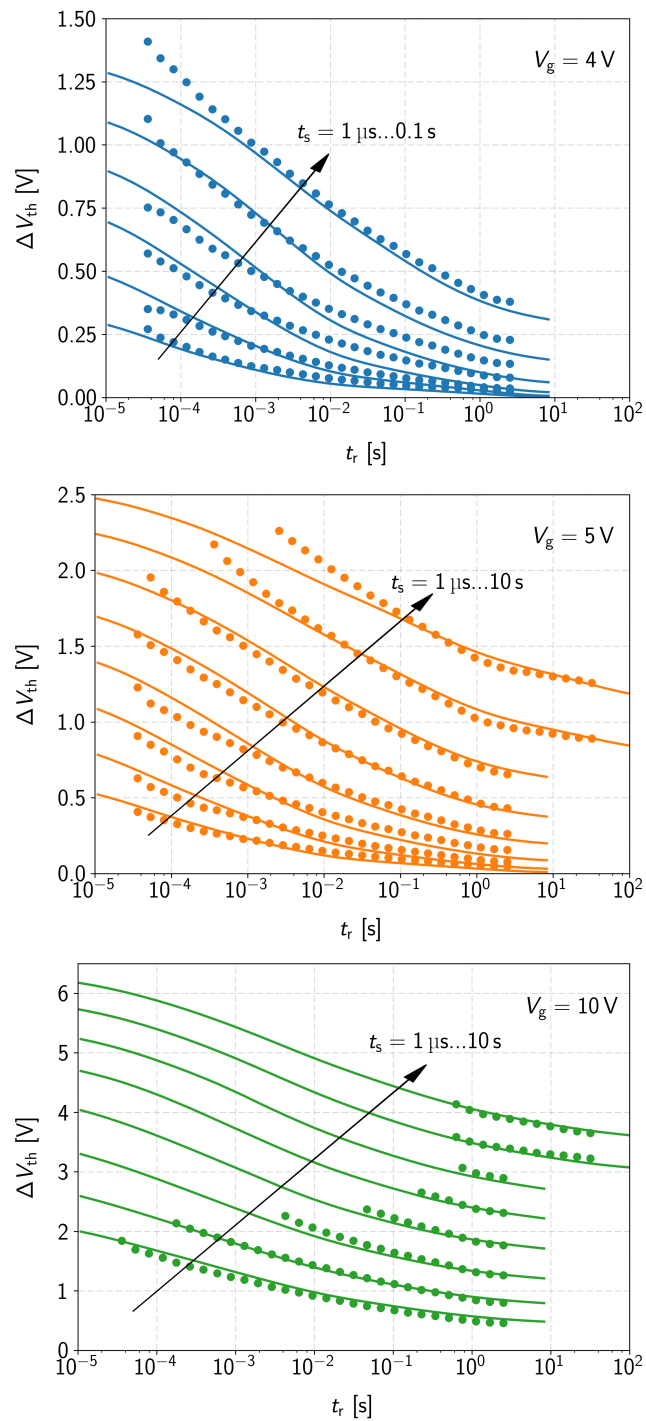


Figure 7.4: Simulated and measured ΔV_{th} recovery traces for eMSM sequences with stress times up to 10 s at room temperature. With two sets of normally-distributed NMP defects, the recovery behavior can be modeled for three different stress voltages of 4 V (blue), 5 V (orange) and 10 V (green) (reprinted from [AGC1]).

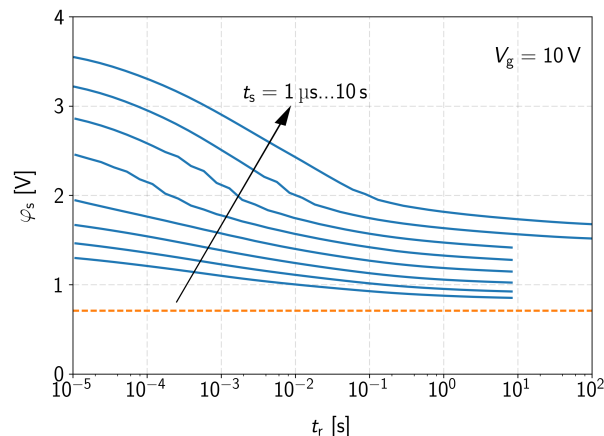


Figure 7.5: The surface potential underneath the gate as a function of recovery time for different stress times. The stress voltage in this case was 10 V. The value in thermal equilibrium is plotted as the dashed orange line. For short recovery times, the transient change in the surface potential can be as high as 3 V (reprinted from [AGC1]).

Electrostatic Charge Feedback

In contrast to silicon technologies, where at typical BTI recovery conditions a medium to strong electron channel is present at the dielectric interface, the situation in GaN/AlGaN MIS-HEMTs is different. Even if the device is still turned on at recovery, the barrier layer separates the dielectric interface from the electron channel. The polarization induced field causes a fully depleted barrier layer even for recovery voltages around 0 V (even more so for more negative gate voltages). In the case of silicon, the surface potential during recovery is pinned by the electron channel and thus its sensitivity on charge trapping is usually negligible. On the other hand, in GaN devices the depleted barrier layer causes a strong sensitivity of the surface potential on charge trapping.

Electrostatic charge feedback (i.e., the change of the surface potential due to charge trapping) is even more amplified because of a large number of charges trapped at the dielectric interface as compared to BTI in silicon devices. The impact of V_{th} drift on the surface potential for a stress bias of 10 V is depicted in Figure 7.5. For long stress times, the values can be up to 3 V larger when compared to the equilibrium value (dashed line). Naturally, this effect has a significant impact on the V_{th} drift characteristics.

On one hand, the trap levels of the defects close to the interface will experience approximately the same amount of shift as the potential at the interface. Since that shift enters the capture and emission times exponentially, the recovery times of all defects will show a strong transient behavior following the partial recovery in the surface potential shift. On the other hand, the active energy area of the defects will also be a strong function of the trap occupancies, causing atypical phenomena like charge capture during recovery and V_{th} recovery during stress.

As the surface potential partly recovers over time, these effects become less and less prominent. Nonetheless, a small portion of NMP defects will not emit their charge, leaving behind a quasi-permanent shift in the surface potential. This also has an effect on defect modeling, as the experiments can no longer be considered to be conducted on a memoryless system. In this case, the defect behavior is a function of the stress history of the device, rendering concepts like defect modeling based on cumulated stress times only useless.

The overall impact of the charge feedback on the capture and emission dynamics of the device

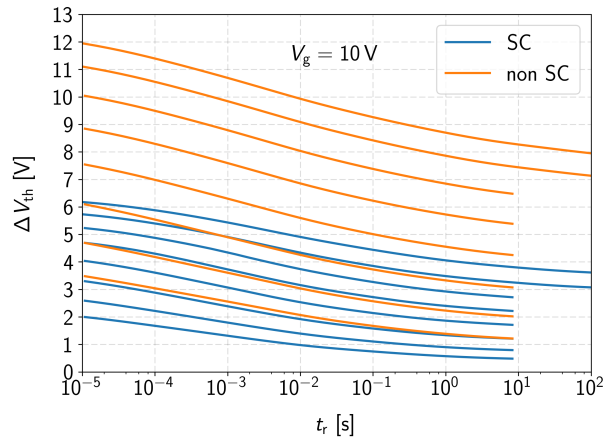


Figure 7.6: The difference between the observed degradation when simulated with or without self-consistent treatment of the defects. If the charges are not considered self-consistently, the degradation is overestimated by about a factor of 2. Note that not only the amount of trapped charges changes but also their dynamic behavior (reprinted from [AGC1]).

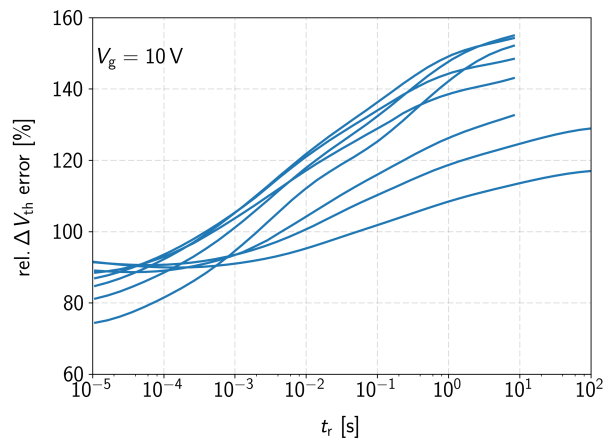


Figure 7.7: The relative error between the degradation simulated with and without charge feedback. The nonlinearities show that not just the amount of trapped charges change but also their kinetics (reprinted from [AGC1]).

can be seen if the same set of defects is simulated once self-consistently and the other time non-self-consistently. As can be seen in Figure 7.6 the amount of degradation is overestimated by a factor of about 2 if the trapped charges are not considered in the simulations. Moreover, it is not just the amount of trapped charges that changes but also their kinetics. These changes in kinetics can be seen best if the relative error between the two simulations is plotted (Figure 7.7). For low recovery times the error, although still huge, is smaller compared to higher recovery times. The general trend can be explained by a simple accumulation effect of the errors made at each instant in time. For a constant relative error, the accumulated errors need to have a linear relation. Thus the nonlinearities show that also the dynamic behavior during recovery changes. Another finding in this figure is that for longer stress times, the relative error decreases (the 100 s traces are related to stress times of 1 s and 10 s). This can be explained by a charge saturation effect similar to [66].

In a first-order approximation, these nonlinearities can be attributed to changes in the active energy area and the trap levels as mentioned earlier in the text. The changes due to charge feedback can be best understood when looking at the CET maps [195] of the defects at different stress bias for different instants in time, see Figure 7.8. At the beginning of the first stress cycle most

of the defects, previously being empty, are brought into the active energy area where they can potentially capture an electron (top row). If a significant amount of them captures a charge, the positive charges present at the dielectric interface decrease, causing an increased surface potential. During recovery, not all charges can be emitted so subsequent stress cycles will experience an increased surface potential. This potential shift causes a higher trap level for the electrons, decreasing the amount of trapped charge as compared to a fresh device. The very same mechanism is responsible for the change in the active energy area too, where the residual surface potential shift lowers the effective oxide field during stress. Because of that, even for longer stress times only a smaller amount of defects is able to capture a charge.

The residual amount of trapped charges of course strongly depends on the applied stress voltage, the stress time and the recovery period allowed between two subsequent stress cycles. In the second row of [Figure 7.8](#) the defects after the last stress cycle are plotted. Electrostatic charge feedback, which was explained in the last paragraph, then already caused a significant shift of the time constants towards larger values. The asymptotic limit of this mechanism can be seen in the third row, where the thermodynamic equilibrium of the defects is shown. Note that the shift of the time constants as compared to the transient results covers additional five orders of magnitude.

Local Potential Perturbations

The changes in the trap level and the active energy area are however not the only feedback mechanisms that can potentially influence the defect kinetics. For example, local changes in the potential caused by a single charge captured by a defect can influence the capture and emission times of neighboring defects. On top of that, the spatial distribution of defects could cause an inhomogeneous lateral surface potential profile. A third mechanism would be band bending due to the vertical distribution of defects in the oxide.

Accurate modeling of those effects is troublesome, as their individual influence on the overall degradation cannot be separated well enough. While the first two effects could in principle be tackled by adjusting the spatial distribution of interface and oxide defects, the latter one is much more difficult to describe. Due to the non-local nature of the NMP defects, corrections for the local potential fluctuations across neighboring grid points would need to be calculated. Concepts like direct interactions between defects are even harder to implement as they would require to solve large systems of equations and accurately extrapolate some of the grid quantities to the place of the defect. A basic discussion about different methods to estimate the influence of a charged defect on the local potential surface can be found in [Section 7.2.4](#).

7.1.4 Conclusions

In this section it was shown that the forward bias V_{th} drift observed in large-area GaN/AlGa_N MIS-HEMTs can accurately be modeled by two sets of normally distributed NMP defects. A careful selection of key parameters for the device simulation like the amount of polarization induced charges, bulk dopings and transport mechanisms in the barrier layer allowed to thoroughly investigate different feedback mechanisms having an influence on the experimentally observed capture and emission times. The most important one was identified as the change in the surface potential due to charge trapping, resulting in increased defect levels and a decreased active energy area seen by the defects.

Due to these effects, modeling approaches like the usage of accumulated stress times cannot be used as the observed defect kinetics are a strong function of previous stress and recovery cycles in

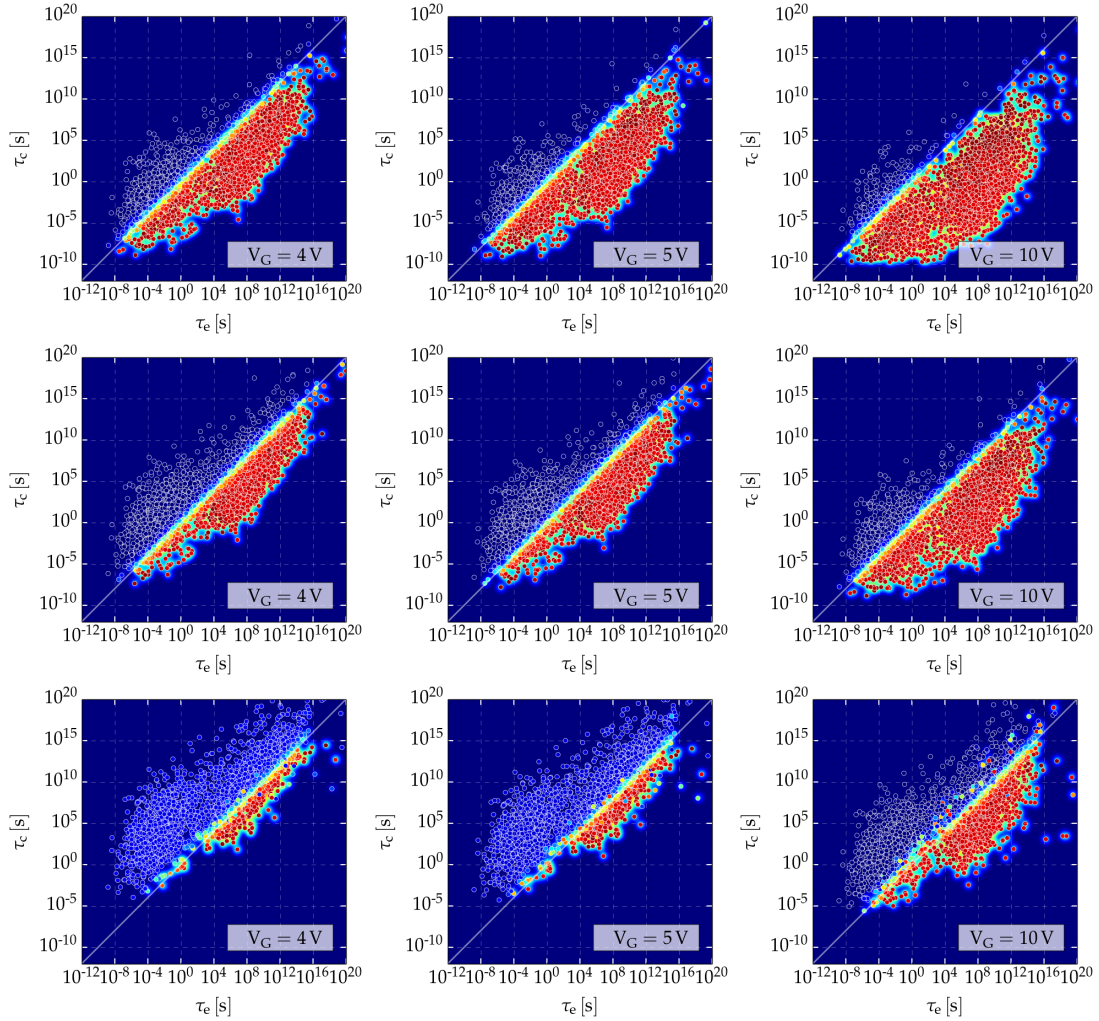


Figure 7.8: The CET maps for stress voltages of 4 V, 5 V, and 10 V (from left to right). The top and mid rows are extracted after accumulated stress times of 1 μ s and 11.11 s, whereas the bottom row shows the thermodynamic equilibrium after infinite recovery. It can be seen that the capture and emission times change due to changes in the trap levels and the active energy area. Unlike in silicon technology, they depend strongly on the stress bias and stress time, but also on the amount of recovery between the stress cycles. With increasing stress times, the charge distributions in the CET map converge toward their equilibrium values (bottom row) (from [AGC1]).

the device. Thus, for this kind of devices, it is of utmost importance to use transient simulations following the experimental bias conditions as closely as possible.

7.2 Single Defect Characterization in GaN/AlGaN MIS-HEMTs

The ability to create normally-off devices within the GaN/AlGaN material system is essential especially for power switching applications. In addition to the previously discussed methods to create normally-off devices, like p-doped structures or recess etched devices (see [Section 2.3.3](#)), another promising approach was presented by using nano-sized fin-MIS-HEMT structures [[196](#)].

In this section, single defect parameters like the trap level and the vertical defect positions are extracted from RTN measurements on these devices at different cryostatic temperatures using spectral maps as introduced in [Section 6.6.3](#). The question if the observed signals emerge from two coupled pairs of two-state defects or a single, more complex defect structure will be answered by evaluating the necessary coupling factors from RTN simulations and comparing them to those calculated for a chosen defect candidate using different methods. Based on these findings, two possible defect structures are investigated by extracting their characteristic time constants using HMM training.

7.2.1 Motivation

Single-defect studies were previously used to identify defect candidates responsible for BTI in silicon devices. While defects like the oxygen vacancy could be ruled out by comparing first-principle simulations to the extracted properties, other ones like the hydrogen bridge or the hydroxyle E' center are still considered to be promising candidates [[86](#), [110](#), [128](#)]. Following the same strategy for GaN technology could help to mitigate the influence of the surface donors on BTI by engineering their electrical properties within the growth process. First attempts to passivate the interface already have been made by using forming gas anneals before the SiN deposition [[197](#), [198](#)] or by using Fluorine treatments [[65](#), [199](#), [200](#)]. These, however, brought only limited improvements regarding the reliability of the devices.

For regular high-power devices, the relatively high defect densities together with the inevitable feedback mechanisms as discussed in [Section 7.1](#) make the experimental characterization of the responsible defects extremely challenging [[6](#), [44](#), [75](#), [AGC1](#)]. Up to date, the main efforts to relate different reliability issues like current collapse, on-state resistance degradation or voltage breakdown to charge trapping were focused on bulk-type defects (see [Section 3.2.1](#) and [[55](#)]). While this strategy seems reasonable for current degradation phenomena and drain voltage breakdown, other reliability issues like BTI and gate voltage breakdown cannot be understood correctly without taking into account the defects within the insulator as well as those at the barrier/insulator interface.

The latter have been shown to be dominant especially at forward gate bias conditions [[44](#), [65](#), [75](#)] when the polarization field in the barrier is compensated by the gate bias (overspill region). The fact that large area devices contain ensembles with a very large number of defects at the interface with broad distributions of capture and emission times further complicate the analysis of individual trap properties [[75](#), [AGC1](#)]. To gain a better insight into the physical mechanisms of charge trapping and possibly the nature of the involved defects, nanoscale devices are very promising because they allow investigating individual single-defect properties.

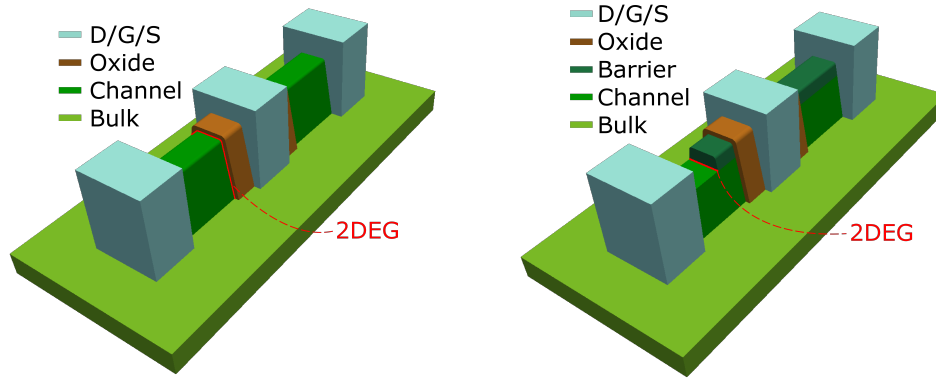


Figure 7.9: Schematic pictures of a regular silicon fin FET (left) and the GaN/AlGaN fin MIS-HEMTs used in this work (right). The main difference is that for the silicon device, the inversion channel is concentrated at the oxide interface, while for the other device the channel is located at the channel/barrier interface.

7.2.2 Electrostatic Device Simulation

In silicon technology, fin FETs have been widely adopted in state of the art technologies due to their improved gate coupling, allowing chip designs with lower gate delays and increased energy efficiency as compared to planar technologies [201–203]. However, the operation principle of the devices investigated in this thesis is entirely different from those in silicon technology due to the native electron channel present at the GaN/AlGaN interface. Figure 7.9 shows schematic pictures of a regular silicon fin FET and the GaN/AlGaN fin MIS-HEMTs used in this work. For the silicon device, the inversion channel is concentrated at the oxide interface, while for the GaN device the channel is located at the channel/barrier interface.

This peculiarity has some interesting consequences. First and foremost it enables to geometrically engineer the sheet carrier density independently from the aluminum content of the barrier as shown in [196, 204]. The mechanisms behind that are thought to be the depletion of the channel from the sidewalls as well as a partial relaxation of the strain induced polarization charges. A simulation study in [204] showed horizontal cuts through the channel for different fin widths. As can be seen in Figure 7.10, for wider fins the sheet carrier density is dominated by the polarization charges. At about 70 nm and below, the channel is more and more depleted, leading to positive threshold voltages for narrow fins.

Unlike the devices investigated in [204] which were fabricated with Schottky gates, the devices in this work were insulated by a 20 nm thick high-quality aluminum oxide layer. The HEMT was formed by a 30 nm thick AlGa_N barrier (30% Al) on top of a 80 nm Ga_N channel layer and a 2 μm thick highly resistive Ga_N buffer. The lateral dimensions of the fin were about 50 nm x 1 μm.

