Die approbierte Originalversion dieser Diplom-/Masterarbeit ist in der Hauptbibliothek der Technischen Universität Wien aufgestellt und zugänglich.
http://www.ub.tuwien.ac.at

The approved original version of this diploma or master thesis is available at the main library of the Vienna University of Technology.
http://www.ub.tuwien.ac.at/eng

DIPLOMARBEIT

Trap Assisted Tunneling und Bandinteraktion mit dem Non-Radiative Multi Phonon Modell

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines

Diplom-Ingenieurs

Unter der Leitung von:

Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Klaus-Tibor Grasser
und
Univ.Ass. Dipl.-Ing. Alexander Grill

E360 - Institut für Mikroelektronik

eingereicht an der Technischen Universität Wien,
Fakultät für Elektrotechnik und Informationstechnik

von

Bernhard Stampfer

0927191 / E 066 508

Humboldtgasse 11/2/11, 1100 Wien
E-Mail: `bernhard.stampfer@student.tuwien.ac.at`

Wien, Jänner 2016

DIPLOMA THESIS

Trap Assisted Tunneling and Band Interaction using the Non-Radiative Multi Phonon Model

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

Approved by:

Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Klaus-Tibor Grasser
and
Univ.Ass. Dipl.-Ing. Alexander Grill

E360 - Institut für Mikroelektronik

submitted to the Vienna University of Technology,
faculty of Electrical Engineering and Information Technology

by

BERNHARD STAMPFER

0927191 / E 066 508

Humboldtgasse 11/2/11, 1100 Vienna
E-Mail: `bernhard.stampfer@student.tuwien.ac.at`

Vienna, January 2016

Kurzfassung

Für die Zuverlässigkeit von Halbleiterbauelementen spielen auf Fehlstellen zurückführbare Phänomene wie Bias Temperature Instability (BTI) und Trap Assisted Tunneling (TAT) eine wichtige Rolle. BTI bezeichnet die reversible und irreversible Änderung der Threshold Spannung mit Temperatur und Bias Stress durch Elektroneneinfang und -emission. TAT bezeichnet Leckströme die ebenfalls durch Elektronenaustausch entstehen. Es wurde eine Reihe von Modellen zur Erklärung von BTI entwickelt. Ursprünglich wurde davon ausgegangen, dass BTI vom Ladungsträgereinfang von unabgesättigten Bindungen der Silizium Atome an der Silizium-Siliziumdioxid Grenzschicht und den von diesen Atomen dissoziierten Wasserstoffspezies resultiert. Neuere Erkenntnisse belegen, dass es sich jedenfalls beim reversiblen Teil um einen Einfang von Ladungen in bereits vorhandene Fehlstellen handelt und sich der Ladungsträgeraustausch durch non-radiative multi phonon (NMP) Übergänge beschreiben lässt.

In dieser Arbeit wird auf ein bestehendes NMP Modell aufgebaut. Die vorhandenen Ratengleichungen beschreiben Ladungsträgeraustausch zwischen Halbleitern oder Metallen mit Isolatoren. Zusätzliche, feldabhängige Ratengleichungen zum Leitungs- und Valenzband im eigenen Material sind notwendig für Simulationen von Fehlstellen in Halbleitermaterialien, z.B. für GaN/AlGaN high electron mobility Transistoren (HEMTs) und High-k Dielektrika. Eine Erweiterung am Modell ermöglicht die Berechnung statischer Tunnel- und transienter Umladeströme erzeugt durch einzelne Fehlstellen.

Mathematische Näherungen in den Ratengleichungen erlauben eine analytische Abschätzung der Raten für verschiedene Bereiche der Feldstärke. Die Temperaturabhängigkeit der Bandraten folgt einem Arrhenius-Gesetz, wobei die scheinbare thermische Barriere mit der Feldstärke sinkt. Ein Vergleich von Simulationen der Capture- und Emission-Zeiten mit und ohne den Bandraten zeigt für welche Situationen das Einbeziehen der Bandraten nicht mehr vernachlässigt werden kann. Mit den erweiterten Raten konnten die Leckströme im Sperrbetrieb in einem GaN/AlGaN HEMT simuliert werden. Die Ergebnisse zeigen ein Verhalten ähnlich dem Frenkel-Poole Modell für niedrige Felder und hohe Temperaturen, und ein Verhalten ähnlich dem Fowler-Nordheim Modell für hohe Felder und tiefe Temperaturen, qualitativ in guter Übereinstimmung mit Messergebnissen aus einschlägiger Literatur. Die transiente Simulation eines MOSFET zeigt Umlade- und Transportströme während Stress- und Erholungsphasen, und den Anteil einzelner Fehlstellen an diesen Strömen.

Abstract

Bias temperature instability (BTI) and trap assisted tunneling (TAT) play an important role for the reliability of semiconductor devices. Both phenomena can be traced back to defects. BTI denotes the reversible and irreversible change in threshold voltage with temperature and bias stress, TAT denotes leakage currents resulting from electron capture and emission of defects. A number of models aiming to describe BTI have been developed. Initially, it was assumed that BTI is a result of dangling bonds at silicon atoms near the silicon-silicon dioxide interface and the hydrogen species dissociating from these atoms. However, newer results show that at least the reversible part of BTI is caused by the capture of charge in pre-existing defects and that the charge transfer can be described by non-radiative multi phonon (NMP) transitions.