The fin structures were formed by electron-beam lithography and a subsequent wet-etch removing the barrier and the channel layers. Due to that, parasitic heterojunction free MOS devices are formed at the sidewalls and the highly resistive Ga_N bulk. A more detailed description of the devices and the fabrication can be found in [196]. The parasitic Ga_N MOSFETs add a second operation regime to these devices where at positive gate voltages majority carriers are accumulated and notably contribute to the overall drain current.

The two operating regimes, regular 2DEG conduction and sidewall accumulation are shown in Figure 7.11. The 3D-simulations were carried out with Minimos-NT [21] on a simplified geometry where the access regions as well as the source and drain regions were scaled in order to speed up the computations. The piezoelectric charges at the hetero-interface were chosen according to

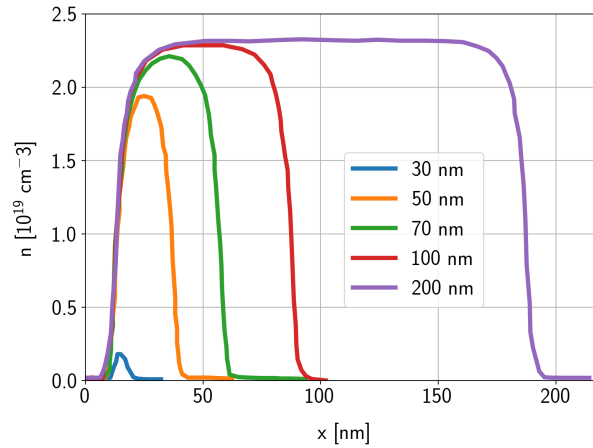


Figure 7.10: Geometric scaling of the sheet carrier densities in GaN/AlGaIn fin HEMTs. For fin widths smaller than 70 nm, the channel is more and more depleted from the sidewalls and partial strain relaxation of the barrier region (data taken from [204]).

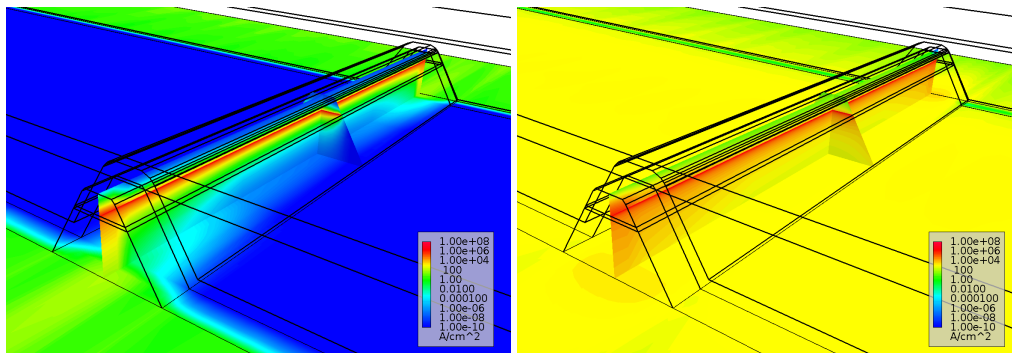


Figure 7.11: The current density for the two operation regimes of the GaN/AlGaIn fin MIS-HEMT used for single-defect measurements. **Left:** For negative gate voltages, the parasitic sidewall MOSFET is depleted and the drain current is concentrated at the heterojunction. **Right:** For positive gate voltages, the sidewall MOSFET is in accumulation, which can be seen also as a second kink in the $I_d(V_g)$ characteristics in these devices (see Figure 7.12 and [196, 205]).

[192] and a thermionic field emission model was used to calculate the transfer of electrons and holes across the heterointerface at the channel. The doping concentrations and the compensation of the piezoelectric charges at the barrier-oxide interface were calibrated using $I_d(V_g)$ characteristics at different temperatures. More precisely, the doping concentrations were obtained by utilizing the sidewall accumulation regime as shown in the right picture in Figure 7.11, which adds to the slope of the on-current in the $I_d(V_g)$ characteristics and produces a second kink at about $V_g = 1$ V (Figure 7.12).

7.2.3 Extraction using Spectral Maps

In this section, the spectral method presented in Section 6.6.3 will be used to obtain the defect properties of four dominant defects observed in a nanoscale GaN/AlGaIn fin MIS-HEMT [AGC5]. Examples of the measured RTN signals of the dominant defects are given in Figure 7.13. At this point, the observed capture and emission events are treated as separately for each level, leading to four different defects called ‘A’, ‘B’, ‘C’ and ‘D’. Later it will be shown that it is rather unlikely that the observed traces were created by two pairs of coupled defects for different reasons.

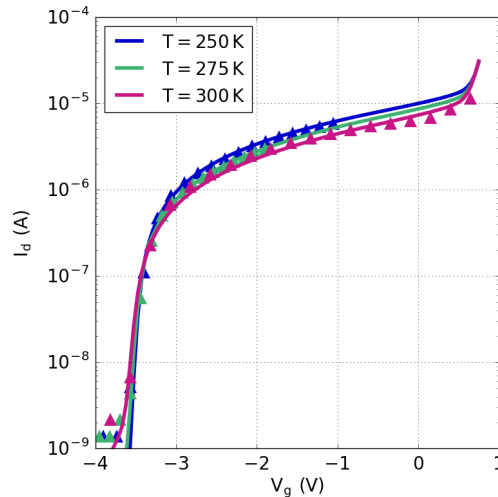


Figure 7.12: The $I_d(V_g)$ characteristics of the device at different temperatures. The accumulation regime at positive gate voltages can be seen as an additional kink in the curve at room temperature. Since the curves have been recorded prior to each RTN measurement sequence, they have been corrected by the accumulated BTI-related V_{th} shift (from [AGC5]).

After step detection and the extraction of the delta times for capture and emission events, spectral maps similar to Section 6.6.3 are constructed to extract the characteristic times of one or more different defects as shown in Figure 7.14. Since the measurement noise also affects the accuracy of the step height extraction, the clusters are spread in both directions. For noise-free signals, the clusters would only be spread in time due to the stochastic nature of the emissions. Note that for RTN signals the mean step heights for charge capture and emission for a specific defect have to be symmetric along the y-axis. To extract the characteristic times, the generated maps are simply binned into histograms according to their step heights. According to (6.20), the capture and emission times for simple two-state defects are exponentially distributed. If only one cluster is found for a specific step height, the histogram consists of a univariate exponential distribution and the characteristic time constant can be simply estimated from the mean value of the observed times. This is possible because the expectation value of an exponential distribution is the characteristic time constant of a two-state Markov chain (see (6.22)).

For more than one defect generating emissions with a certain step height, the mean values could be taken independently for each cluster. However, if the separation between the clusters in time is not sufficient, especially the minor distributions with less samples (i.e. the ones with the larger time constants) often cannot be extracted accurately. This case is depicted for the emission times in the spectral map shown in Figure 7.14. A more reliable approach is then to fit multivariate exponential distributions directly to the histogram. The only parameter that has to be chosen beforehand is the number of defects (i.e. the number of distributions). Note that the number of visible defects can potentially change with voltage and temperature.

In the right picture of Figure 7.14, the resulting histogram together with the fitted distributions are shown. The mean values of the distributions correspond to the capture and emission times of the defects for a certain voltage and temperature. If this procedure is repeated for all bias conditions and temperatures, one can obtain the voltage dependence and temperature activation of all (observable) defects. The results of this extraction can be seen in Figure 7.15. Note that those characteristics were obtained without any assumptions regarding a specific physical model except the system being memoryless (i.e. being a Markov chain).

To estimate the most likely location of the traps based on the observed step heights, TCAD

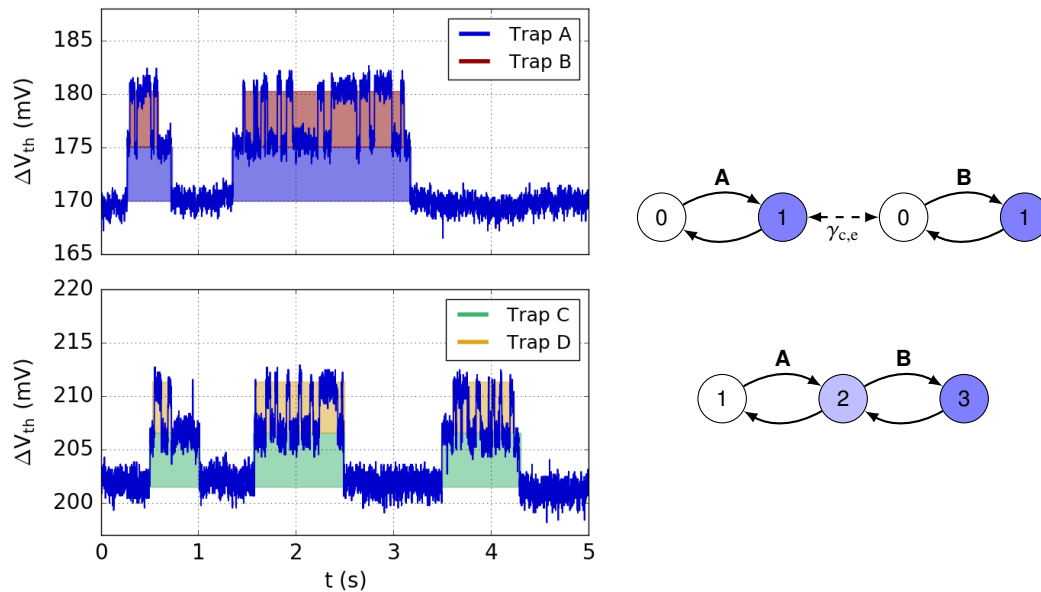


Figure 7.13: **Left:** The measured RTN signal after mapping to ΔV_{th} . The emissions ‘B’ and ‘D’ are only active if in ‘A’ and ‘C’ an electron has been captured. When treated as separate defects, this fact together with the similar step heights indicates that the traps would be in immediate vicinity to each other. The sampling frequency for all measurements was 10 kHz. For illustration purposes, the signal is shown after filtering with a median filter. **Right:** Two possible Markov chains resembling the observed correlated RTN signals. The similar step heights together with the correlated emissions either indicate a strong coupling between two independent two-state defects (top) or a single three-state defect with one electron captured for each state (from [AGC5]).

simulations with single charges along a horizontal cut 0.5 nm above the channel and a vertical cut through the mid-section of the device were conducted. Single electron fixed charges were placed at different positions along the cuts and their electrostatic influence on V_{th} was calculated directly from the corresponding $I_d(V_g)$ curves. Simulations with randomly distributed dopants were conducted, leading to a distribution of step heights, compare shaded areas in Figure 7.16. The results indicate that the defects most likely reside close to the barrier/channel interface. Deviations from the observed step heights could either be explained by the trap locations being at an especially critical place on the percolation path, double-emitting traps [206] fast enough to be sampled out from the measurement signal or some other source of variability missing in the simulations. Defect candidates for double-ionized defects have been predicted in first-principle simulations for silicon [207] and GaN technology [57]. However, the likelihood of a double-emitting defect with the intermediate transition fast enough to be consistently missed due to sampling can be considered low because of the related structural relaxations happening between the double-charged state and the neutral state.

More information about the defects despite their empirical voltage and temperature behavior can also be extracted directly from the curves shown in Figure 7.15 if a defect model is chosen. The bias conditions at which the RTN was observed indicate that the barrier is highly depleted, thus acting as a quasi-insulator. For that reason, band bending can be neglected and a constant field across the barrier region can be assumed. If the most simple case of a two-state NMP trap is chosen for the defect model, this enables the extraction of the trap levels, the trap positions and the activation energy of the defects observed in the measurements (see Section 5.2 and band diagram in Figure 7.17).

Starting from the reaction barriers \mathcal{E}_c for electron capture and \mathcal{E}_e for electron emission, their

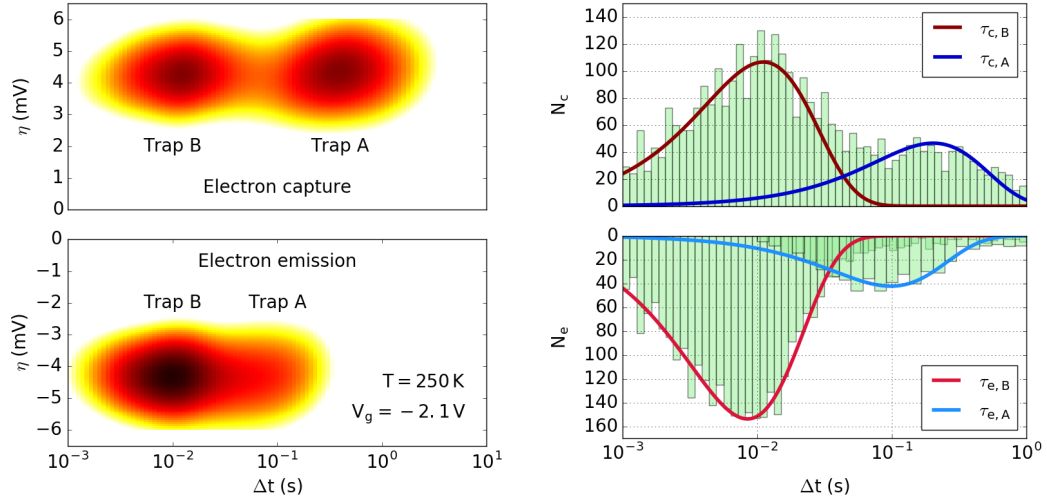


Figure 7.14: **Left:** After step detection and the extraction of the delta times for capture and emission events, a spectral map consisting of the step heights versus the delta times is plotted. In order to extract the time constants of different defects, those spectral maps are split into one or more histograms by the observed step heights. For RTN signals, the step heights extracted for capture and emission have to be symmetric around 0. **Right:** The histograms of the capture (top) and emission (bottom) times show the number of transition events at a certain gate voltage and temperature. The capture and emission times of each individual trap can be obtained most reliably by fitting multivariate exponential distributions to the data. Because of (6.20), the resulting parameters are equal to the characteristic capture and emission times of the defects (from [AGC5]).

bias dependence is used to extract the aforementioned properties. The Arrhenius equation in its logarithmic form reads:

$$\ln\left(\frac{1}{\tau_{c,e}}\right) = \ln(\sigma) - \frac{\mathcal{E}_{c,e}}{k_B T} \quad (7.4)$$

The bias dependence of the reaction barriers is given by the partial derivative with respect to V_g and cancels out the exponential pre-factor $\ln(\sigma)$ (assumed to be independent of the gate bias).

$$\begin{aligned} \frac{\partial}{\partial V_g} \left[\ln(1) - \ln(\tau_{c,e}) \right] &= \frac{\partial}{\partial V_g} \left[\ln(\sigma) - \frac{\mathcal{E}_{c,e}}{k_B T} \right] \\ k_B T \frac{\partial \ln(\tau_{c,e})}{\partial V_g} &= \frac{\partial \mathcal{E}_{c,e}}{\partial V_g} \end{aligned} \quad (7.5)$$

The energy difference between capture and emission barriers is equivalent to the shift of the barrier potential at the location of the trap because the ionization energy of the trap to the local conduction band edge being constant. Thus the trap level shifts together with the local barrier potential (see Figure 7.17). This can be used to relate the bias dependence of the surface potential to the bias dependence of the capture and emission barriers.

$$\frac{\partial \mathcal{E}}{\partial V_g} = \frac{\partial \mathcal{E}_c}{\partial V_g} - \frac{\partial \mathcal{E}_e}{\partial V_g} = \frac{\partial \varphi_s}{\partial V_g} \quad (7.6)$$

As mentioned before, the bias conditions suggest a highly depleted barrier. Thus a capacitive voltage divider can be used to calculate the potential at the trap position from the gate voltage.

$$\varphi_s = \frac{1}{1 + \frac{\epsilon_{\text{bar}} t_{\text{ox}}}{\epsilon_{\text{ox}} t_{\text{bar}}}} \frac{y_{\text{trap}}}{t_{\text{bar}}} V_g \quad (7.7)$$

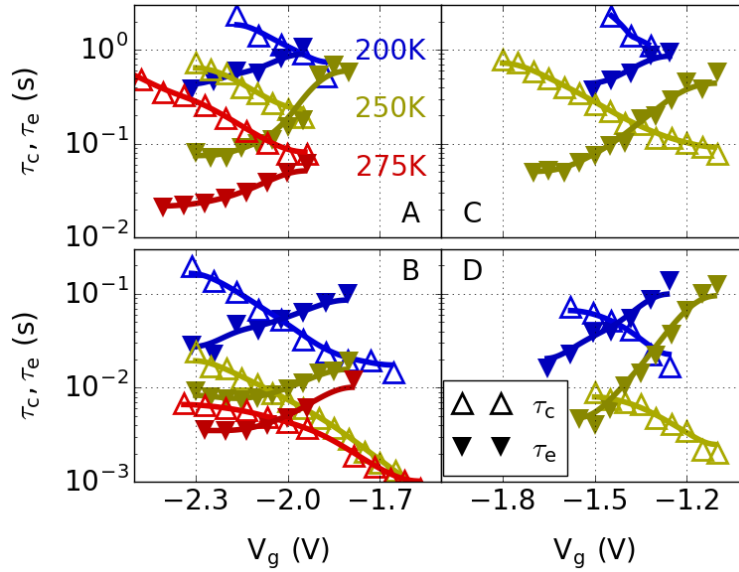


Figure 7.15: The extracted capture and emission times for the two pairs of coupled defects. The weak voltage dependence suggests a defect location close to the channel. The defects ‘A’ and ‘B’ as well as ‘C’ and ‘D’ share the same intersections and voltage dependency, a further argument in favor of a strong coupling. Based on the selection of a certain defect type, in [Section 7.2.4](#) realistic coupling factors will be derived. In [Section 7.2.5](#), alternative defect structures in terms of their Markov states are proposed and their time constants are extracted by HMM training (from [\[AGC5\]](#)).

The local shift of the trap level is then a function of the vertical trap position and can be calculated from (7.6) and (7.7). Solving for the trap position y_{trap} yields:

$$y_{\text{trap}} = \left(t_{\text{bar}} + t_{\text{ox}} \frac{\varepsilon_{\text{bar}}}{\varepsilon_{\text{ox}}} \right) \frac{\partial \mathcal{E}}{\partial V_g} \quad (7.8)$$

At the bias conditions of the intersection point of the curves in [Figure 7.15](#), the trap level is equal to the Fermi-level at the position of the trap (compare band diagram in [Figure 7.17](#)). This fact can be used to calculate the trap level if the effective conduction band potential of the GaN/AlGaN interface φ_{ch} and the surface voltage at $V_g = 0$ V, $\varphi_{\text{s},0}$, are known. In fact, the surface voltage at any gate voltage would be sufficient, as long as the barrier is still sufficiently depleted to neglect band bending, (i.e., the capacitive voltage divider in (7.7) is justified). In this case, the surface potential and the band offset were taken from TCAD simulations. With $V_{\text{g,int}}$ being the gate voltage at the intersection point of τ_c and τ_e , the equation to calculate the trap level E_T reads:

$$E_T = \frac{\partial \varphi_s}{\partial V_g} V_{\text{g,int}} + (\varphi_{\text{s},0} - \varphi_{\text{ch}}) \frac{y_{\text{trap}}}{t_{\text{bar}}} + \varphi_{\text{ch}} \quad (7.9)$$

Another thing which can be extracted from the measurements is the temperature activation of the process. When assuming equal curvatures of the parabolas from the two-state NMP process, the energy barriers for electron capture and emission can be calculated from the energy barriers for the relaxation energy \mathcal{E}_R and the minimum-energy difference \mathcal{E} :

$$\mathcal{E}_c = \frac{(\mathcal{E}_R + \mathcal{E})^2}{4\mathcal{E}_R}, \quad \mathcal{E}_e = \frac{(\mathcal{E}_R - \mathcal{E})^2}{4\mathcal{E}_R} \quad (7.10)$$

The squares in equation (7.10) can be expanded if $\mathcal{E}_R \gg \mathcal{E}$ is assumed, which is the case in

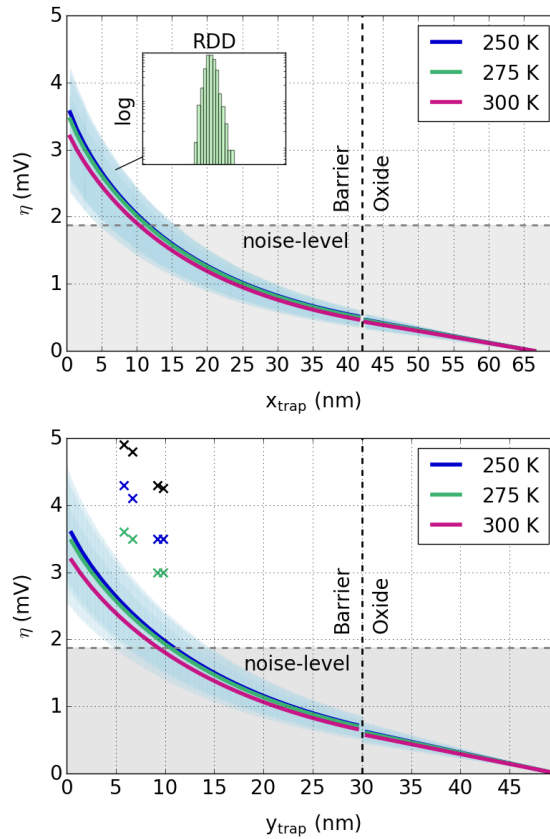


Figure 7.16: With random dopant simulations are used to estimate the step heights caused by individual defects for single fixed charges along the channel interface and vertically through the middle of the channel. The threshold voltage shifts were directly extracted from the corresponding $I_d(V_g)$ curves (from [AGC5]).

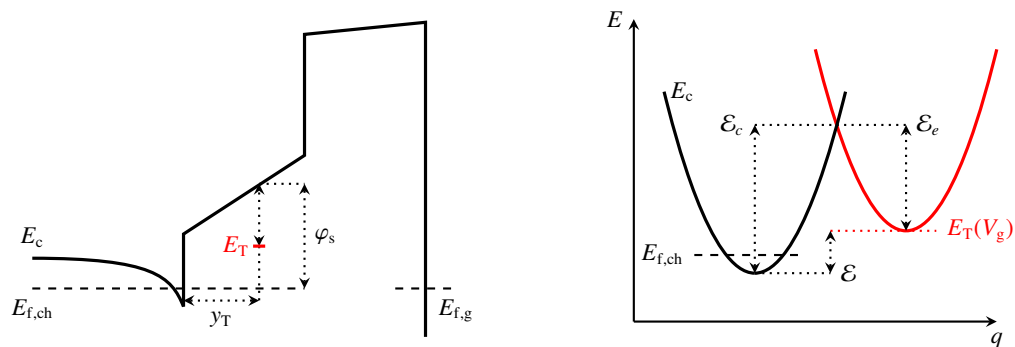


Figure 7.17: Left: The schematic conduction band diagram of the fin-MIS-HEMT structure with the extracted quantities. The surface potential at $V_g = 0$ V was taken from the simulations conducted in Section 7.2.2.

Right: Schematic adiabatic potential energy surface of a two-state NMP defect with equal curvatures. If band-bending in the barrier can be neglected, the trap level depends linearly on the applied gate voltage (from [AGC5]).

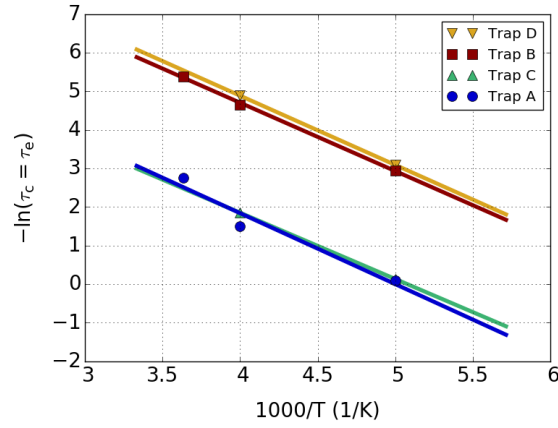


Figure 7.18: The Arrhenius plot of the four investigated defects. All defects share about the same temperature activation energy of $\mathcal{E}_a = 0.15$ eV. This observation also supports the validity of the extracted trap positions because of the crystalline AlGa_N layer having a narrower distribution of defect properties than the amorphous oxide (from [AGC5]).

Trap	y_{trap} (nm)	E_T (eV)	k_0 (s ⁻¹)	E_R (eV)
A	6.7	0.63	1.1×10^7	0.63
B	5.8	0.59	5.5×10^9	0.61
C	9.2	0.68	7.3×10^6	0.59
D	9.8	0.72	5.2×10^9	0.62

Table 7.1: The defect properties extracted from Figure 7.15. The extracted vertical trap positions show that the correlated trap pairs ‘AB’ and ‘CD’ are most likely in close proximity to each other. The extracted trap levels of the defect pairs also match closely, confirming their similar positions within the barrier. If the assumption of two pairs of defects is correct, their similar intersections, their voltage dependence and their correlated behavior suggest that the potential energy surface around defects ‘B’ and ‘D’ has to be altered if defects ‘A’ and ‘C’ have captured an electron.

the strong electron-phonon coupling regime. The temperature activation of the process is then dominated by the relaxation energy and the apparent activation energy can be estimated by $\mathcal{E}_R \approx 4\mathcal{E}_A$ as can be seen in (7.11). The Arrhenius plot of the four investigated defects in Figure 7.18 shows that all of them share about the same activation energy $\mathcal{E}_A \approx 0.15$ eV.

$$\mathcal{E}_c \approx \frac{1}{4}\mathcal{E}_R + \frac{1}{2}\mathcal{E}, \quad \mathcal{E}_e \approx \frac{1}{4}\mathcal{E}_R - \frac{1}{2}\mathcal{E} \quad (7.11)$$

The summary of the extracted trap properties is shown in Table 7.1. Since the intersections of the correlated defects pairs are very similar, also the vertical positions of the traps seem to be in close proximity to each other. For the same reason, they also share the same trap levels although the observed time constants differ by one order of magnitude. The only remaining parameters to describe the differing time constants are the exponential pre-factor k_0 . It can be seen that the factors differ by two orders of magnitude between the slow and fast time constants, while the two defect pairs almost share the same pre-factors.

Given that the assumption of two coupled pairs of defects holds, the validity of the extracted trap parameters before primarily depends on (i) the accuracy of the step detection, (ii) the accuracy

of the algorithm to extract the delta times for capture and emission times and (iii) the signal-noise ratio of the traps under investigation. The next section will investigate how to model the electrostatic coupling between two defects and how to estimate the coupling factors necessary to obtain RTN signals comparable to [Figure 7.13](#).

7.2.4 Electrostatic Coupling of Defects

In a first order approximation, the electrostatic coupling between two independent defects can be modeled as a shift of the local potential due to the Coulomb potential of a single point charge [208–211]. Other approaches calculated the change in the Coulomb energy of a given MOS system if a single defect has captured a charge [212–215].

The perturbation of the local potential caused by one defect can act on the other in two different ways, either as an additional local Coulomb barrier shifting the initial trap level or remotely via two distant traps being on the same percolation path [156, 157]. In the case of the correlated RTN observed in this work, probably the first mechanism dominates since the coupling of the time constants seems to be rather large. This is mainly because the additional Coulomb barrier enters the equations exponentially and the supply of carriers (i.e. the percolation path) only linearly. Other arguments speaking against a lateral separation of the coupled defects are their similar step heights (i.e. trap depths) as well as the bias region they were observed at. The coupling via percolation path would be strongest at weak channels for voltages around V_{th} , however, the experiments were conducted at bias conditions quite far from that value.

The magnitude of the potential perturbation and thus the trap level shift caused by a nearby trap is hard to predict if only classical or semi-classical simulations are available. One obvious approach is to treat defects as point charges and add them to the discretized Poisson equation. It has to be noted at this point that 3D-simulations are required for that task in order to obtain correct results. This is because the solution in two dimensions gives the potential of an infinitely long line charge. For one dimension the solution will resemble an infinitely large sheet charge, independently of discretization.

In the case of a point charge, the potential is inversely proportional to the distance to the charge. With r being the distance to the charge, the solution is given by the well-known equation for the Coulomb potential:

$$\varphi = -\frac{q_0}{4\pi\epsilon_0\epsilon_r r} \quad (7.12)$$

On the other hand, the solution for an infinitely long line charge gives a logarithmic dependence of the potential with λ being the charge density and r the orthogonal distance from the line:

$$\varphi = -\frac{\lambda}{2\pi\epsilon_0\epsilon_r} \ln(r) \quad (7.13)$$

[Figure 7.19](#) depicts the numerical and analytical solutions for the potential of a point charge in two and three dimensions. As mentioned before, 2D-simulations cannot be used because of the wrong overall trend (inversely proportional vs. logarithmic). In the three dimensional case and with sufficiently fine grids, the near-field impact of a single charge is overestimated whereas the far-field impact is underestimated. To overcome this problem and to be independent of grid spacing, a first order quantum correction model like the density gradient (DG) model [216–218] can be used. Other approaches like the Conwell-Weisskopf model [219] divide the Coulomb potential into short-range and long-range terms, where only the long-range term enters the right-hand side of the Poisson equation. A more recent approach makes use of an analytical expression for the short-range force acting on a particle at a distance r from the charge. It resembles the

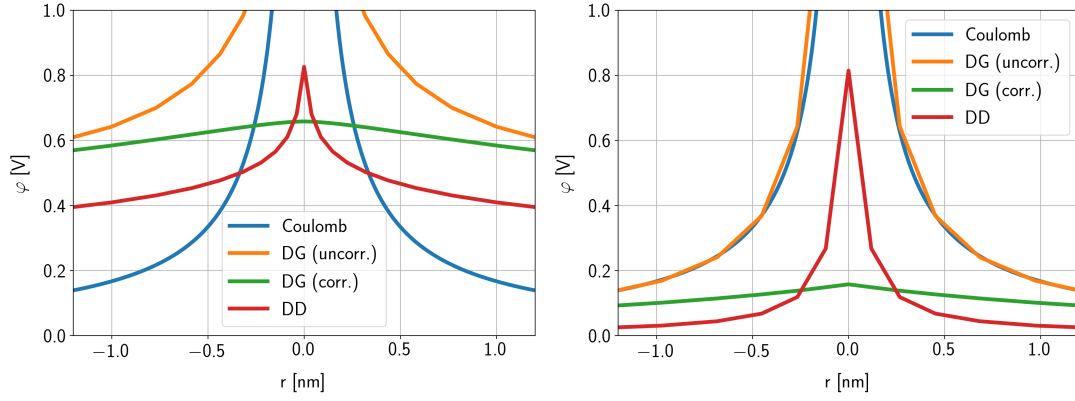


Figure 7.19: The numerical solution of a point charge in lightly doped GaN at $r = 0$ in two and three dimensions. Drift-diffusion (DD) simulations are known to predict unphysical charge crowding for low grid-spacing. To overcome this problem and to be independent of grid spacing, a first order quantum correction model like the DG model can be used. **Left:** In two dimensions, the numerical solution is the potential of an infinitely long line charge. This leads to an overestimation of the local impact of a single charge. **Right:** The three-dimensional solution correctly converges towards a Coulomb potential for intrinsic semiconductors. The density gradient model makes the local impact of a single charge independent of the grid and only the long-range part of the Coulomb potential is considered in the Poisson equation.

Coulomb force for large distances whereas at short distances it decreases to zero, removing the singularity and rapidly changing components from the equation [220].

$$E(r) = \frac{q_0 r}{4\pi\epsilon(r^2 + 0.5r_c^2)^{3/2}} \quad (7.14)$$

This force term has its maximum at the cutoff radius r_c and afterwards decreases monotonically towards the point of the defect. The value of the cutoff radius should be chosen according to the physical nature of the problem. The work of Alexander [211] assumes that the maximum of the electrical field associated with a single donor should be at the effective Bohr radius of the ground state of a donor. The effective Bohr radius for AlGaIn can be calculated using the relative permittivity $\epsilon_r = 8.6$ and the relative electron mass $m_e^* = 0.26m_e$:

$$r_c = \frac{4\pi\hbar^2\epsilon_0\epsilon_r}{m_e^*q_0^2} = 1.75 \text{ nm} \quad (7.15)$$

From the short-range force seen by an electron, the potential change and thus the local shift of the trap energy can easily be derived by integrating (7.14).

$$\varphi(r) = -\frac{q_0}{4\pi\epsilon_0\epsilon_r\sqrt{2r^2 + r_c^2}} \quad (7.16)$$

A straightforward way to check the validity of the solution provided in (7.16) is to compare it to the quantum mechanical solution for the ground state of a hydrogen atom with an effective Bohr radius as given in (7.15). The starting point of the derivation is the radial-symmetric charge density of an electron in the ground state.

$$\rho(r) = q_0|\psi_{100}|^2 = \frac{q_0}{\pi r_c^3} \exp\left(-\frac{2r}{r_c}\right) \quad (7.17)$$

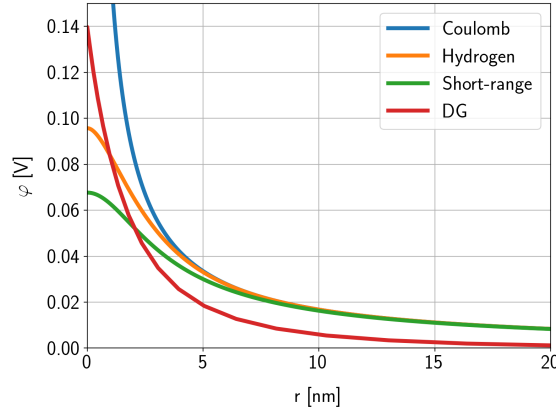


Figure 7.20: The results of the quantum corrected drift-diffusion simulations versus the analytic short-range potential used in [220] and the solution of the Hydrogen model. The two analytic potentials match very well and converge towards the solution of the Coulomb potential at larger distances. The difference to the numerical solution can most likely be explained by the fact that the simulations were conducted in semi-insulating GaN while the analytical solutions assume insulators.

By applying Gauss' law with a spherical ansatz, the absolute value of the electric field of the electron cloud is found by:

$$|E(r)| = \frac{q_0}{4\pi\epsilon_0\epsilon_r r^3} \left[1 - \exp\left(-\frac{2r}{r_c}\right) \left(1 + \frac{2r}{r_c} + \frac{2r^2}{r_c^2} \right) \right] \quad (7.18)$$

The potential of the electron of a hydrogen atom is again calculated from the electric field by integration of (7.18).

$$\varphi(r) = \frac{q_0}{4\pi\epsilon_0\epsilon_r} \left[\frac{\exp\left(-\frac{2r}{r_c}\right) - 1}{r} + \frac{\exp\left(-\frac{2r}{r_c}\right)}{r_c} \right] \quad (7.19)$$

The results of the analytic potentials versus the corrected numerical solution from Figure 7.19 can be seen in Figure 7.20. The deviation of the numerical solution from the short-range correction and the Hydrogen model could possibly be explained by two issues. The first one is due to the simulation itself, where additional screening due to the lightly doped semiconductor can influence the results. The second and probably more important one is the lack of a well-calibrated set of parameters for the density gradient model at the time of writing.

The results provided in Table 7.1 unfortunately cannot provide any information on the lateral distance between the traps, so the only way to estimate the coupling factors is to choose a suitable defect candidate based on its trap level. Due to the similar trap levels of the two coupled defects, the best guess is to make a worst-case assumption by assuming the very same type of defect in a nearest neighbor manner. Taking the values from TCAD simulations in MinimosNT, the conduction band minimum in the AlGa_N barrier is at approximately 3.7 eV. Based on the data in Figure 3.1 and assuming the same absolute energy levels in the barrier as compared to GaN, the most likely candidates turn out to be either a dislocation or a nitrogen vacancy. Since the nitrogen vacancy is one of the most common defects and likely to be responsible for the n-type conduction in GaN, it is chosen for the following extraction of the coupling factors.

To calculate the minimum distance between two nearest neighbor nitrogen vacancies it has to

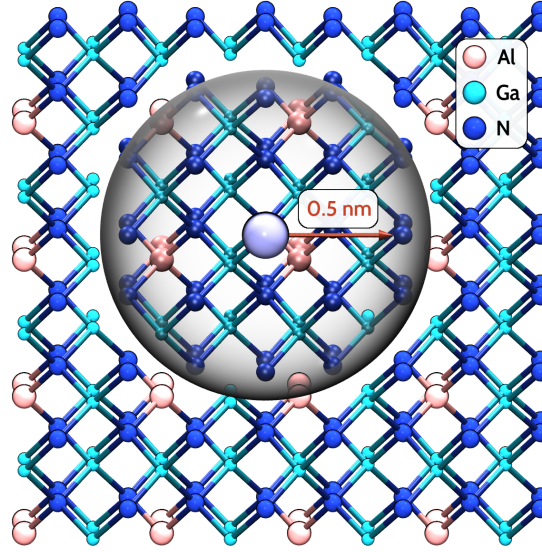


Figure 7.21: The distance between potential nitrogen vacancy sites in an ideal crystal structure of AlGaN. The minimum distance between two nitrogen sites not sharing the same Ga atom is approximately 5 Å. The light-blue atom represents one nitrogen vacancy, while the grey sphere indicates the minimum distance to the first nitrogen site not sharing the same gallium atom.

be taken into account that the two nitrogen vacancies should not share one Ga atom. This kind of defect would distort the overall structure too much and thus is unlikely to be stable. The lattice constants of wurtzite AlGaN alloy can theoretically be calculated based on their alloy composition x [27]:

$$a_{\text{Al}_x\text{Ga}_{1-x}\text{N}} = (3.1986 - 0.0891x) \text{ \AA} \quad (7.20)$$

$$c_{\text{Al}_x\text{Ga}_{1-x}\text{N}} = (5.2262 - 0.2323x) \text{ \AA} \quad (7.21)$$

In the case of the investigated devices the lattice constants for $x = 0.3$ are calculated to $a = 3.172 \text{ \AA}$ and $c = 5.157 \text{ \AA}$. From the crystal structure in Figure 7.21, the worst-case distance between two nitrogen sites not sharing the same Ga atom was calculated to be around 5 Å, thus the second nearest defect site has a distance of approximately 10 Å. Table 7.2 gives the resulting energy shifts for those two defects, which can easily be translated into coupling factors using the Arrhenius law for the appropriate temperatures.

$$\gamma = \exp\left(\frac{\Delta E_T}{k_B T}\right) \quad (7.22)$$

The calculated coupling factors are given in Table 7.3. If the results of the unscreened analytic Coulomb potential from (7.12) are neglected, realistic coupling factors are in the range of ~ 15 to ~ 500 . It should be noted that the provided results are a somewhat crude approximation because a structural defect in reality will always change the local configuration of atoms as well as their bonding lengths. On top of that, the hydrogen model is an oversimplification of the local potential surface which can only be provided by first principle simulations. The results however allow quantifying a range of realistic coupling factors that could be present in a worst-case scenario. A nice first principle study on the properties of other point defects in GaN combining DFT and quantum molecular dynamics simulations can be found in [57]. Now that a range of coupling factors is known, the coupling factors required to make the vast majority of the modified RTN emissions too fast for measurements with a fixed sample rate have to be calculated. For a fixed sampling time, the mean time to emission of a defect has to be approximately one order of

Model	ΔE_T	
	0.5 nm	1 nm
Coulomb	335 meV	167 meV
Hydrogen	92 meV	84 meV
Short-range	66 meV	63 meV
DG	105 meV	83 meV

Table 7.2: The trap energy shifts estimated from the data in [Figure 7.20](#) for the first and second nearest defect sites of nitrogen vacancies. The potentials of hydrogen model and the short-range model only show a weak distance dependency below the critical radius. On the other hand the Coulomb potential and the density gradient model predict a much larger difference between the first and second neighbor.

	γ (0.5 nm)			γ (1 nm)		
	200 K	250 K	275 K	200 K	250 K	275 K
Coulomb	2.76×10^8	5.67×10^6	1.38×10^6	16 152	2326	1150
Hydrogen	208	72	49	131	49	35
Short-range	46	21	16	39	19	14
DG	442	131	84	123	48	33

Table 7.3: The calculated coupling factors for the first and second nearest defect sites of nitrogen vacancies for the three different temperatures. If the results of the Coulomb potential are neglected, realistic coupling factors are in the range of ~ 15 to ~ 500 .

magnitude below that value because of the stochastic nature of the process. In order to do that, the HMM library presented in [Appendix A](#) can be used to simulate a coupled pair of defects using the extracted time constants from [Figure 7.15](#) with different coupling factors. The simulation results can be seen in [Figure 7.22](#) for a gate voltage of -1.95 V and a temperature of 250 K. To reliably suppress emissions for a sampling rate of 10 kHz as used in the measurements, a coupling factor of about $\gamma = 100$ is needed which is already at the upper limit of the calculated factors in [Table 7.3](#). Given all the uncertainties in the derivation of the time constants and the perturbation potentials, this result is still in range for the nearest-neighbor nitrogen vacancy. There is little literature on single defect studies focused on the coupling of traps in terms of their time constants [[156](#), [221](#), [222](#)]. These studies suggest that for strongly coupled defects, coupling factors between 10 and 20 seem to be realistic. Unfortunately they do not provide the temperature used during their measurements. Quite interestingly, the calculated coupling factors for 300 K and 1 nm are also in the range from ~ 11 to ~ 25 for the last three models in [Table 7.3](#), closely matching their observations.

7.2.5 Extraction using the Hidden Markov Model

As already briefly mentioned in [Section 6.6.3](#), the spectral method (together with all the other methods discussed in [6.6](#)) share the shortcoming that the real structure of a defect is obfuscated if the defect contains thermal transitions without charge transfer. In a best-case scenario, the time constants of the inactive state are much larger than the ones of the active states and are of the same order as the measurement time for one trace. Then the spectral method completely misses that state and the emission times for the fast state are not affected at all. On the other hand, if

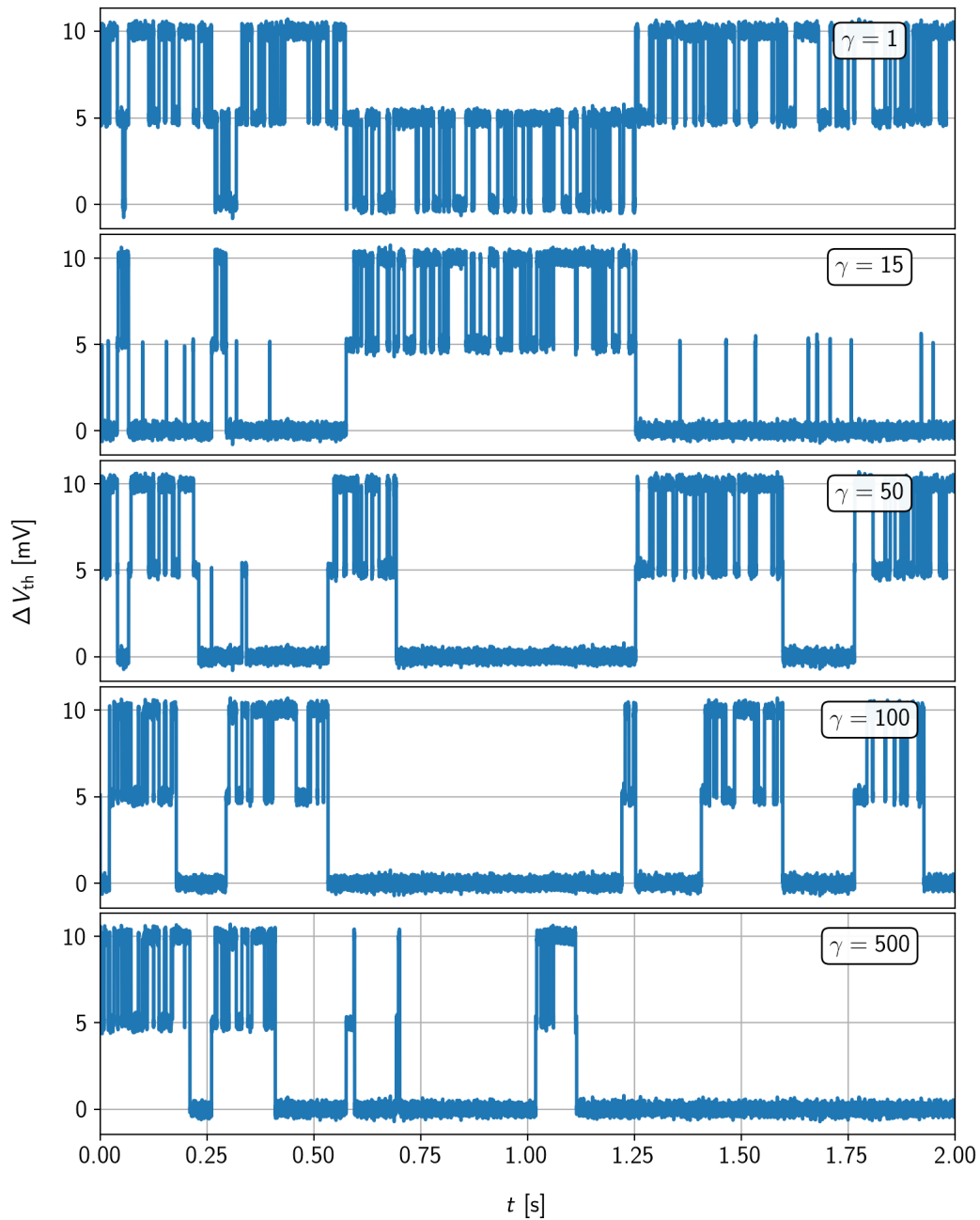


Figure 7.22: Simulation results for two coupled two-state defects with different coupling factors. The time constants were taken from [Figure 7.15](#) for a gate voltage of -1.95 V and a temperature of 250 K. It can be seen that a coupling factor of about $\gamma = 100$ is needed in order to reliably sample the modified RTN signal out of the signal.

the slow state captures and emits within the measurement window, it is going to be added as an emission of the fast state. This potentially results in a severe overestimation of the emission times of certain defects containing thermal transitions.

[Figure 7.23](#) shows the measurement data from 50 subsequent RTN measurements used to extract the time constants of traps ‘A’ and ‘B’ in [Section 7.2.2](#), merged together into one trace. It can be seen that the RTN signal becomes inactive from time to time, which is usually referred to as anomalous RTN [[108](#), [112](#), [113](#)].

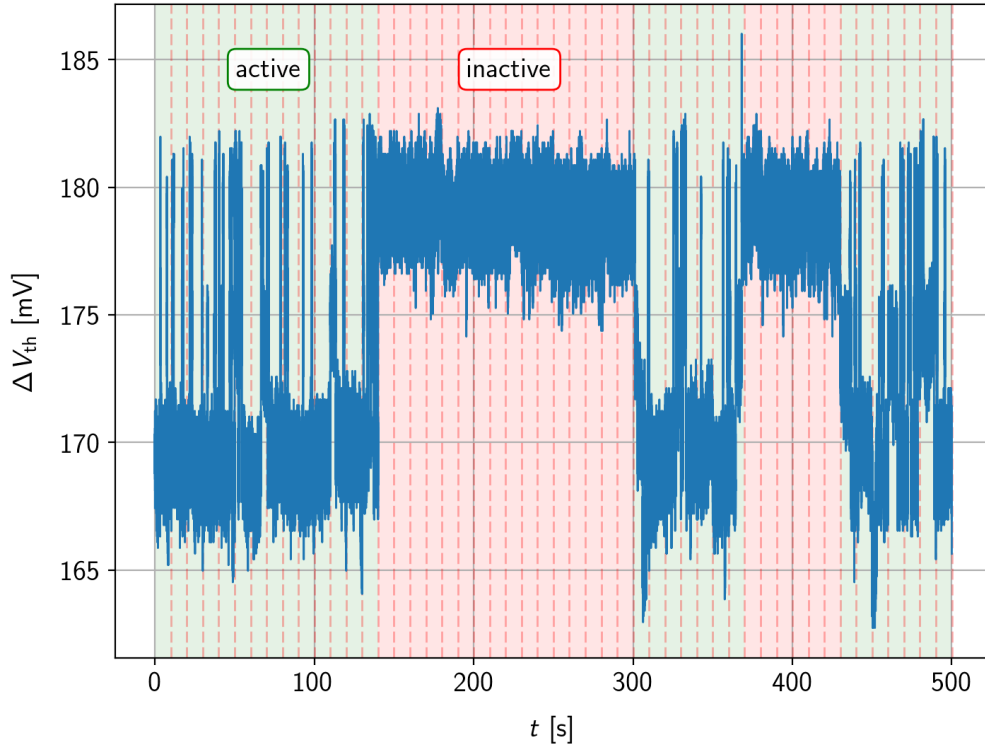


Figure 7.23: A merged measurement sequence producing anomalous RTN. Within the green labeled regions, the defects are active. The red regions mark some inactive state, which happens to have exactly the same level as the top level of the signal, ruling out any another uncorrelated slow defect. The red dashed lines mark the merging points. Note that due to the time needed for data transfer for one trace they represent a blind spot on the order of several seconds.

Naturally the question arises, on how to judge if a very slow emission event is just an unlucky sample of the fast state or rather should be considered as a separate thermal state. For this, the PDF of the exponential distribution given in (7.23) can be evaluated.

$$P(\tau) = \int_{\tau}^{\infty} \frac{1}{\bar{\tau}} \exp(-t/\bar{\tau}) dt \quad (7.23)$$

Figure 7.24 shows that for a trap with a mean emission time of 10 ms, the probability to observe 1 s emissions is $\sim 10^{-44}$, which is already practically zero. Note that in the recorded RTN traces, the observed inactive times are at least three orders of magnitude larger than the extracted emission time constants of the fast trap.

Another issue that can be seen in Figure 7.23 is that the inactive state always follows the captured state of trap 'B'. This has serious consequences on the hypothesis of having two independent but electrostatically coupled defects. If this were the case, defect 'A' would have to emit its charge again during the inactivity of defect 'B' which should be seen in the measurement signal as a transition to the level of defect 'A'. However, this is obviously not the case as the RTN signal stays inactive for more than 100 s at the highest level. This effectively rules out the case of two independent two-state defects, independently of the required coupling factors.

In the next step, the capture and emission times of the inactive state have to be determined. In order to be able to train the HMM, the Markov chain of the defect needs to be determined. Since the hypothesis of two coupled two-state traps was ruled out by the observation of an inactive state, it is much more likely that the observed signal is produced by a more complex defect. First

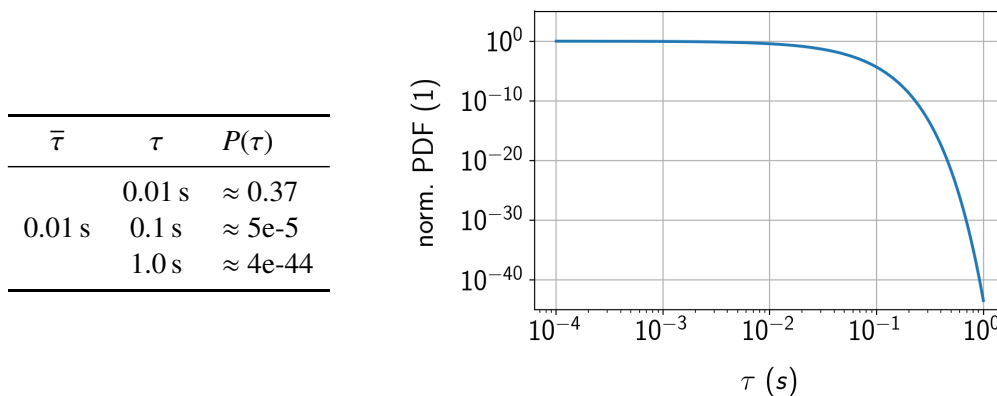


Figure 7.24: The right picture shows the PDF for an exponential distribution with a mean emission time $\bar{\tau} = 10$ ms. The table on the left gives the probabilities to observe emissions being a factor 10 or 100 larger. With a probability of about one in 200000, a 100 ms emission could still be seen as a (very) unlucky sample of the same distribution, an emission of 1 s from the same distribution is already practically impossible to happen in reality.

of all, the similar step heights strongly suggest that the defect is actually capable of capturing two electrons. Because the initial charge of the defect cannot be detected, the emitting states will be plotted as neutral in its ground state 1, negative in its slow state 2 and double-negative in its highest state 3 without any loss of generality. Secondly, since the inactive state always appears from the highest, double-negatively charged state, the thermal state is added as state 4, coming from state 3.

An exemplary picture of merged RTN traces useful for the selection of the Markov chain is shown in [Figure 7.25](#). The “regular” operating regime already covered in [Section 7.2.3](#) is labeled with 1 – 3, the inactive thermal state with 4. Another interesting finding is marked with 5, eventually representing another thermal state. There, after emitting an electron from state 3, another region of inactivity shows up. This state is seen explicitly only on a few occasions. This can possibly be explained by a significantly larger transition time from state 2 to state 5 as compared to the emission times from state 2 to the other two connected states. On the other hand, the observed time constants are approximately larger by only one order of magnitude compared to the emission times of trap ‘A’ extracted in [Figure 7.15](#), which still is in range of a very unlucky sample of the regular distribution (compare [Figure 7.24](#)).

The two most likely Markov chains are presented in the bottom part of [Figure 7.25](#). The thermal state(s) are added to the HMM as tied states, meaning that they possess the same charge state (i.e. no observable emissions). In the sense of the atomic defect structure, those states most likely resemble the same kind of structural relaxation as observed in the NMP four-state model from states 2’ to 2, see [Section 5.2.2](#).

On this occasion it should be stressed that a proper selection of the Markov chains is one of the most important tasks because HMM training does not include any physical reasoning. In other words, the HMM will always stick to the pre-selected defect structure no matter how unlikely a sequence of observations will be. Sometimes the MAP probability is used to determine the number of (two-state) defects [223]. This is however a problematic approach since more defects (i.e. more levels) tend to match the long-term drift and measurement noise better and thus will have a larger probability. The best solution to this problem is to compare the temperature and bias dependence of the individual defects or defect states and judge if the Markov training delivers physically reasonable results.

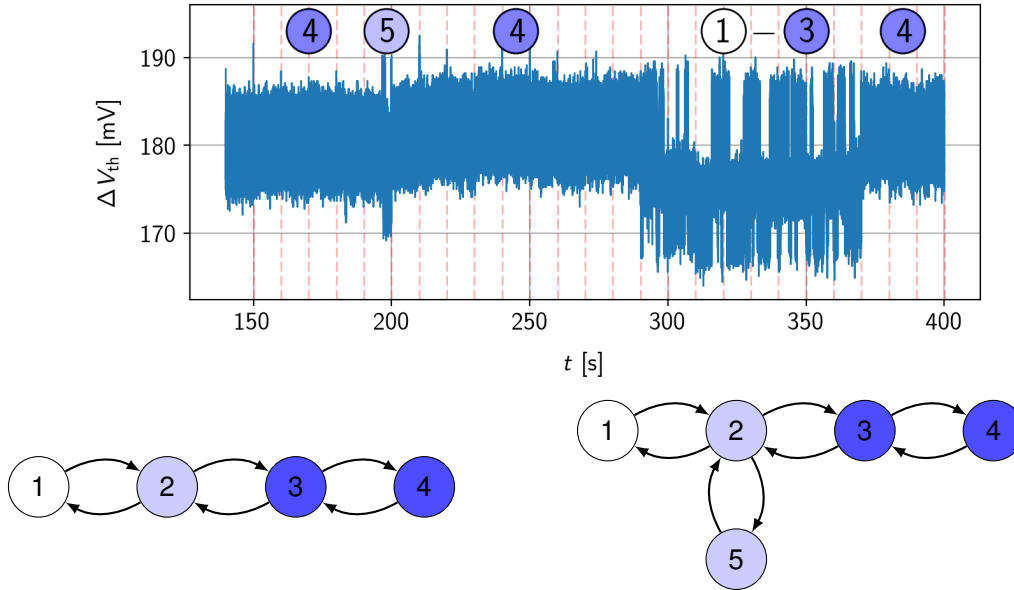


Figure 7.25: The bottom pictures show the two most likely Markov chains producing the measured RTN signals given in the upper picture. The states 1 and 2 produce the correlated RTN shown in [Figure 7.13](#). State 3 is a thermal state causing the inactive phases from state 2. Another thermal state 4 was added because of inactivities from the fast emissions for about 5 s observed at the same level as state 1.

The training of the HMM was performed with the original measurement data for each temperature and bias condition using the four and five state defect structures mentioned above. Because of the dependence of the Baum-Welch algorithm presented in [Section 6.7](#) on the initial parameters, the training was performed with 100 seeds using Gaussian distributions for the initial step-heights as well as the initial capture and emission times. The mean values of the Gaussian distributions for capture and emission times were taken from the results in [Figure 7.15](#).

For baseline correction, the LOWESS algorithm was chosen because of its versatility and robustness against different long-term drift patterns. After training, the results of the MAP probabilities of all data points and the corresponding step heights were examined for both defect configurations (see [Figures 7.26](#) and [7.27](#)). This was done primarily to check the accuracy of the extracted defect parameters for the different voltages. It is evident that the logarithmic probabilities decrease towards the regions where $\tau_c \gg \tau_e$ or vice versa. In the region of $\tau_c \ll \tau_e$ around -2.3 V, the defect is inactive most of the time and thus the number of capture and emission events is quite small, decreasing the overall accuracy. For the other region at around -1.7 V, the emission times are much larger than the capture times. This leads to a condition where the defect is rarely in its ground state which poses another problem on top of the reduced statistics, namely the baseline correction. To catch medium-frequency signal drifts within one trace, the algorithm also needs to be set to a sensitivity where the estimated baseline can change accordingly. This however tends to pull the baseline towards the regions where the defect is in a charged state.

The extracted step heights of the defects lie between 4 mV and 5 mV with a slightly lower average value for elevated temperatures which is in line with the values given in [Figure 7.16](#). They also possess a slight voltage dependence with decreasing values for increasing gate bias. This can probably be explained by the fact that for a stronger channel, the screening of the defect is increased and thus its individual impact on V_{th} decreases.

To judge which of the two defect candidates is more likely, the differences in the MAP probabilities can be plotted. If one of the defects shows a significantly higher probability across the gate

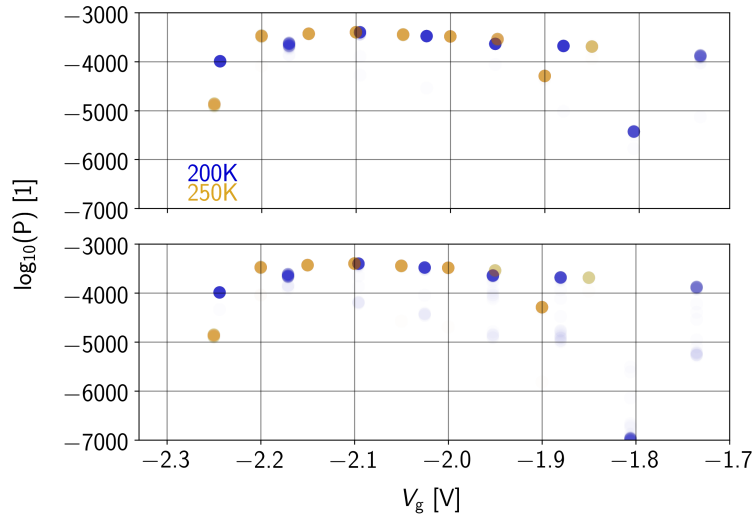


Figure 7.26: The resulting MAP probabilities for the five-state (upper) and the four-state (lower) defect configurations. Towards the border regions, the results diverge because of unfavorable duty cycles of the recorded signals and problems with the baseline estimation.

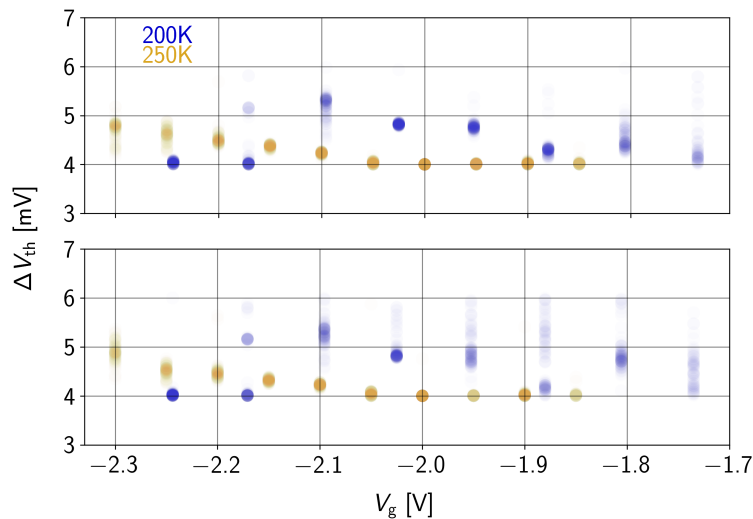


Figure 7.27: The extracted step heights for the five-state (upper) and the four-state (lower) defect configurations. A weak voltage dependence can be seen for both of the defect candidates, which can be explained by an increased screening of the defect with a stronger channel. The absolute values of ΔV_{th} as well as their temperature dependence are in the same range like the independently obtained values in [Figure 7.16](#), confirming the validity of the extracted parameters.

voltages, this would be a strong argument in favor of the respective configuration. In [Figure 7.28](#) the differences of the probabilities and step-heights of the five-state defect and the four-state defect are plotted. A positive value of ΔP thereby means that the solution for five-state defect is more likely. The differential probabilities do not show a clear trend regarding which of the two candidates is more likely to produce the measured signals although on average the five-state defect has slightly higher probabilities across many gate biases. This could also be a consequence of the additional parameter for the HMM die to the extra state of the five-state defect. Additionally, the step height differences are pretty much centered around zero, not favoring any of the candidates. Different results in the step heights would point to a badly aligned baseline or a

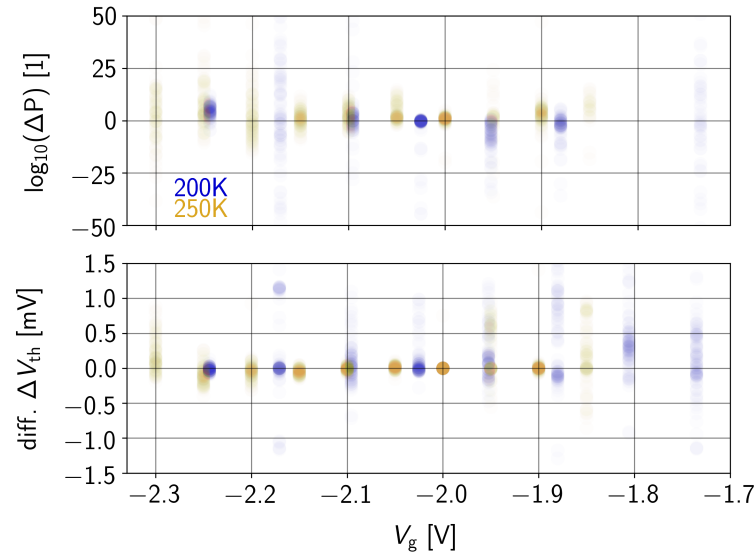


Figure 7.28: The differences of the MAP probabilities and the step heights between the five-state defect and the four-state defect from Figure 7.25. A positive value of ΔP thereby means that the solution for the five-state defect is more likely. Differences in the step heights would point to a badly aligned baseline or a significant amount of missed steps. Neither of the two results show a clear tendency in favor of one of the defects, making the judgment on which of them is more likely difficult.

significant amount of missed steps.

Even after careful selection of the two defect candidates and evaluating their probabilities to produce the observed measurements, none of the two could be excluded so far. The last possibility left to prefer one over the other is thus the extracted time constants. The average capture and emission times were calculated from the 100 seeds using weighted averages. For the weight factors, the inverse MAP probabilities of the seeds were chosen in order to give less weight to more unlikely results. At that point, it has to be noted that the weights are likely to overestimate unlikely training results as they were calculated using logarithmic probabilities. If on the other hand the actual probabilities had been used, most likely only a few of the seeds with the highest probability would have defined the result because the probabilities can differ by several orders of magnitude.

The extracted time constants for the four and five-state defects are plotted separately for each of the states in Figure 7.29, with the darker color being the emission time and the lighter one being the capture time of the state. The only difference between the two defects are the transitions between states 2 and 5, hence only the transitions τ_{21} and τ_{23} are affected by the additional state. Judging by the bias dependence of the extracted time constants, the five-state defect possesses a smoother, nearly exponential behavior which is typical for RTN defects. On the other hand, the transitions τ_{25} and τ_{52} show a significant voltage dependence for 200 K, which normally should not be the case for thermal transitions.

As mentioned before, the measurements show only few explicit transitions pointing to a (hypothetical) state 5. The observed voltage dependence thus could also be caused by an insufficient number of transitions, backed by the rather large error bars for this state in Figure 7.29. Overall, a clear tendency towards one of the proposed defect structures cannot be seen in the results. The differential probabilities in Figure 7.28 and the bias dependencies, however, make a five-state defect structure look a bit more likely.

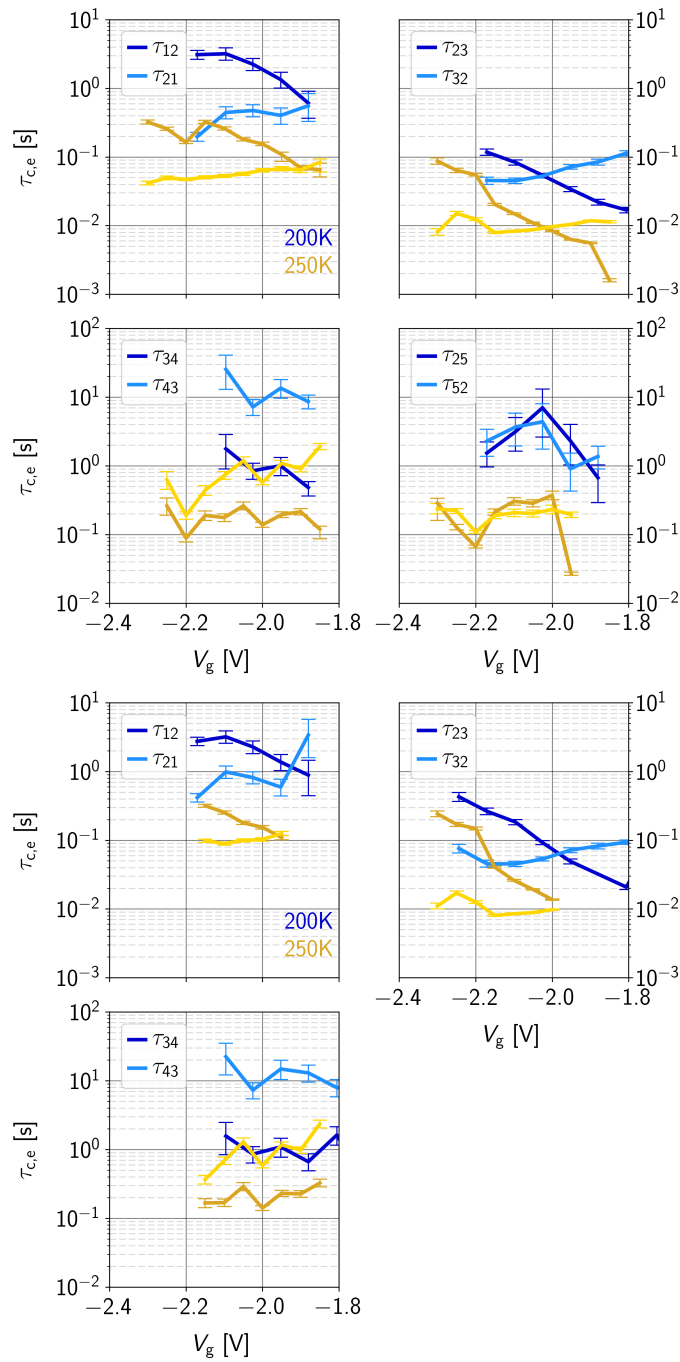


Figure 7.29: The time constants of the five-state (top) and four-state (bottom) defects proposed in [Figure 7.25](#) obtained by HMM training. The smoother, nearly exponential behavior of the emitting states put the five-state structure slightly in favor of the other. The clear bias dependence of the thermal state 5 for 200 K can possibly be explained by a lack of observed transitions to this state.

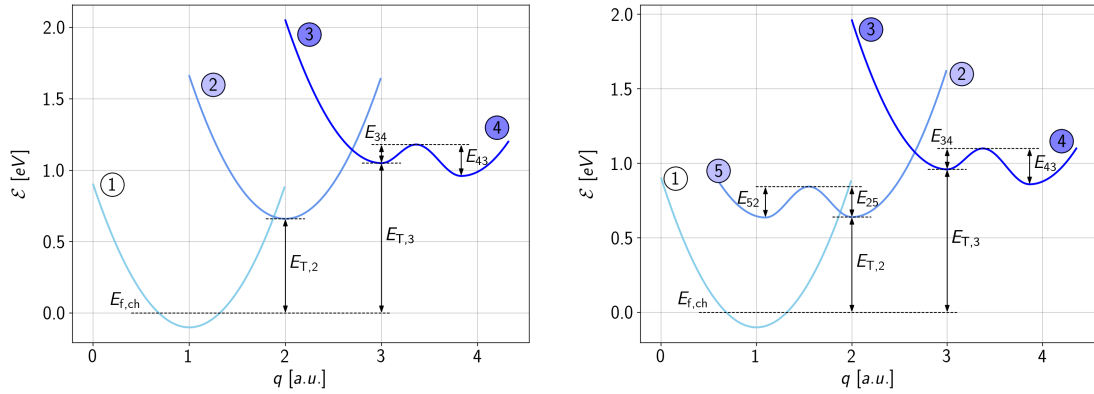


Figure 7.30: Schematic configuration coordinate diagrams for the proposed four-state (left) and five-state (right) defect structures. The extracted barriers and energy levels were taken from Tables 7.5 and 7.5.

It should be noted at that point, that the noise-level and the associated long-term drift for the measurements at 275 K was too large for an extraction with the presented HMM library. An inspection of the resulting time constants showed almost equal capture and emission times for all states across the whole bias range. This suggests accidental fitting of measurement noise due to either a bad baseline estimation or too much noise. The MAP paths indeed revealed a lot of wrongly asserted emissions mostly due to long-term drift of the signal, and thus the results for 275 K were discarded.

Finally, a slightly modified defect parameter extraction for the trap position and the trap levels as shown in Section 7.2.3 can be done with the two defect candidates shown in Figure 7.25. One difference is that the voltage dependence of the trap level of the double-negatively charged state 3 is twice as high compared to that of state 2. Consequently, the slopes of the capture and emission barriers in (7.6) also need to be multiplied by two. The other difference are the thermal barriers to the states 4 and 5 not present in the initial extraction. Those were calculated separately for capture and emission by a simple Arrhenius law for both of the states. Schematic CC diagrams for the two defect candidates are shown in Figure 7.30.

The results of this extraction can be seen in Table 7.4 and 7.5. Not very surprisingly, state 2 closely matches defect ‘A’ in Table 7.1. The main difference here is the larger values for the defect positions, which are caused by slightly different intersection points of the calculated time constants. The values for defect ‘B’ cannot be compared directly for two reasons. First, the defect position has to be the same which is still backed by the similar values of ΔV_{th} in the RTN signal. Additionally, the double-negatively charged state 3 by definition has twice the voltage dependence of state 2, which forces the extracted trap levels to be different.

7.2.6 Conclusions

In this section, single defect parameters at different cryostatic temperatures were extracted from RTN measurements on a GaN/AlGaIn fin MIS-HEMT.

In Section 7.2.2, first the differences between regular fin FETs and the devices used in this work are laid out. The transfer characteristics of the measured device recorded at different temperatures are then used to calibrate the electrostatic device simulations.

The characteristic time constants of two pairs of coupled RTN producing defects are extracted in Section 7.2.3 with the spectral maps method introduced in Section 6.6.3. Additionally, important

Trap	T (K)	y_{trap} (nm)	$E_{T,2}$ (eV)	$E_{T,3}$ (eV)
4-state	200	7.7	0.69	1.01
	250	6.6	0.63	1.12
5-state	200	6.9	0.63	0.91
	250	7.1	0.65	1.06

Table 7.4: The defect positions and trap levels for the two proposed defect candidates. Note that due to its double-negative charge, the voltage dependence of state 3 is twice as high as the one of state 2. Consequently, the slopes of the capture and emission barriers in (7.6) also need to be multiplied by two for the extraction of $E_{T,3}$.

Trap	k_{12} (s^{-1})	$E_{R,12}$ (eV)	k_{23} (s^{-1})	$E_{R,23}$ (eV)	k_{34} / k_{43} (s^{-1})	E_{34} / E_{43} (eV)	k_{25} / k_{52} (s^{-1})	E_{25} / E_{52} (eV)
4-state	1.7×10^5	0.85	5.5×10^5	0.72	1.6×10^3	0.13	–	–
					2.4×10^4	0.22	–	–
5-state	8.3×10^4	0.75	9.0×10^5	0.59	4.7×10^3	0.14	6.1×10^4	0.20
					1.1×10^5	0.20	7.7×10^4	0.21

Table 7.5: The temperature related parameters from the defects. The relaxation energies $E_{R,12}$ and $E_{R,23}$ and the corresponding pre-factors are calculated according to (7.11). The thermal barriers for capture and emission are extracted from the Arrhenius law.

parameters like the trap level, the vertical position and the relaxation energy of the defects are extracted in Table 7.1 using a two-state NMP model.

The question if the observed signals emerge from two coupled pairs of two-state defects or a single, more complex defect structure are tried to be answered by evaluating the necessary coupling factors from RTN simulations and comparing them to those calculated for a chosen defect candidate using different methods.

In Section 7.2.4, three different approaches to estimate the short-range potential perturbation of one defect capturing a charge were explored. Based on the extracted trap level in Section 7.2.3, the most likely defect candidate was identified to be the nitrogen vacancy. Theoretical coupling factors were extracted in a worst-case sense, namely for the neighboring and the second to next neighbor nitrogen vacancies. On the other hand, HMM simulations were used to identify the required coupling factors to observe the measured coupled RTN signals. The required coupling factor of about 100 is at the upper limit of the theoretically extracted data for the nearest neighbor defect. The uncertainties in the extracted potentials, however, are quite large and enter the coupling factors exponentially. Quite interestingly, the coupling factors reported in literature closely match the values in Table 7.3 at room temperature for defects being 1 nm apart from each other.

Finally, two possible defect structures, a four-state and a five-state defect, are investigated by extracting their characteristic time constants using HMM training in Section 7.2.5. Both defect structures are compared to each other in terms of their MAP probabilities and the extracted defect distributions to find the most likely candidate. As none of them can be discarded within a reasonable likelihood, a slightly modified version of the parameter extraction introduced in Section 7.2.3 is carried out for both defect structures (see Table 7.4 and 7.5).

Chapter 8

Conclusions and Outlook

The first section of this chapter collects the main findings across this thesis into a single section, where they are briefly summarized and brought into context. Based on these findings, the second section lists some ideas for future directions of research.

8.1 Conclusions

The main topics investigated in this thesis were charge feedback effects on the V_{th} drift of GaN/AlGaN MIS-HEMTs at forward gate-bias stress, different methods for the extraction of the characteristic time-constants, and the extraction of single-defect parameters from nano-scale GaN/AlGaN fin-MIS-HEMTs.

To investigate charge feedback effects on large-area devices, a simulation study on GaN/AlGaN MIS-HEMTs was performed using the NMP four-state model. The effects of different feedback mechanisms on the experimentally observed capture and emission times were estimated, which led to the following conclusions:

- Transient changes in the surface potential caused by the charge feedback of the defects lead to increased defect levels and a decreased active energy area seen by the defects.
- It is not just the amount of trapped charges which changes due to charge feedback, but also the kinetics of charge capture and emission events.
- The different kinetics are caused by the change of the trap levels and the active energy area, but also by local potential perturbations influencing the characteristic time-constants of neighbouring defects.
- Due to these effects, common modeling approaches like the usage of accumulated stress times cannot be used as the observed defect kinetics are a strong function of previous stress and recovery cycles in the device. Thus, for such of devices, it is of utmost importance to use transient simulations following the experimental bias conditions as closely as possible.

To focus on the microscopic properties of individual defects, a theoretical investigation of Markov processes and the stochastic nature of charge emissions was used to put forward two novel methods to extract the characteristic capture and emission times from RTN signals:

- One method is based on spectral maps and is insensitive to long-term drift of the signal, suited best for basic RTN signals with high to medium SNR.
- The other method is based on a modified HMM and capable of handling more complex RTN signals coming from multiple defects. It can handle significantly more noisy signals and was implemented as a Python library which is able to extract the characteristic time

constants of a system of defects by fitting the model to a set of measurement data. Furthermore, the library allows simulating the stochastic RTN emissions of a given system of defects.

Both methods were used to identify the bias-dependent capture and emission times of RTN producing defects in nano-scale GaN/AlGaIn fin-MIS-HEMTs.

- The spectral method was used to calculate single-defect parameters like trap levels, vertical defect positions, and the apparent activation energy of two sets of coupled defects.
- The hypothesis of having a pair of coupled defects or a more complex defect structure was checked by estimating the required coupling factors of the defects from HMM simulations and comparing them to the factors calculated from a shielded Coulomb potential.
- Finally, two alternative defect structures were deduced from careful examination of the merged RTN traces. Their characteristic time constants were extracted from HMM training and compared to each other. As none of the two could be discarded with absolute certainty, a slightly modified parameter extraction to obtain their vertical positions, trap levels and thermal activation energies was carried out for both of the candidates.

This work contributed to a more profound physical understanding of the defects responsible for BTI in GaN/AlGaIn MIS-HEMTs. For large-area devices, the importance of charge feedback effects on a proper interpretation of the observed time constants in BTI measurements in GaN technology was highlighted. Furthermore, two innovative methods to obtain the stochastic capture and emission time constants from defects were introduced in this work. With these methods, for the first time, the microscopic structure, vertical positions, energy levels, and temperature activation of RTN producing defects could be extracted for GaN.

In general, the presented methods for the extraction of the characteristic time constants are formulated universally enough to be useful for single-defect investigations in many different semiconductor technologies. This is mainly justified by the fact that neither the spectral method nor the HMM depends on any physical defect model except the Markov property (i.e., being a memoryless system).

8.2 Outlook

The identification of the microscopic defects responsible for charge trapping will be a major key to find recipes to improve the reliability of future GaN technology. The results derived throughout this work should thus be seen as a basis for further studies in the field of the reliability of GaN technology. The following paragraphs list some ideas for future research based on the findings above.

Identification of Surface Donors

Although several types of bulk defects affecting the reliability of GaN have been found by measurements and are widely accepted throughout the community, one of the most crucial defect(s) necessary for HEMT devices, the surface donors, remain largely unknown. To identify the structure and the origin of these defects, future studies should combine the extraction of defect bands from large area devices with single-defect experiments and first-principle simulations.

From a technological point of view, this requires the fabrication of large-area devices together with nano-scale devices in the same process to ensure compatibility. Furthermore, the device

layout should be kept as simple as possible (i.e., planar devices) to simplify the calibration of the device in TCAD simulations. For large-area devices, the feedback of charges captured in defects plays a crucial role. This should be taken into account already when designing eMSM measurements, but also in device simulations, as the observed degradation will be a function of the stress history of the device.

As the devices usually suffer from large instabilities already at nominal operating conditions, measurements at cryostatic temperatures allow to reduce thermal noise and trigger less response of defects. For single-defect measurements, the barrier layer should be designed as thin as possible in order to raise the average step-heights of defects at the interface to a level which is detectable by the equipment. The usage of constant-current setups can also help to detect defects with smaller step-heights as the full measurement resolution is available across the whole current range. Alternatively, MIS-HEMTs with small oxide thicknesses could be used to monitor the gate current which in turn allows defect modeling based on the observed leakage currents.

The impact of the barrier on the observed time-constants of the surface donors is another topic, which is widely neglected in reliability investigations of GaN HEMTs up to date. Separating the influence of the barrier from native defect properties, however, could be one of the keys to identify the origin and structure of the surface donors.

Hidden Markov Model Library

The HMM library was so far only tested on real measurement data from GaN technology. Although it should be independent of the technology the data was recorded on, a broader set of tests should help to reveal eventual bugs and further improve the general robustness of the algorithms.

The first and most valuable improvement would be the implementation of *finite* HMMs, which would also enable the processing of TDDS data. Currently, the HMM does not consider an explicit state which ends a sequence. In the case of TDDS data, such an explicit end-state would naturally be given when all defects emitted their charge. The average occupancies of the defects after stress can be directly obtained from the starting probabilities of the HMM after training.

Another major improvement would be the implementation of a *factorial* HMM, which limits the size of the system from being the factor of the states to the addition of the states of the underlying defects. A state space of a system of three three-state defects thus would be reduced from $3^3 = 27$ to 9. This would allow to use a larger number of defects at the expense of complicating the addition of thermal states.

One obvious improvement is related to the training speed of the library. In the current state, about half of the running time is dedicated the baseline estimation which cannot be processed in parallel yet. A clever parallelization algorithm of this part of the library or faster baseline estimation algorithms would dramatically improve run-times. Moreover, the baseline estimation algorithms themselves need to be optimized further in terms of both, their speed and robustness against different measurements.

Up to date, the Baum-Welch algorithm to find the sequence of states is independent of the baseline estimation algorithm. Combining these two parts into a single expectation maximization algorithm for the combined system *including* the baseline could help to improve both, the convergence and the quality of the training results. Further, this algorithm could also be designed to add other constraints to the Baum-Welch algorithm which prohibit the necessity of splitting up and re-assembling the system of defects at every iteration. A more formal inclusion of these constraints could also help to judge on their impact on the global training results, which is largely unknown up to this point.

Appendix A

The Hidden Markov Model Library

The HMM library consists of two main class objects, **Trap** and another class derived from it called **Traps**. The first one has three mandatory arguments, the name of the defect, its step height ΔV_{th} and the number of emitting states. Depending on the number of states, keyword arguments for the forward and backward rates for all states have to be provided in Hz. States without emission (i.e. thermal states, see Section 5.2.2) is defined with the method `addTiedState`. Additionally, limits for ΔV_{th} and the rates, the initial voltage offset as well as a list with the corresponding charge of each state can be set optionally.

The **Traps** class takes a list of traps, from which it calculates the combined system using the data provided by the **Trap** objects in the list. The probably most important method used to train the model is `fit`, which takes a sequence of observations and uses a modified Baum-Welch algorithm to fit the individual defects independently of each other. A simplified flow chart of the training algorithm is given in [Figure A.1](#).

It should be noted that the implementations of the Viterbi algorithm, the forward and backward algorithm and some other methods are provided by the pomegranate HMM library [224]. One example of the training of a two-defect system shows the simple usage of the provided classes:

```

from hmm import Trap,Traps

# reading and processing of measurement data
<...>

two = TwoState(name='twostate', dVth=5.0,
               dVthRange=(4., 6.), k12=1e3, k21=1e2)
four = Trap(name='fourstate', nrStates=3,
            dVth=3.0, charge = [0,1,2],
            k12=0.1, k21=0.5, k23=10.0,
            k32=3.0)
# add thermal state to state 2 with tau_c=2s, tau_e=10s
four.addTiedState(2, 1./2, 1./10)

defects = Traps([two,four], sigma=0.5)

# train the HMM
defects.fit([data1,data2,data3], update=1e-5,
           iterations=(5, 100), jobs=threads)

# plot results
print defects
defects.plot()

```

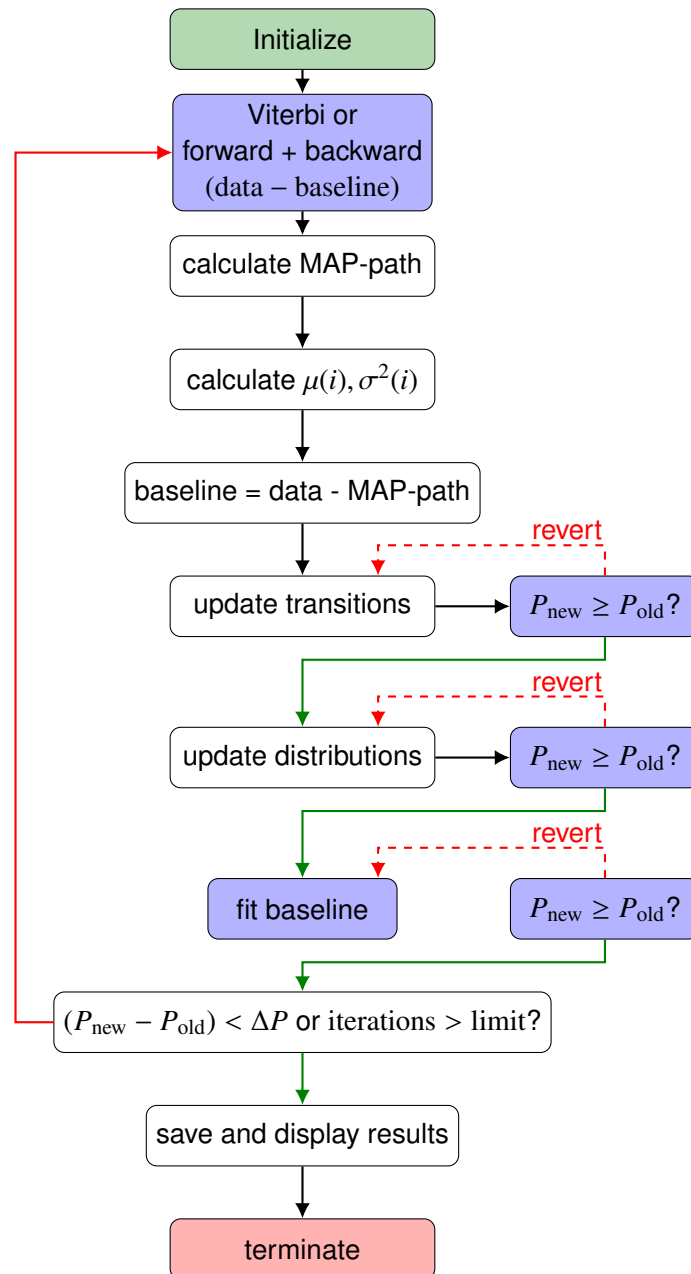


Figure A.1: A simplified flowchart of the modified Baum-Welch algorithm used for training of the HMM library. The blue rectangles mark the parts of the algorithm which are processed in parallel. The implementations of the Viterbi algorithm, the forward and backward algorithm and some other methods are provided by the pomegranate HMM library [224].

Bibliography

- [1] International Energy Agency. *Key World Energy Statistics 2017*. 2017.
- [2] D. Ueda. “Properties and Advantages of Gallium Nitride”. In: *Power GaN Devices: Materials, Applications and Reliability*. Ed. by M. Meneghini, G. Meneghesso, and E. Zanoni. Springer International Publishing, 2017, pp. 1–26.
- [3] B. J. Baliga. “Semiconductors for high-voltage, vertical channel field-effect transistors”. In: *Journal of Applied Physics* 53.3 (Mar. 1982), pp. 1759–1764.
- [4] B. J. Baliga. *Fundamentals of power semiconductor devices*. Springer Science & Business Media, 2010.
- [5] E. Johnson. “Physical limitations on frequency and power parameters of transistors”. In: *1958 IRE International Convention Record*. Vol. 13. Mar. 1965, pp. 27–34.
- [6] P. W. Lagger. *Physics and characterization of the gate stack in gallium nitride based MIS-HEMTs*. 2014.
- [7] H. Morkoç. *Handbook of Nitride Semiconductors and Devices*. Wiley-VCH Verlag GmbH & Co. KGaA, Mar. 2008.
- [8] R. Quay. *Gallium nitride electronics*. Vol. 96. Springer Science & Business Media, 2008.
- [9] T. P. Chow, I. Omura, M. Higashiwaki, H. Kwarada, and V. Pala. “Smart Power Devices and ICs Using GaAs and Wide and Extreme Bandgap Semiconductors”. In: *IEEE Transactions on Electron Devices* 64.3 (Mar. 2017), pp. 856–873.
- [10] O. Ambacher. “Growth and applications of Group III-nitrides”. In: *Journal of Physics D: Applied Physics* 31.20 (1998), p. 2653.
- [11] S. Keller. “Substrates and Materials”. In: *Power GaN Devices: Materials, Applications and Reliability*. Ed. by M. Meneghini, G. Meneghesso, and E. Zanoni. Springer International Publishing, 2017, pp. 27–52.
- [12] T. Lei, M. Fanciulli, R. J. Molnar, T. D. Moustakas, R. J. Graham, and J. Scanlon. “Epitaxial growth of zinc blende and wurtzitic gallium nitride thin films on (001) silicon”. In: *Applied Physics Letters* 59.8 (1991), pp. 944–946.
- [13] M. J. Paisley, Z. Sitar, J. B. Posthill, and R. F. Davis. “Growth of cubic phase gallium nitride by modified molecular-beam epitaxy”. In: *Journal of Vacuum Science & Technology A* 7.3 (1989), pp. 701–705.
- [14] M. Mizuta, S. Fujieda, Y. Matsumoto, and T. Kawamura. “Low Temperature Growth of GaN and AlN on GaAs Utilizing Metalorganics and Hydrazine”. In: *Japanese Journal of Applied Physics* 25.Part 2, No. 12 (Dec. 1986), pp. L945–L948.
- [15] Solid State. <https://de.wikipedia.org/wiki/Wurtzit>. 2008.
- [16] N.-Q. Zhang, B. Moran, S. DenBaars, U. Mishra, X. Wang, and T. Ma. “Effects of surface traps on breakdown voltage and switching speed of GaN power switching HEMTs”. In: *International Electron Devices Meeting. Technical Digest (Cat. No.01CH37224)*. IEEE, Dec. 2001, pp. 25.5.1–25.5.4.
- [17] A. Krost and A. Dadgar. “GaN-based optoelectronics on silicon substrates”. In: *Materials Science and Engineering: B* 93.1-3 (May 2002), pp. 77–84.
- [18] A. Dadgar, C. Hums, A. Diez, F. Schulze, J. Bläsing, and A. Krost. “Epitaxy of GaN LEDs on large substrates: Si or sapphire?” In: *Proc. SPIE 6355, Advanced LEDs for*

- Solid State Lighting*. Ed. by C.-H. Hong, T. Taguchi, J. Han, and L. Chen. Sept. 2006, 63550R.
- [19] J. J. Xu, Y.-F. Wu, S. Keller, S. Heikman, B. J. Thibeault, U. K. Mishra, and R. A. York. “1 - 8-GHz GaN-based power amplifier using flip-chip bonding”. In: *IEEE Microwave and Guided Wave Letters* 9.7 (July 1999), pp. 277–279.
- [20] T. Paskova and K. R. Evans. “GaN Substrates - Progress, Status, and Prospects”. In: *IEEE Journal of Selected Topics in Quantum Electronics* 15.4 (July 2009), pp. 1041–1052.
- [21] *Minimos-NT User Manual - Release 2018.04*.
- [22] S. Vitanov. “Simulation of high electron mobility transistors”. PhD thesis. Institute for Microelektronics, TU Wien, 2010.
- [23] W. Soluch, E. Brzozowski, M. Lysakowska, and J. Sadura. “Determination of mass density, dielectric, elastic, and piezoelectric constants of bulk GaN crystal”. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 58.11 (Nov. 2011), pp. 2469–2474.
- [24] D. M. Caughey and R. E. Thomas. “Carrier mobilities in silicon empirically related to doping and field”. In: *Proceedings of the IEEE* 55.12 (Dec. 1967), pp. 2192–2193.
- [25] F. Schwierz. “An electron mobility model for wurtzite GaN”. In: *Solid-State Electronics* 49.6 (June 2005), pp. 889–895.
- [26] V. Palankovski. “Simulation of Heterojunction Bipolar Transistors”. PhD thesis. Institute for Microelektronics, TU Wien, 2000.
- [27] O. Ambacher, J. Majewski, C. Miskys, A. Link, M. Hermann, M. Eickhoff, M. Stutzmann, F. Bernardini, V. Fiorentini, V. Tilak, B. Schaff, and L. F. Eastman. “Piezoelectric properties of Al(In)GaN/GaN hetero- and quantum well structures”. In: *Journal of Physics: Condensed Matter* 14.13 (2002), p. 3399.
- [28] A. Zoroddu, F. Bernardini, P. Ruggerone, and V. Fiorentini. “First-principles prediction of structure, energetics, formation enthalpy, elastic constants, polarization, and piezoelectric constants of AlN, GaN, and InN: Comparison of local and gradient-corrected density-functional theory”. In: *Physical Review B - Condensed Matter and Materials Physics* 64.4 (2001), pp. 1–6.
- [29] S. F. Chichibu and S. Nakamura. *Introduction to nitride semiconductor blue lasers and light emitting diodes*. CRC Press, 2014.
- [30] M. A. Khan, J. M. Van Hove, J. N. Kuznia, and D. T. Olson. “High electron mobility GaN/Al_xGa_{1-x}N heterostructures grown by low-pressure metalorganic chemical vapor deposition”. In: *Applied Physics Letters* 58.21 (May 1991), pp. 2408–2410.
- [31] M. Asif Khan, A. Bhattarai, J. N. Kuznia, and D. T. Olson. “High electron mobility transistor based on a GaN-Al_xGa_{1-x}N heterojunction”. In: *Applied Physics Letters* 63.9 (Aug. 1993), pp. 1214–1215.
- [32] J. P. Ibbetson, P. T. Fini, K. D. Ness, S. P. DenBaars, J. S. Speck, and U. K. Mishra. “Polarization effects, surface states, and the source of electrons in AlGaIn/GaN heterostructure field effect transistors”. In: *Applied Physics Letters* 77.2 (July 2000), pp. 250–252.
- [33] O. Ambacher, J. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Murphy, W. J. Schaff, L. F. Eastman, R. Dimitrov, L. Wittmer, M. Stutzmann, W. Rieger, and J. Hilsenbeck. “Two-dimensional electron gases induced by spontaneous and piezoelectric polarization charges in N- and Ga-face AlGaIn/GaN heterostructures”. In: *Journal of Applied Physics* 85.6 (Mar. 1999), pp. 3222–3233.
- [34] M. A. Khan, J. N. Kuznia, D. T. Olson, W. J. Schaff, J. W. Burm, and M. S. Shur. “Microwave performance of a 0.25 μm gate AlGaIn/GaN heterostructure field effect transistor”. In: *Applied Physics Letters* 65.9 (1994), pp. 1121–1123.

- [35] Y. F. Wu, B. P. Keller, S. Keller, D. Kapolnek, S. P. Denbaars, and U. K. Mishra. "Measured microwave power performance of AlGa_N/Ga_N MODFET". In: *IEEE Electron Device Letters* 17.9 (Sept. 1996), pp. 455–457.
- [36] B. M. Green, K. K. Chu, E. M. Chumbes, J. A. Smart, J. R. Shealy, and L. F. Eastman. "The effect of surface passivation on the microwave characteristics of undoped AlGa_N/Ga_N HEMTs". In: *IEEE Electron Device Letters* 21.6 (June 2000), pp. 268–270.
- [37] A. Vertiatchikh, L. F. Eastman, W. J. Schaff, and I. Prunty. "Effect of surface passivation of AlGa_N/Ga_N heterostructure field-effect transistor". In: *Electronics Letters* 38.8 (Apr. 2002), pp. 388–389.
- [38] G. Koley, V. Tilak, L. F. Eastman, and M. G. Spencer. "Slow transients observed in AlGa_N/Ga_N HFETs: effects of SiN_x passivation and UV illumination". In: *IEEE Transactions on Electron Devices* 50.4 (Apr. 2003), pp. 886–893.
- [39] Y. Ando, Y. Okamoto, H. Miyamoto, T. Nakayama, T. Inoue, and M. Kuzuhara. "10-W/mm AlGa_N-Ga_N HFET with a field modulating plate". In: *IEEE Electron Device Letters* 24.5 (May 2003), pp. 289–291.
- [40] Y. F. Wu, A. Saxler, M. Moore, R. P. Smith, S. Sheppard, P. M. Chavarkar, T. Wisleder, U. K. Mishra, and P. Parikh. "30-W/mm Ga_N HEMTs by field plate optimization". In: *IEEE Electron Device Letters* 25.3 (Mar. 2004), pp. 117–119.
- [41] H. Zhang, E. J. Miller, and E. T. Yu. "Analysis of leakage current mechanisms in Schottky contacts to Ga_N and Al_{0.25}Ga_{0.75}N/Ga_N grown by molecular-beam epitaxy". In: *Journal of Applied Physics* 99.2 (Jan. 2006), p. 023703.
- [42] S. Ganguly, A. Konar, Z. Hu, H. Xing, and D. Jena. "Polarization effects on gate leakage in InAlN/AlN/Ga_N high-electron-mobility transistors". In: *Applied Physics Letters* 101.25 (Dec. 2012), p. 253519.
- [43] M. A. Khan, G. Simin, J. Yang, J. Zhang, A. Koudymov, M. S. Shur, R. Gaska, X. Hu, and A. Tarakji. "Insulating gate III-N heterostructure field-effect transistors for high-power microwave and switching applications". In: *IEEE Transactions on Microwave Theory and Techniques* 51.2 (Feb. 2003), pp. 624–633.
- [44] R. Stradiotto. "Characterization of Electrically Active Defects at III-N/Dielectric Interfaces". PhD thesis. Institute for Microelectronics, TU Wien, 2016.
- [45] W. Saito, Y. Takada, M. Kuraguchi, K. Tsuda, and I. Omura. "Recessed-gate structure approach toward normally off high-voltage AlGa_N/Ga_N HEMT for power electronics applications". In: *IEEE Transactions on Electron Devices* 53.2 (Feb. 2006), pp. 356–362.
- [46] T. Oka and T. Nozawa. "AlGa_N/Ga_N Recessed MIS-Gate HFET With High-Threshold-Voltage Normally-Off Operation for Power Electronics Applications". In: *IEEE Electron Device Letters* 29.7 (July 2008), pp. 668–670.
- [47] R. Chu, A. Corrion, M. Chen, R. Li, D. Wong, D. Zehnder, B. Hughes, and K. Boutros. "1200-V Normally Off Ga_N-on-Si Field-Effect Transistors With Low Dynamic on-Resistance". In: *IEEE Electron Device Letters* 32.5 (May 2011), pp. 632–634.
- [48] C. Liu, S. Yang, S. Liu, Z. Tang, H. Wang, Q. Jiang, and K. J. Chen. "Thermally Stable Enhancement-Mode Ga_N Metal-Isolator-Semiconductor High-Electron-Mobility Transistor With Partially Recessed Fluorine-Implanted Barrier". In: *IEEE Electron Device Letters* 36.4 (Apr. 2015), pp. 318–320.
- [49] Y. Cai, Y. Zhou, K. J. Chen, and K. M. Lau. "High-performance enhancement-mode AlGa_N/Ga_N HEMTs using fluoride-based plasma treatment". In: *IEEE Electron Device Letters* 26.7 (July 2005), pp. 435–437.

- [50] S. Jia, Y. Cai, D. Wang, B. Zhang, K. M. Lau, and K. J. Chen. "Enhancement-mode AlGa_N/Ga_N HEMTs on silicon substrate". In: *physica status solidi (c)* 3.6 (June 2006), pp. 2368–2372.
- [51] K. J. Chen and C. Zhou. "Enhancement-mode AlGa_N/Ga_N HEMT and MIS-HEMT technology". In: *physica status solidi (a)* 208.2 (Feb. 2011), pp. 434–438.
- [52] Y. Uemoto, M. Hikita, H. Ueno, H. Matsuo, H. Ishida, M. Yanagihara, T. Ueda, T. Tanaka, and D. Ueda. "Gate Injection Transistor (GIT) - A Normally-Off AlGa_N/Ga_N Power Transistor Using Conductivity Modulation". In: *IEEE Transactions on Electron Devices* 54.12 (Dec. 2007), pp. 3393–3399.
- [53] R. Kajitani, K. Tanaka, M. Ogawa, H. Ishida, M. Ishida, and T. Ueda. "Novel high-current density Ga_N-based normally off transistor with tensile-strained quaternary In-AlGa_N barrier". In: *Japanese Journal of Applied Physics* 54.4S (2015), 04DF09.
- [54] L. Romano. *Defects in Semiconductors*. Elsevier Science, 2015.
- [55] I. Rossetto, D. Bisi, C. de Santi, A. Stocco, G. Meneghesso, E. Zanoni, and M. Meneghini. "Performance-Limiting Traps in Ga_N-Based HEMTs: From Native Defects to Common Impurities". In: *Power Ga_N Devices: Materials, Applications and Reliability*. Ed. by M. Meneghini, G. Meneghesso, and E. Zanoni. Springer International Publishing, 2017, pp. 197–236.
- [56] T. Zhu and R. A. Oliver. "Unintentional doping in Ga_N". In: *Physical Chemistry Chemical Physics* 14.27 (2012), p. 9558.
- [57] Z. Xie, Y. Sui, J. Buckeridge, C. R. A. Catlow, T. W. Keal, P. Sherwood, M. R. Farrow, D. O. Scanlon, S. M. Woodley, and A. A. Sokol. "Donor and Acceptor Characteristics of Native Point Defects in Ga_N". In: *arXiv:1803.06273* (2018).
- [58] D. V. Lang. "Deep-level transient spectroscopy: A new method to characterize traps in semiconductors". In: *Journal of Applied Physics* 45.7 (July 1974), pp. 3023–3032.
- [59] A. M. Armstrong and R. J. Kaplar. "Deep-Level Characterization: Electrical and Optical Methods". In: *Power Ga_N Devices: Materials, Applications and Reliability*. Ed. by M. Meneghini, G. Meneghesso, and E. Zanoni. Springer International Publishing, 2017, pp. 145–163.
- [60] A. Chantre, G. Vincent, and D. Bois. "Deep-level optical spectroscopy in GaAs". In: *Phys. Rev. B* 23.10 (May 1981), pp. 5335–5359.
- [61] J. D. McNamara, M. A. Foussekis, A. A. Baski, X. Li, V. Avrutin, H. Morkoç, J. H. Leach, T. Paskova, K. Udvary, E. Preble, and M. A. Reshchikov. "Electrical and optical properties of bulk Ga_N substrates studied by Kelvin probe and photoluminescence". In: *physica status solidi (c)* 10.3 (Mar. 2013), pp. 536–539.
- [62] A. R. Arehart. "Investigation of electrically active defects in Ga_N, AlGa_N, and AlGa_N/Ga_N high electron mobility transistors". PhD thesis. Ohio State University, 2009.
- [63] A. M. Armstrong. "Investigation of deep level defects in Ga_N:C, Ga_N:Mg and pseudomorphic AlGa_N/Ga_N films". PhD thesis. Ohio State University, 2006.
- [64] A. R. Arehart, A. A. Allerman, and S. A. Ringel. "Electrical characterization of n-type Al_{0.30}Ga_{0.70}N Schottky diodes". In: *Journal of Applied Physics* 109.11 (2011), p. 114506.
- [65] C. Ostermaier, P. Lager, M. Reiner, and D. Pogany. "Review of bias-temperature instabilities at the III-N/dielectric interface". In: *Microelectronics Reliability* 82 (Mar. 2018), pp. 62–83.
- [66] P. Lager, P. Steinschifter, M. Reiner, M. Stadtmüller, G. Denifl, A. Naumann, J. Müller, L. Wilde, J. Sundqvist, D. Pogany, and C. Ostermaier. "Role of the dielectric for the charging dynamics of the dielectric/barrier interface in AlGa_N/Ga_N based metal-insulator-semiconductor structures under forward gate bias stress". In: *Applied Physics Letters* 105.3 (July 2014), p. 033512.

- [67] J. del Alamo and J. Joh. “GaN HEMT reliability”. In: *Microelectronics Reliability* 49.9-11 (Sept. 2009), pp. 1200–1206.
- [68] D. Bisi, M. Meneghini, M. Van Hove, D. Marcon, S. Stoffels, T.-L. Wu, S. Decoutere, G. Meneghesso, and E. Zanoni. “Trapping mechanisms in GaN-based MIS-HEMTs grown on silicon substrate”. In: *physica status solidi (a)* 212.5 (May 2015), pp. 1122–1129.
- [69] M. Meneghini, A. Tajalli, P. Moens, A. Banerjee, E. Zanoni, and G. Meneghesso. “Trapping phenomena and degradation mechanisms in GaN-based power HEMTs”. In: *Materials Science in Semiconductor Processing* 78 (May 2018), pp. 118–126.
- [70] G. Meneghesso, G. Verzellesi, R. Pierobon, F. Rampazzo, A. Chini, U. Mishra, C. Canali, and E. Zanoni. “Surface-Related Drain Current Dispersion Effects in AlGaIn-GaN HEMTs”. In: *IEEE Transactions on Electron Devices* 51.10 (Oct. 2004), pp. 1554–1561.
- [71] J. Joh and J. A. del Alamo. “A Current-Transient Methodology for Trap Analysis for GaN High Electron Mobility Transistors”. In: *IEEE Transactions on Electron Devices* 58.1 (Jan. 2011), pp. 132–140.
- [72] A. Santarelli, R. Cignani, G. P. Gibiino, D. Niessen, P. A. Traverso, C. Florian, D. M. M. P. Schreurs, and F. Filicori. “A Double-Pulse Technique for the Dynamic I/V Characterization of GaN FETs”. In: *IEEE Microwave and Wireless Components Letters* 24.2 (Feb. 2014), pp. 132–134.
- [73] D. Bisi, M. Meneghini, C. de Santi, A. Chini, M. Dammann, P. Brückner, M. Mikulla, G. Meneghesso, and E. Zanoni. “Deep-Level Characterization in GaN HEMTs-Part I: Advantages and Limitations of Drain Current Transient Measurements”. In: *IEEE Transactions on Electron Devices* 60.10 (Oct. 2013), pp. 3166–3175.
- [74] G. Meneghesso, M. Meneghini, C. De Santi, M. Ruzzarin, and E. Zanoni. “Positive and negative threshold voltage instabilities in GaN-based transistors”. In: *Microelectronics Reliability* 80 (Jan. 2018), pp. 257–265.
- [75] P. Lager, M. Reiner, D. Pogany, and C. Ostermaier. “Comprehensive Study of the Complex Dynamics of Forward Bias-Induced Threshold Voltage Drifts in GaN Based MIS-HEMTs by Stress/Recovery Experiments”. In: *IEEE Transactions on Electron Devices* 61.4 (Apr. 2014), pp. 1022–1030.
- [76] C.-Y. Hu and T. Hashizume. “Non-localized trapping effects in AlGaIn/GaN heterojunction field-effect transistors subjected to on-state bias stress”. In: *Journal of Applied Physics* 111.8 (2012), p. 84504.
- [77] G. Meneghesso, G. Verzellesi, F. Danesin, F. Rampazzo, F. Zanon, A. Tazzoli, M. Meneghini, and E. Zanoni. “Reliability of GaN High-Electron-Mobility Transistors: State of the Art and Perspectives”. In: *IEEE Transactions on Device and Materials Reliability* 8.2 (June 2008), pp. 332–343.
- [78] D. Bisi, C. D. Santi, M. Meneghini, S. Wienecke, M. Guidry, H. Li, E. Ahmadi, S. Keller, U. K. Mishra, G. Meneghesso, and E. Zanoni. “Observation of Hot Electron and Impact Ionization in N-Polar GaN MIS-HEMTs”. In: *IEEE Electron Device Letters* 39.7 (July 2018), pp. 1007–1010.
- [79] Y. Dong, R. M. Feenstra, and J. E. Northrup. “Oxidized GaN(0001) surfaces studied by scanning tunneling microscopy and spectroscopy and by first-principles theory”. In: *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena* 24.4 (2006), pp. 2080–2086.
- [80] Y. Dong, R. M. Feenstra, and J. E. Northrup. “Electronic states of oxidized GaN(0001) surfaces”. In: *Applied Physics Letters* 89.17 (2006), p. 171920.
- [81] H. Marchand, J. P. Ibbetson, P. T. Fini, P. Kozodoy, S. Keller, S. DenBaars, J. S. Speck, and U. K. Mishra. “Atomic force microscopy observation of threading dislocation

- density reduction in lateral epitaxial overgrowth of gallium nitride by MOCVD". In: *MRS Internet Journal of Nitride Semiconductor Research* 3 (1998), e3.
- [82] M. Jupina and P. Lenahan. "A Spin-Dependent Recombination Study of Radiation-Induced P_{b1} Centers at the (001) Si/SiO₂ Interface". In: *IEEE Transactions on Nuclear Science* 36.6 (Dec. 1989), pp. 1800–1807.
- [83] W. E. Carlos, E. R. Glaser, T. A. Kennedy, and S. Nakamura. "Paramagnetic resonance in GaN-based light emitting diodes". In: *Applied Physics Letters* 67.16 (1995), pp. 2376–2378.
- [84] A. Stesmans, B. Nouwen, and V. V. Afanas'ev. " P_{b1} interface defect in thermal (100)Si/SiO₂ : ²⁹Si hyperfine interaction". In: *Physical Review B* 58.23 (Dec. 1998), pp. 15801–15809.
- [85] W. E. Carlos, J. A. Freitas, M. A. Khan, D. T. Olson, and J. N. Kuznia. "Electron-spin-resonance studies of donors in wurtzite GaN". In: *Phys. Rev. B* 48.24 (Dec. 1993), pp. 17878–17884.
- [86] W. Goes, Y. Wimmer, A.-M. El-Sayed, G. Rzepa, M. Jech, A. Shluger, and T. Grasser. "Identification of oxide defects in semiconductor devices: A systematic approach linking DFT to rate equations and experimental evidence". In: *Microelectronics Reliability* 87 (Aug. 2018), pp. 286–320.
- [87] A. Kerber and E. Cartier. "Bias Temperature Instability Characterization Methods". In: *Bias Temperature Instability for Devices and Circuits*. Ed. by T. Grasser. Springer New York, 2014, pp. 3–31.
- [88] G. Rzepa. "Efficient physical modeling of bias temperature instability". PhD thesis. Institute for Microelectronics, TU Wien, 2018.
- [89] B. Ullmann. "Mixed negative bias temperature instability and hot-carrier stress". PhD thesis. Institute for Microelectronics, TU Wien, 2018.
- [90] S. Imam, S. Sabri, and T. Szkopek. "Low-frequency noise and hysteresis in graphene field-effect transistors on oxide". In: *Micro & Nano Letters* 5.1 (Feb. 2010), p. 37.
- [91] K. Chatty, S. Banerjee, T. P. Chow, and R. J. Gutmann. "Hysteresis in transfer characteristics in 4H-SiC depletion/accumulation-mode MOSFETs". In: *IEEE Electron Device Letters* 23.6 (June 2002), pp. 330–332.
- [92] Y. Y. Illarionov, G. Rzepa, M. Waltl, T. Knobloch, A. Grill, M. M. Furchi, T. Mueller, and T. Grasser. "The role of charge trapping in MoS₂/SiO₂ and MoS₂/hBN field-effect transistors". In: *2D Materials* 3.3 (July 2016), p. 035004.
- [93] M. Capriotti, P. Lagger, C. Fleury, M. Oposich, O. Bethge, C. Ostermaier, G. Strasser, and D. Pogany. "Modeling small-signal response of GaN-based metal-insulator-semiconductor high electron mobility transistor gate stack in spill-over regime: Effect of barrier resistance and interface states". In: *Journal of Applied Physics* 117.2 (2015).
- [94] P. Lagger, A. Schiffmann, G. Pobegen, D. Pogany, and C. Ostermaier. "Very fast dynamics of threshold voltage drifts in GaN-based MIS-HEMTs". In: *IEEE Electron Device Letters* 34.9 (2013), pp. 1112–1114.
- [95] C. Mizue, Y. Hori, M. Miczek, and T. Hashizume. "Capacitance-Voltage Characteristics of Al₂O₃/AlGaN/GaN Structures and State Density Distribution at Al₂O₃/AlGaN Interface". In: *Japanese Journal of Applied Physics* 50.2R (2011), p. 21001.
- [96] C. Ostermaier, H.-C. Lee, S.-Y. Hyun, S.-I. Ahn, K.-W. Kim, H.-I. Cho, J.-B. Ha, and J.-H. Lee. "Interface characterization of ALD deposited Al₂O₃ on GaN by CV method". In: *physica status solidi (c)* 5.6 (May 2008), pp. 1992–1994.
- [97] E. H. Nicollian and A. Goetzberger. "The Si-SiO₂ Interface - Electrical Properties as Determined by the Metal-Insulator-Silicon Conductance Technique". In: *Bell System Technical Journal* 46.6 (July 1967), pp. 1055–1133.

- [98] X. Liu, H. C. Chin, L. S. Tan, and Y. C. Yeo. "In situ Surface Passivation of Gallium Nitride for Metal-Organic Chemical Vapor Deposition of High-Permittivity Gate Dielectric". In: *IEEE Transactions on Electron Devices* 58.1 (Jan. 2011), pp. 95–102.
- [99] M. Ershov, S. Saxena, H. Karbasi, S. Winters, S. Minehane, J. Babcock, R. Lindley, P. Clifton, M. Redford, and A. Shibkov. "Dynamic recovery of negative bias temperature instability in p-type metal-oxide-semiconductor field-effect transistors". In: *Applied Physics Letters* 83.8 (2003), pp. 1647–1649.
- [100] B. Kaczer, T. Grasser, J. Roussel, J. Martin-Martinez, R. O'Connor, B. J. O'Sullivan, and G. Groeseneken. "Ubiquitous relaxation in BTI stressing – New evaluation and insights". In: *2008 IEEE International Reliability Physics Symposium*. IEEE, Apr. 2008, pp. 20–27.
- [101] H. Reisinger, O. Blank, W. Heinrigs, A. Muhlhoff, W. Gustin, and C. Schlunder. "Analysis of NBTI Degradation- and Recovery-Behavior Based on Ultra Fast VT-Measurements". In: *2006 IEEE International Reliability Physics Symposium Proceedings*. Mar. 2006, pp. 448–453.
- [102] M. Wlatl. "Characterization of Bias Temperature Instabilities in Modern Transistor Technologies". PhD thesis. Institute for Microelectronics, TU Wien, 2016.
- [103] M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, Y. Rey-Tauriac, and N. Revil. "On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's". In: *IEDM Technical Digest. IEEE International Electron Devices Meeting, 2004*. IEEE, Dec. 2004, pp. 109–112.
- [104] R. Stradiotto, G. Pobegen, C. Ostermaier, and K.-T. Grasser. "On The Fly Characterization of Charge Trapping Phenomena at GaN/Dielectric and GaN/AlGaIn/Dielectric Interfaces Using Impedance Measurements". In: *Proceedings of the 45th European Solid State Device Research Conference*. 2015, pp. 218–225.
- [105] T. Grasser, P. J. Wagner, P. Hehenberger, W. Goes, and B. Kaczer. "A Rigorous Study of Measurement Techniques for Negative Bias Temperature Instability". In: *IEEE Transactions on Device and Materials Reliability* 8.3 (Sept. 2008), pp. 526–535.
- [106] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth, and D. M. Tennant. "Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1/f) Noise". In: *Phys. Rev. Lett.* 52.3 (Jan. 1984), pp. 228–231.
- [107] M. J. Uren, D. J. Day, and M. J. Kirton. "1/f and random telegraph noise in silicon metal-oxide-semiconductor field-effect transistors". In: *Applied Physics Letters* 47.11 (1985), pp. 1195–1197.
- [108] M. J. Uren, M. J. Kirton, and S. Collins. "Anomalous telegraph noise in small-area silicon metal-oxide-semiconductor field-effect transistors". In: *Physical Review B* 37.14 (1988), pp. 8346–8350.
- [109] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel. "Understanding negative bias temperature instability in the context of hole trapping". In: *Microelectronic Engineering* 86.7-9 (July 2009), pp. 1876–1882.
- [110] T. Grasser. "Stochastic charge trapping in oxides: From random telegraph noise to bias temperature instabilities". In: *Microelectronics Reliability* 52.1 (Jan. 2012), pp. 39–70.
- [111] C.-Y. Chen, Q. Ran, Hyun-Jin Cho, A. Kerber, Y. Liu, M.-R. Lin, and R. W. Dutton. "Correlation of I_{d-} and I_{d-} Random Telegraph Noise to Positive Bias Temperature Instability in Scaled High- κ /Metal Gate n-type MOSFETs". In: *2011 International Reliability Physics Symposium*. IEEE, Apr. 2011, 3A.2.1–3A.2.6.
- [112] S. Guo, R. Wang, D. Mao, Y. Wang, and R. Huang. "Anomalous random telegraph noise in nanoscale transistors as direct evidence of two metastable states of oxide traps". In: *Scientific Reports* 7.1 (Dec. 2017), p. 6239.

- [113] R. Wang, S. Guo, Z. Zhang, J. Zou, D. Mao, and R. Huang. “Complex Random Telegraph Noise (RTN): What Do We Understand?” In: *2018 IEEE 25th International Symposium on the Physical and Failure Analysis of Integrated Circuits*. 2018.
- [114] A. Karwath and M. Schulz. “Deep level transient spectroscopy on single, isolated interface traps in field-effect transistors”. In: *Applied Physics Letters* 52.8 (1988), pp. 634–636.
- [115] H. Reisinger, T. Grasser, W. Gustin, and C. Schlünder. “The statistical analysis of individual defects constituting NBTI and its implications for modeling DC- and AC-stress”. In: *2010 IEEE International Reliability Physics Symposium*. May 2010, pp. 7–15.
- [116] T. Grasser, H. Reisinger, P.-J. Wagner, F. Schanovsky, W. Goes, and B. Kaczer. “The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability”. In: *2010 IEEE International Reliability Physics Symposium*. Vol. 1835. IEEE, 2010, pp. 16–25.
- [117] H. Reisinger. “The Time-Dependent Defect Spectroscopy”. In: *Bias Temperature Instability for Devices and Circuits*. Ed. by T. Grasser. Springer New York, 2014, pp. 75–109.
- [118] T. Grasser. “The Capture/Emission Time Map Approach to the Bias Temperature Instability”. In: *Bias Temperature Instability for Devices and Circuits*. Ed. by T. Grasser. Springer New York, 2014, pp. 447–481.
- [119] H. Reisinger, T. Grasser, K. Ermisch, H. Nielen, W. Gustin, and C. Schlünder. “Understanding and modeling AC BTI”. In: *2011 International Reliability Physics Symposium*. Apr. 2011, 6A.1.1–6A.1.8.
- [120] C. H. Henry and D. V. Lang. “Nonradiative capture and recombination by multiphonon emission in GaAs and GaP”. In: *Phys. Rev. B* 15.2 (Jan. 1977), pp. 989–1016.
- [121] P. T. Landsberg. “Non-Radiative Transitions in Semiconductors”. In: *physica status solidi (b)* 41.2 (1970), pp. 457–489.
- [122] D. V. Lang and C. H. Henry. “Nonradiative Recombination at Deep Levels in GaAs and GaP by Lattice-Relaxation Multiphonon Emission”. In: *Phys. Rev. Lett.* 35.22 (Dec. 1975), pp. 1525–1528.
- [123] A. M. Stoneham. “Non-radiative transitions in semiconductors”. In: *Reports on Progress in Physics* 44.12 (1981), p. 1251.
- [124] S. Makram-Ebeid and M. Lannoo. “Quantum model for phonon-assisted tunnel ionization of deep levels in a semiconductor”. In: *Phys. Rev. B* 25.10 (May 1982), pp. 6406–6424.
- [125] T. L. Tewksbury. “Relaxation effects in MOS devices due to tunnel exchange with near-interface oxide traps”. PhD thesis. Massachusetts Institute of Technology, 1992.
- [126] T. L. Tewksbury and H.-S. Lee. “Characterization, modeling, and minimization of transient threshold voltage shifts in MOSFETs”. In: *IEEE Journal of Solid-State Circuits* 29.3 (Mar. 1994), pp. 239–252.
- [127] M. Born and R. Oppenheimer. “Zur Quantentheorie der Molekeln”. In: *Annalen der Physik* 389.20 (1927), pp. 457–484.
- [128] Y. Wimmer. “Hydrogen Related Defects in Amorphous SiO₂ and the Negative Bias Temperature Instability”. PhD thesis. Institute for Microelectronics, TU Wien, 2017.
- [129] E. Condon. “A Theory of Intensity Distribution in Band Systems”. In: *Phys. Rev.* 28.6 (Dec. 1926), pp. 1182–1201.
- [130] K. Huang and A. Rhys. “Theory of Light Absorption and Non-Radiative Transitions in F-Centres”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 204.1078 (Dec. 1950), pp. 406–423.

- [131] M. B. Weissman. “ $1/f$ noise and other slow, nonexponential kinetics in condensed matter”. In: *Rev. Mod. Phys.* 60.2 (Apr. 1988), pp. 537–571.
- [132] J. P. Campbell, J. Qin, K. P. Cheung, L. C. Yu, J. S. Suehle, A. Oates, and K. Sheng. “Random telegraph noise in highly scaled nMOSFETs”. In: *2009 IEEE International Reliability Physics Symposium*. Apr. 2009, pp. 382–388.
- [133] T. Nagumo, K. Takeuchi, T. Hase, and Y. Hayashi. “Statistical characterization of trap position, energy, amplitude and time constants by RTN measurement of multiple individual traps”. In: *2010 International Electron Devices Meeting*. IEEE, Dec. 2010, pp. 28.3.1–28.3.4.
- [134] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel. “A two-stage model for negative bias temperature instability”. In: *2009 IEEE International Reliability Physics Symposium*. Apr. 2009, pp. 33–44.
- [135] G. Rzepa, W. Goes, G. Rott, K. Rott, M. Karner, C. Kernstock, B. Kaczer, H. Reisinger, and T. Grasser. “Physical modeling of NBTI: From individual defects to devices”. In: *2014 International Conference on Simulation of Semiconductor Processes and Devices*. Sept. 2014, pp. 81–84.
- [136] G. Rzepa, M. Walzl, W. Goes, B. Kaczer, and T. Grasser. “Microscopic oxide defects causing BTI, RTN, and SILC on high-k FinFETs”. In: *2015 International Conference on Simulation of Semiconductor Processes and Devices*. IEEE. IEEE, Sept. 2015, pp. 144–147.
- [137] Y. Y. Illarionov, M. Walzl, G. Rzepa, J.-S. Kim, S. Kim, A. Dodabalapur, D. Akinwande, and T. Grasser. “Long-Term Stability and Reliability of Black Phosphorus Field-Effect Transistors”. In: *ACS Nano* 10.10 (Oct. 2016), pp. 9543–9549.
- [138] Y. Y. Illarionov, T. Knobloch, M. Walzl, G. Rzepa, A. Pospischil, D. K. Polyushkin, M. M. Furchi, T. Mueller, and T. Grasser. “Energetic mapping of oxide traps in MoS₂ field-effect transistors”. In: *2D Materials* 4.2 (June 2017), p. 025108.
- [139] T. Knobloch, G. Rzepa, Y. Y. Illarionov, M. Walzl, D. Polyushkin, A. Pospischil, M. Furchi, T. Mueller, and T. Grasser. “Impact of Gate Dielectrics on the Threshold Voltage in MoS₂ Transistors”. In: *ECS Transactions* 80.1 (Aug. 2017), pp. 203–217.
- [140] L. Landau and E. Lifshitz. *Quantum Mechanics*. Addison-Wesley, 1958.
- [141] R. H. Fowler and L. Nordheim. “Electron emission in intense electric fields”. In: *Proc. R. Soc. Lond. A* 119.781 (1928), pp. 173–181.
- [142] W. Gös. “Hole Trapping and the Negative Bias Temperature Instability”. PhD thesis. Institute for Microelectronics, TU Wien, 2011.
- [143] S. E. Rauch. “Review and Reexamination of Reliability Effects Related to NBTI-Induced Statistical Variations”. In: *IEEE Transactions on Device and Materials Reliability* 7.4 (Dec. 2007), pp. 524–530.
- [144] B. Kaczer, T. Grasser, P. J. Roussel, J. Franco, R. Degraeve, L.-A. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger. “Origin of NBTI variability in deeply scaled pFETs”. In: *2010 IEEE International Reliability Physics Symposium*. IEEE, May 2010, pp. 26–32.
- [145] J. M. M. de Nijs, K. G. Druijf, V. V. Afanas’ev, E. van der Drift, and P. Balk. “Hydrogen induced donor-type Si/SiO₂ interface states”. In: *Applied Physics Letters* 65.19 (1994), pp. 2428–2430.
- [146] A. Stesmans and V. V. Afanas’ev. “¹³C - ESR of interfaces and nanolayers in semiconductor heterostructures”. In: *Characterization of Semiconductor Heterostructures and Nanostructures*. Ed. by C. Lamberti. Elsevier, 2008, pp. 435–482.
- [147] T. Grasser, M. Walzl, Y. Wimmer, W. Goes, R. Kosik, G. Rzepa, H. Reisinger, G. Pobegen, A. El-Sayed, A. Shluger, and B. Kaczer. “Gate-sided hydrogen release as

- the origin of "permanent" NBTI degradation: From single defects to lifetimes". In: *2015 IEEE International Electron Devices Meeting*. Dec. 2015, pp. 20.1.1–20.1.4.
- [148] M. Herrmann and A. Schenk. "Field and high-temperature dependence of the long term charge loss in erasable programmable read only memories: Measurements and modeling". In: *Journal of Applied Physics* 77.9 (1995), pp. 4522–4540.
- [149] L. Larcher. "Statistical simulation of leakage currents in MOS and flash memory devices with a new multiphonon trap-assisted tunneling model". In: *IEEE Transactions on Electron Devices* 50.5 (May 2003), pp. 1246–1253.
- [150] A. Padovani, L. Larcher, S. Verma, P. Pavan, P. Majhi, P. Kapur, K. Parat, S. Bersuker, and K. Saraswat. "Statistical Modeling of Leakage Currents Through SiO₂/High- κ Dielectrics Stacks for Non-Volatile Memory Applications". In: *2008 IEEE International Reliability Physics Symposium*. Apr. 2008, pp. 616–620.
- [151] L. Vandelli, A. Padovani, L. Larcher, R. G. Southwick, W. B. Knowlton, and G. Bersuker. "A Physical Model of the Temperature Dependence of the Current Through SiO₂/HfO₂ Stacks". In: *IEEE Transactions on Electron Devices* 58.9 (Sept. 2011), pp. 2878–2887.
- [152] J. Frenkel. "On Pre-Breakdown Phenomena in Insulators and Electronic Semiconductors". In: *Physical Review* 54.8 (Oct. 1938), pp. 647–648.
- [153] D. T. Gillespie. *Markov processes : an introduction for physical scientists*. Academic Press, 1992, p. 565.
- [154] O. C. Ibe. *Markov processes for stochastic modeling*. Elsevier insights. Elsevier, 2013, p. 493.
- [155] D. J. Frank and H. Miki. "Analysis of Oxide Traps in Nanoscale MOSFETs using Random Telegraph Noise". In: *Bias Temperature Instability for Devices and Circuits*. Ed. by T. Grassler. Springer New York, 2014, pp. 111–134.
- [156] Z. Jibin, W. Runsheng, G. Shaofeng, L. Mulong, Y. Zhuoqing, J. Xiaobo, R. Pengpeng, W. Jianping, L. Jinhua, W. Jingang, W. Waisum, Y. Shaofeng, W. Hanming, L. Shiuh-Wuu, W. Yangyuan, and H. Ru. "New understanding of state-loss in complex RTN: Statistical experimental study, trap interaction models, and impact on circuits". In: *2014 IEEE International Electron Devices Meeting*. IEEE, Dec. 2014, pp. 34.5.1–34.5.4.
- [157] L. Gerrer, J. Ding, S. Amoroso, F. Adamu-Lema, R. Hussin, D. Reid, C. Millar, and A. Asenov. "Modelling RTN and BTI in nanoscale MOSFETs from device to circuit: A review". In: *Microelectronics Reliability* 54.4 (Apr. 2014), pp. 682–697.
- [158] M. Huymajer. "Cluster Detection Algorithm to Study Single Charge Trapping Events in TDDS". masterthesis. Institute for Microelectronics, TU Wien, 2017.
- [159] T. Nagumo, K. Takeuchi, S. Yokogawa, K. Imai, and Y. Hayashi. "New analysis methods for comprehensive understanding of Random Telegraph Noise". In: *2009 IEEE International Electron Devices Meeting*. IEEE, Dec. 2009, pp. 1–4.
- [160] Y. Yuzhelevski, M. Yuzhelevski, and G. Jung. "Random telegraph noise analysis in time domain". In: *Review of Scientific Instruments* 71.4 (Apr. 2000), pp. 1681–1688.
- [161] A. K. Jain. "Data clustering: 50 years beyond K-means". In: *Pattern Recognition Letters* 31.8 (June 2010), pp. 651–666.
- [162] T. Obara, A. Teramoto, A. Yonezawa, R. Kuroda, S. Sugawa, and T. Ohmi. "Analyzing correlation between multiple traps in RTN characteristics". In: *2014 IEEE International Reliability Physics Symposium*. IEEE, June 2014, 4A.6.1–4A.6.7.
- [163] M. Maestro, J. Diaz, A. Crespo-Yepes, M. Gonzalez, J. Martin-Martinez, R. Rodriguez, M. Nafria, F. Campabadal, and X. Aymerich. "New high resolution Random Telegraph Noise (RTN) characterization method for resistive RAM". In: *Solid-State Electronics* 115 (Jan. 2016), pp. 140–145.

- [164] T. Grasser, H. Reisinger, P.-J. Wagner, and B. Kaczer. “Time-dependent defect spectroscopy for characterization of border traps in metal-oxide-semiconductor transistors”. In: *Physical Review B* 82.24 (Dec. 2010), p. 245318.
- [165] J. Canny. “A Computational Approach to Edge Detection”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8.6 (Nov. 1986), pp. 679–698.
- [166] H. Steinhaus. “Sur la division des corp materiels en parties”. In: *Bull. Acad. Polon. Sci* 1.804 (1956), p. 801.
- [167] S. Lloyd. “Least squares quantization in PCM”. In: *IEEE Transactions on Information Theory* 28.2 (Mar. 1982), pp. 129–137.
- [168] G. H. Ball and D. J. Hall. *ISODATA, a novel method of data analysis and pattern classification*. Tech. rep. Stanford research inst Menlo Park CA, 1965.
- [169] J. MacQueen et al. “Some methods for classification and analysis of multivariate observations”. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 14. Oakland, CA, USA. 1967, pp. 281–297.
- [170] L. E. Baum and T. Petrie. “Statistical Inference for Probabilistic Functions of Finite State Markov Chains”. In: *The Annals of Mathematical Statistics* 37.6 (1966), pp. 1554–1563.
- [171] L. E. Baum, T. Petrie, G. Soules, and N. Weiss. “A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov”. In: *The Annals of Mathematical Statistics* 41.1 (1970), pp. 164–171.
- [172] L. Rabiner. “A tutorial on hidden Markov models and selected applications in speech recognition”. In: *Proceedings of the IEEE* 77.2 (1989), pp. 257–286.
- [173] M. Stamp. *A revealing introduction to hidden Markov models*. Tech. rep. 2004, p. 20.
- [174] F. G. David. “The viterbi algorithm”. In: *Proceedings of the IEEE* 61.3 (1973), pp. 302–309.
- [175] L. E. Baum. “An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes”. In: *Inequalities III: Proceedings of the Third Symposium on Inequalities*. Ed. by O. Shisha. University of California, Los Angeles: Academic Press, 1972, pp. 1–8.
- [176] G. Rodriguez. *Smoothing and Non-Parametric Regression*. Tech. rep. 2001, p. 12.
- [177] P. Dierckx. *Curve and Surface Fitting with Splines*. Oxford University Press, Inc., 1993.
- [178] W. S. Cleveland. “Robust Locally Weighted Regression and Smoothing Scatterplots”. In: *Journal of the American Statistical Association* 74.368 (1979), pp. 829–836.
- [179] W. S. Cleveland and S. J. Devlin. “Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting”. In: *Source Journal of the American Statistical Association* 83.403 (1988), pp. 596–610.
- [180] C. J. Vogel. *A faster version of the lowess function in statsmodels.nonparametric.lowess*. 2012.
- [181] C. D. Boor. “Splines as linear combinations of B-splines. A Survey”. In: *Approximation Theory Conference*. 1976, pp. 1–47.
- [182] P. Dierckx. “An algorithm for smoothing, differentiation and integration of experimental data using spline functions”. In: *Journal of Computational and Applied Mathematics* 1.3 (1975), pp. 165–184.
- [183] P. Dierckx. “A Fast Algorithm for Smoothing Data on a Rectangular Grid While Using Spline Functions”. In: *SIAM Journal on Numerical Analysis* 19.6 (1982), pp. 1286–1304.
- [184] S.-J. Baek, A. Park, Y.-J. Ahn, and J. Choo. “Baseline correction using asymmetrically reweighted penalized least squares smoothing”. In: *The Analyst* 140.1 (2015), pp. 250–257.

- [185] P. H. C. Eilers. “A Perfect Smoother”. In: *Analytical Chemistry* 75.14 (July 2003), pp. 3631–3636.
- [186] P. H. C. Eilers and H. F. M. Boelens. “Baseline correction with asymmetric least squares smoothing”. In: *Leiden University Medical Centre Report* 1.1 (2005), p. 5.
- [187] Z.-M. Zhang, S. Chen, and Y.-Z. Liang. “Baseline correction using adaptive iteratively reweighted penalized least squares”. In: *The Analyst* 135.5 (2010), p. 1138.
- [188] S. M. Ross. *Introduction to probability and statistics for engineers and scientists*. Elsevier, 2009.
- [189] T. Imada, M. Kanamura, and T. Kikkawa. “Enhancement-mode GaN MIS-HEMTs for power supplies”. In: *The 2010 International Power Electronics Conference*. IEEE, June 2010, pp. 1027–1033.
- [190] P. Lagger, C. Ostermaier, G. Pobegen, and D. Pogany. “Towards understanding the origin of threshold voltage instability of AlGaIn/GaN MIS-HEMTs”. In: *2012 International Electron Devices Meeting*. IEEE, Dec. 2012, pp. 13.1.1–13.1.4.
- [191] P. Lagger, C. Ostermaier, and D. Pogany. “Enhancement of V_{th} drift for repetitive gate stress pulses due to charge feedback effect in GaN MIS-HEMTs”. In: *2014 IEEE International Reliability Physics Symposium*. IEEE, June 2014, pp. 6C.3.1–6C.3.6.
- [192] O. Ambacher, B. Foutz, J. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Murphy, A. J. Sierakowski, W. J. Schaff, L. F. Eastman, R. Dimitrov, A. Mitchell, and M. Stutzmann. “Two dimensional electron gases induced by spontaneous and piezoelectric polarization in undoped and doped AlGaIn/GaN heterostructures”. In: *Journal of Applied Physics* 87.1 (Jan. 2000), pp. 334–344.
- [193] T. Simlinger. “Simulation von Heterostruktur-Feldeffekttransistoren”. PhD thesis. Institute for Microelectronics, TU Wien, 1996.
- [194] B. S. Eller, J. Yang, and R. J. Nemanich. “Electronic surface and dielectric interface states on GaN and AlGaIn”. In: *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* 31.5 (Sept. 2013), p. 050807.
- [195] T. Grasser, P.-J. Wagner, H. Reisinger, T. Aichinger, G. Pobegen, M. Nelhiebel, and B. Kaczer. “Analytic modeling of the bias temperature instability using capture/emission time maps”. In: *2011 International Electron Devices Meeting*. IEEE, Dec. 2011, pp. 27.4.1–27.4.4.
- [196] K.-S. Im, H.-S. Kang, J.-H. Lee, S.-J. Chang, S. Cristoloveanu, M. Bawedin, and J.-H. Lee. “Characteristics of GaN and AlGaIn/GaN FinFETs”. In: *Solid-State Electronics* 97 (July 2014), pp. 66–75.
- [197] M. Reiner, G. Denifl, M. Stadtmueller, R. Pietschnig, and C. Ostermaier. “Through-layer XPS investigations of the $\text{Si}_3\text{N}_4/\text{AlGaIn}$ interface”. In: *physica status solidi (b)* 253.10 (Oct. 2016), pp. 2009–2014.
- [198] L. C. Grabow, J. J. Uhlrich, T. F. Kuech, and M. Mavrikakis. “Effectiveness of in situ NH_3 annealing treatments for the removal of oxygen from GaN surfaces”. In: *Surface Science* 603.2 (2009), pp. 387–399.
- [199] M. Reiner, P. Lagger, G. Prechtel, P. Steinschifter, R. Pietschnig, D. Pogany, and C. Ostermaier. “Modification of “native” surface donor states in AlGaIn/GaN MIS-HEMTs by fluorination: Perspective for defect engineering”. In: *2015 IEEE International Electron Devices Meeting*. Dec. 2015, pp. 35.5.1–35.5.4.
- [200] M. Reiner, J. Schellander, G. Denifl, M. Stadtmueller, M. Schmid, T. Frischmuth, U. Schmid, R. Pietschnig, and C. Ostermaier. “Physical-chemical stability of fluorinated III-N surfaces: Towards the understanding of the (0001) $\text{Al}_x\text{Ga}_{1-x}\text{N}$ surface donor modification by fluorination”. In: *Journal of Applied Physics* 121.22 (2017), p. 225704.

- [201] D. Hisamoto, T. Kaga, and E. Takeda. "Impact of the vertical SOI 'DELTA' structure on planar device technology". In: *IEEE Transactions on Electron Devices* 38.6 (June 1991), pp. 1419–1424.
- [202] Chenming Hu, J. Bokor, Tsu-Jae King, E. Anderson, C. Kuo, K. Asano, H. Takeuchi, J. Kedzierski, Wen-Chin Lee, and D. Hisamoto. "FinFET-a self-aligned double-gate MOSFET scalable to 20 nm". In: *IEEE Transactions on Electron Devices* 47.12 (2000), pp. 2320–2325.
- [203] Xuejue Huang, Wen-Chin Lee, C. Kuo, D. Hisamoto, Leland Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Yang-Kyu Choi, K. Asano, V. Subramanian, Tsu-Jae King, J. Bokor, and Chenming Hu. "Sub-50 nm P-channel FinFET". In: *IEEE Transactions on Electron Devices* 48.5 (May 2001), pp. 880–886.
- [204] M. A. Alsharif, R. Granzner, and F. Schwierz. "Theoretical Investigation of Trigate AlGaIn/GaN HEMTs". In: *IEEE Transactions on Electron Devices* 60.10 (Oct. 2013), pp. 3335–3341.
- [205] K.-S. Im, H.-S. Kang, D.-K. Kim, S. Vodapally, Y. Park, J.-H. Lee, Y.-T. Kim, S. Cristoloveanu, and J.-H. Lee. "Temperature-dependent characteristics of AlGaIn/GaN FinFETs with sidewall MOS channel". In: *Solid-State Electronics* 120 (2016), pp. 47–51.
- [206] T. Grasser, K. Rott, H. Reisinger, M. Walzl, and W. Goes. "Evidence for defect pairs in SiON pMOSFETs". In: *Proceedings of the 21th International Symposium on the Physical and Failure Analysis of Integrated Circuits*. IEEE, June 2014, pp. 258–263.
- [207] A.-M. El-Sayed and A. L. Shluger. "Atomistic Modeling of Defects Implicated in the Bias Temperature Instability". In: *Bias Temperature Instability for Devices and Circuits*. Ed. by T. Grasser. Springer New York, 2014, pp. 305–321.
- [208] H.-p. Komsa, T. T. Rantala, and A. Pasquarello. "Finite-size supercell correction schemes for charged defect calculations". In: *Physical Review B* 86.4 (2012), pp. 1–16.
- [209] A. Asenov, R. Balasubramaniam, A. Brown, J. Davies, and S. Saini. "Random telegraph signal amplitudes in sub 100 nm (decanano) MOSFETs: a 3D 'Atomistic' simulation study". In: *International Electron Devices Meeting 2000. Technical Digest. IEDM (Cat. No.00CH37138)*. IEEE, 2000, pp. 279–282.
- [210] A. Asenov, A. Brown, J. Davies, S. Kaya, and G. Slavcheva. "Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs". In: *IEEE Transactions on Electron Devices* 50.9 (Sept. 2003), pp. 1837–1852.
- [211] C. Alexander, A. Brown, J. Watling, and A. Asenov. "Impact of scattering in 'atomistic' device simulations". In: *Solid-State Electronics* 49.5 (May 2005), pp. 733–739.
- [212] M. Schulz. "Coulomb energy of traps in semiconductor space-charge regions". In: *Journal of Applied Physics* 74.4 (Aug. 1993), pp. 2649–2657.
- [213] M.-P. Lu and M.-J. Chen. "Oxide-trap-enhanced Coulomb energy in a metal-oxide-semiconductor system". In: *Physical Review B* 72.23 (Dec. 2005), p. 235417.
- [214] M.-P. Lu. "Coulomb-energy featured capture kinetics in graphene nanoribbon field-effect transistors". In: *Physical Review B* 86.4 (July 2012), p. 045433.
- [215] P. Ren, P. Hao, C. Liu, R. Wang, X. Jiang, Y. Qiu, R. Huang, S. Guo, M. Luo, J. Zou, M. Li, J. Wang, J. Wu, J. Liu, W. Bu, W. Wong, S. Yu, H. Wu, S.-W. Lee, and Y. Wang. "New observations on complex RTN in scaled high- κ /metal-gate MOSFETs - The role of defect coupling under DC/AC condition". In: *2013 IEEE International Electron Devices Meeting*. IEEE, Dec. 2013, pp. 31.4.1–31.4.4.
- [216] A. Asenov, G. Slavcheva, A. Brown, J. Davies, and S. Saini. "Increase in the random dopant induced threshold fluctuations and lowering in sub-100 nm MOSFETs due to

- quantum effects: a 3-D density-gradient simulation study”. In: *IEEE Transactions on Electron Devices* 48.4 (Apr. 2001), pp. 722–729.
- [217] M. G. Ancona and G. J. Iafrate. “Quantum correction to the equation of state of an electron gas in a semiconductor”. In: *Phys. Rev. B* 39.13 (May 1989), pp. 9536–9540.
- [218] A. Wettstein, A. Schenk, and W. Fichtner. “Quantum device-simulation with the density-gradient model on unstructured grids”. In: *IEEE Transactions on Electron Devices* 48.2 (Feb. 2001), pp. 279–284.
- [219] E. Conwell and V. F. Weisskopf. “Theory of Impurity Scattering in Semiconductors”. In: *Physical Review* 77.3 (Feb. 1950), pp. 388–390.
- [220] C. Alexander, G. Roy, and A. Asenov. “Random-Dopant-Induced Drain Current Variation in Nano-MOSFETs: A Three-Dimensional Self-Consistent Monte Carlo Simulation Study Using “Ab Initio” Ionized Impurity Scattering”. In: *IEEE Transactions on Electron Devices* 55.11 (Nov. 2008), pp. 3251–3258.
- [221] S. M. Amoroso, L. Gerrer, F. Adamu-Lema, S. Markov, and A. Asenov. “Statistical Study of Bias Temperature Instabilities by Means of 3D “Atomistic” Simulation”. In: *Bias Temperature Instability for Devices and Circuits*. Ed. by T. Grasser. Springer New York, 2014, pp. 323–348.
- [222] P. Hao, D. Mao, R. Wang, S. Guo, P. Ren, and R. Huang. “On the frequency dependence of oxide trap coupling in nanoscale MOSFETs: Understanding based on complete 4-state trap model”. In: *2016 13th IEEE International Conference on Solid-State and Integrated Circuit Technology*. IEEE, Oct. 2016, pp. 981–983.
- [223] H. Miki, N. Tega, M. Yamaoka, D. J. Frank, A. Bansal, M. Kobayashi, K. Cheng, C. P. D’Emic, Z. Ren, S. Wu, J.-B. Yau, Y. Zhu, M. A. Guillorn, D.-G. Park, W. Haensch, E. Leobandung, and K. Torii. “Statistical measurement of random telegraph noise and its impact in scaled-down high- κ /metal-gate MOSFETs”. In: *2012 International Electron Devices Meeting*. IEEE, Dec. 2012, pp. 19.1.1–19.1.4.
- [224] J. Schreiber. “pomegranate: Fast and Flexible Probabilistic Modeling in Python”. In: *The Journal of Machine Learning Research* 18.164 (Oct. 2018), pp. 1–6.

List of Publications

Scientific Journals

- [AGJ1] C. Ostermaier, P. W. Lagger, G. Prechtel, **A. Grill**, K.-T. Grasser, and D. Pogany. “Dynamics of carrier transport via AlGa_N barrier in AlGa_N/Ga_N MIS-HEMTs”. In: *Applied Physics Letters* 110.17 (Apr. 2017), p. 173502.
- [AGJ2] A. A. Makarov, S. E. Tyaginov, B. Kaczer, M. Jech, A. Chasin, **A. Grill**, G. Hellings, M. I. Vexler, D. Linten, and T. Grasser. “Analysis of the Features of Hot-Carrier Degradation in FinFETs”. In: *Semiconductors* 52.10 (Oct. 2018), pp. 1298–1302.
- [AGJ3] M. Waltl, G. Rzepa, **A. Grill**, W. Gös, J. Franco, B. Kaczer, L. Witters, J. Mitard, N. Horiguchi, and K.-T. Grasser. “Superior NBTI in High-k SiGe Transistors - Part I: Experimental”. In: *IEEE Transactions on Electron Devices* 64.5 (2017), pp. 2092–2098.
- [AGJ4] M. Waltl, G. Rzepa, **A. Grill**, W. Gös, J. Franco, B. Kaczer, L. Witters, J. Mitard, N. Horiguchi, and K.-T. Grasser. “Superior NBTI in High-k SiGe Transistors - Part II: Theory”. In: *IEEE Transactions on Electron Devices* 64.5 (2017), pp. 2099–2105.
- [AGJ5] Y. Illarionov, G. Rzepa, M. Waltl, T. Knobloch, **A. Grill**, M. M. Furchi, T. Müller, and K.-T. Grasser. “The Role of Charge Trapping in MoS₂/SiO₂ and MoS₂/hBN Field-Effect Transistors”. In: *2D Materials* 3.3 (2016), pp. 35004–35010.
- [AGJ6] B. Stampfer, F. Zhang, Y. Illarionov, T. Knobloch, P. Wu, M. Waltl, **A. Grill**, J. Appenzeller, and K.-T. Grasser. “Characterization of Single Defects in Ultrascalded MoS₂ Field-Effect Transistors”. In: *ACS Nano* 12.6 (2018), pp. 5368–5375.
- [AGJ7] R. Stradiotto, G. Pobegen, C. Ostermaier, M. Waltl, **A. Grill**, and K.-T. Grasser. “Characterization of Interface Defects With Distributed Activation Energies in GaN-Based MIS-HEMTs”. In: *IEEE Transactions on Electron Devices* 64.3 (Mar. 2017), pp. 1045–1052.
- [AGJ8] M. Reiche, M. Kittler, E. Pippel, H. Uebensee, H. Kosina, **A. Grill**, Z. Stanojevic, and O. Baumgartner. “Impact of Defect-Induced Strain on Device Properties”. In: *Advanced Engineering Materials* 18.12 (2016), pp. 1–4.
- [AGJ9] M. Jech, B. Ullmann, G. Rzepa, S. Tyaginov, **A. Grill**, M. Waltl, D. Jabs, and C. Jungemann. “Impact of Mixed Negative Bias Temperature Instability and Hot Carrier Stress on MOSFET Characteristics - Part II: Theory”. In: *IEEE Transactions on Electron Devices* (2018), accepted for publication.

Conference Proceedings

- [AGC1] **A. Grill**, G. Rzepa, P. Lagger, C. Ostermaier, H. Ceric, and T. Grasser. “Charge feedback mechanisms at forward threshold voltage stress in GaN/AlGa_N HEMTs”. In: *2015 IEEE International Integrated Reliability Workshop*. IEEE, Oct. 2015, pp. 41–45.

- [AGC2] A. Makarov, S. E. Tyaginov, B. Kaczer, M. Jech, A. Chasin, **A. Grill**, G. Hellings, M. Vexler, D. Linten, and K.-T. Grasser. “Hot-Carrier Degradation in FinFETs: Modeling, Peculiarities, and Impact of Device Topology”. In: *Proceedings of the 2017 International Electron Devices Meeting Technical Digest*. 2017, pp. 310–313.
- [AGC3] G. Rzepa, J. Franco, A. Subirats, M. Jech, A. Chasin, **A. Grill**, M. Walzl, T. Knobloch, B. Stampfer, T. Chiarella, N. Horiguchi, L. A. Ragnarsson, D. Linten, B. Kaczer, and K.-T. Grasser. “Efficient Physical Defect Model Applied to PBTI in High- κ Stacks”. In: *2017 IEEE International Reliability Physics Symposium*. 2017, XT-11.1—XT-11.6.
- [AGC4] M. Walzl, **A. Grill**, G. Rzepa, W. Gös, J. Franco, B. Kaczer, J. Mitard, and K.-T. Grasser. “Nanoscale Evidence for the Superior Reliability of SiGe High- κ pMOS-FETs”. In: *2016 IEEE International Reliability Physics Symposium*. 2016, XT-02-1—XT-02-6.
- [AGC5] **A. Grill**, B. Stampfer, M. Walzl, K.-S. Im, J.-H. Lee, C. Ostermaier, H. Ceric, and T. Grasser. “Characterization and modeling of single defects in GaN/AlGaIn fin-MIS-HEMTs”. In: *2017 IEEE International Reliability Physics Symposium*. IEEE, Apr. 2017, 3B-5.1–3B-5.5.
- [AGC6] O. Baumgartner, Z. Stanojevic, L. Filipovic, **A. Grill**, K.-T. Grasser, H. Kosina, and M. Karner. “Investigation of Quantum Transport in Nanoscaled GaN High Electron Mobility Transistors”. In: *Proceedings of the 19th International Conference on Simulation of Semiconductor Processes and Devices*. 2014, pp. 117–120.
- [AGC7] B. Ullmann, M. Jech, S. E. Tyaginov, M. Walzl, Y. Illarionov, **A. Grill**, K. Puschkarsky, H. Reisinger, and K.-T. Grasser. “The Impact of Mixed Negative Bias Temperature Instability and Hot Carrier Stress on Single Oxide Defects”. In: *2017 IEEE International Reliability Physics Symposium*. 2017, XT-10.1—XT-10.6.
- [AGC8] S. E. Tyaginov, M. Jech, G. Rzepa, **A. Grill**, A.-M. El-Sayed, G. Pobegen, A. Makarov, and K.-T. Grasser. “Border Trap Based Modeling of SiC Transistor Transfer Characteristics”. In: *2018 IEEE International Integrated Reliability Workshop*. 2018, accepted for publication.

Master Theses

- [AGT1] B. Stampfer. “Trap Assisted Tunneling and Band Interaction using the Non-Radiative Multi Phonon Model”. Master Thesis (supervised by **A. Grill** and T. Grasser). Institute for Microelectronics, TU Wien, 2016.
- [AGT2] **A. Grill**. “A framework for simulation and parameter optimization of a 90nm CMOS process in Sentaurus”. Master Thesis (supervised by K. Schweiger and H. Zimmermann). Institute for Circuit and Microwave Engineering, TU Wien, 2013.

Curriculum Vitae



ALEXANDER GRILL

Researcher

Institute for Microelectronics
TU Wien, Austria

Born: April 4th, 1982 – Wiener Neustadt, Austria

Education

- since 2013 **Doctoral (PhD) programme in Engineering Sciences**
Microelectronics, TU Wien.
- 2008–2013 **Master program**
Microelectronics, TU Wien, Graduation with honors..
- 2001–2008 **Bachelor of Science**
Electrical Engineering, TU Wien.
- 1996–2001 **High School Diploma**
Electrical engineering, HTL Wiener Neustadt, Graduation with honors..

Professional Experience

Academic

- since 03/2013 **Assistant Professor**
Vienna University of Technology, Vienna.
Reliability of Nitride based heterostructure devices (A.o. Prof. Tibor Grasser, Institute for Microelectronics)
- 2010–2011 **Technical assistant**
Vienna University of Technology, Vienna.
Set up of an optical parametric amplifier
(Univ.Prof. Andrius Baltuska, PhD and Tadas Balciunas, MSc), August–March
- 2010 **Technical assistant**
Vienna University of Technology, Vienna.
Design of a wireless pulse oximeter for the P-Stim project
(A.o. Univ.Prof. Eugenijus Kanisusas), February–April
- 01/08–12/08 **Technical assistant**
Vienna University of Technology, Vienna.
Programming and improvement of the electrical device simulator MINIMOS in C++
(A.o. Prof. Tibor Grasser).

2007 **Technical assistant**
Vienna University of Technology, Vienna.
Design of an elevator model for the course “Microcomputer Laboratory”
(Dipl.Ing. Josef Wenninger), October–February

Teaching

- 2013–2017 **Assistant Professor**
Vienna University of Technology, Vienna.
Supervision of undergraduate students in programming, device simulation and quantum electronics
- 2016 **Assistant Professor**
Vienna University of Technology, Vienna.
Supervision of Master thesis “Trap Assisted Tunneling and Band Interaction using the Non-Radiative Multi Phonon Model” from Bernhard Stampfer
- 2011–2017 **Tutor**
Vienna University of Technology, Vienna.
Organization and supervision of several \LaTeX courses
- 2008 **Tutor**
Vienna University of Technology, Vienna.
Supervision of undergraduate students in the laboratory part of the course “Electrical Engineering 2” (Dipl.-Ing. Dr.techn. Peter Schönhuber)

Vocational

- 2005–2007 **Electrical engineer**
Lunatone, Vienna.
Development and design of equipment for functional testing of different products.
Functional testing and repair of products.

Internships

- June 2004 **Programmer**
Wertheim AG, Vienna.
Programming of database queries in SAP.
- Aug. 2001 **Electrician**
July 2000 *Elektro-Feymann GmbH, Stoob.*
July 1999 Electrical installations and wiring
- June 2000 **Electrical engineer**
Siemens AG, Amberg, Germany.
Manufacturing and repair of different types of programmable logic controllers
- July 1998 **Electrician**
BEWAG AG, Neutal.
Building of electrical house connections and laying of underground cables
- July 1997 **Engineer**
ABB AG, Neutal.
Product assembly

Voluntary activities

Member of the executive committee of the IEEE SB Vienna.

- Lecturer in several workshops (welding, \LaTeX)
- Co-founder of an electronics laboratory for students (IEEE Student-Lab)
- Organization of a Central European student branch meeting in Vienna
- Member of the IEEE student branch and Gold congress in London 2008

Member of the students representation at TU Wien

- Representation of graduate and undergraduate students at official committees
- Member of several committees for professorship castings
- Participation in several seminars for group dynamics and presentation techniques
- Tutor in voluntary courses for beginning students

Languages

German native speaker

English fluent (spoken and written)

Computer skills

Engineering software FreeCAD, Cadence, MATLAB, Spice, Sentaurus, Global TCAD Solutions

Programming skills Assembler, Bash, TCL/TK, Python, C, C++, Java

Operating systems Windows, Linux

Other Office, \LaTeX

Personal Interests

Sports tennis, football, bike riding

Engineering Invention, design and repair of electronic devices

Other Technical officer at the volunteer fire brigade in Stoob