This work is based on an existing NMP model. The existing rate equations of the NMP model describe the charge exchange between semiconductors or metals and insulators. Additional, field dependent rate equations to conduction and valence band near the defect are necessary for simulation of traps in semiconducting materials, e.g. for GaN/AlGaN high electron mobility transistors (HEMTs) and transistors with high-k dielectrics. An extension to the model enables the simulation of static tunnel and transient displacement currents for individual traps.

Mathematical approximations in the rate equations allow an estimation of the rates for different regimes of the electric field. The temperature dependence of the band rates follows an Arrhenius-law, whereby the apparent thermal barrier decreases with field strength. A comparison of simulations of capture and emission times with and without band rates shows for which situations the band rates can not be neglected. With the extended rates reverse leakage currents in a GaN/AlGaN HEMT could be simulated. The results show Frenkel-Poole like behaviour for low fields and high temperatures and Fowler-Nordheim like behaviour for high fields and low temperatures in good agreement with measurement results in literature. The transient simulation of a MOSFET shows the displacement and transport currents during stress and recovery cycles, and the contribution for each of the traps.

Acknowledgment

At this point I want to thank everybody who made this work possible.

First and foremost, I want to thank my family, and especially my parents, Irene and Vinzenz Stampfer for supporting me, both mentally and financially, throughout my studies.

I'd like to thank my supervisors, Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Klaus-Tibor Grasser and Univ.Ass. Dipl.-Ing. Alexander Grill for their time supervising my work, providing input, helping me with questions I had when working on this thesis, and proof reading it.

I started studying at the Technische Universität Wien six years ago and made many friends here during that time, all of whom I'd like to thank for the time they shared with me, both in and out of the university. A special thank you goes to my girlfriend, Carmen Reischl, who is also part of this group.

I also want to thank the people working on the Institute for Microelectronics who are not directly involved in my work but whom I talked with during the work on this thesis. When I started my work Univ.Ass. Dipl.-Ing. Gerhard Rzepa and Projektass. Dipl.-Ing. Dr.techn. Wolfgang Gös, who worked with the four-state model before, gave me important input for this work. Benedikt Limbacher, a student who wrote his bachelor thesis with Gerhards supervision provided me with some useful python scripts. Later, Gerhard handed me a calibrated simulation set-up to use to test my implementation of transport currents in the simulator.

Finally, I want to mention some notable and interesting courses and lectures I visited during the course of my master studies and thank everybody involved. Most related to this work is of course "Modellierung elektronischer Bauelemente" which played an important part in my decision to do my master thesis at the Institute for Microelectronics, others were "Fourier Optics", "Nonlinear Optics", "Photonik 2" and "Quantenelektronik".

Contents

List of Figures **ix**

List of Tables **xi**

List of Acronyms **xiii**

1 Introduction **1**
- 1.1 Bias Temperature Instability . . . 1
  - 1.1.1 Characterization of BTI . . . 2
  - 1.1.2 Modeling of NBTI . . . 3
- 1.2 Trap Assisted Tunneling . . . 5
  - 1.2.1 Reverse Leakage in HEMTs . . . 6
  - 1.2.2 Stress Induced Leakage Current . . . 7

2 Theory **9**
- 2.1 Markov Chains . . . 9
- 2.2 The Four-State Model . . . 10
  - 2.2.1 Trap State Occupancies and Charge . . . 11
  - 2.2.2 Carrier Transport, Trap Assisted Tunneling . . . 12
- 2.3 Non-Radiative Multi Phonon Transitions . . . 14
  - 2.3.1 NMP Transitions . . . 14
  - 2.3.2 Transitions to a Band of States . . . 16
  - 2.3.3 Band Edge Approximation . . . 17
  - 2.3.4 Field-assisted Transitions . . . 18
  - 2.3.5 Practical Considerations For Implementation . . . 23
- 2.4 Thermal Transitions . . . 29

3 Results **31**
- 3.1 Capture and Emission Rates . . . 31
  - 3.1.1 Field Dependence . . . 35
  - 3.1.2 Temperature Dependence . . . 43
  - 3.1.3 Energy Distribution . . . 45
- 3.2 Significance of the Local Band Interactions . . . 49

3.3 Trap Assisted Tunneling 55
  3.3.1 Reverse Leakage Currents in a HEMT 55
  3.3.2 Stress and Recovery Currents in a MOS Transistor 62

4 Summary, Outlook **69**

Bibliography **71**

List of Figures

1.1 Gate voltage and measurement points for the extended MSM method . . 3
1.2 Schematic drawing of the processes happening in the reaction-diffusion model, adopted from [2] and [8] 4
1.3 Band diagrams for trap assisted tunneling models, adopted from [15] . . 5
1.4 Schematic geometry of a HEMT and possible leakage paths 7

2.1 Markov chains of a two-state model and the four-state model 10
2.2 Band diagram of an example GaN/AlGaN heterostructure 13
2.3 Adiabatic potential energy surfaces in the configuration coordinate diagram for example states i and j 15
2.4 Transition from a band of states to a trap in the configuration coordinate diagram and the band diagram 16
2.5 Possible paths of interaction between trap and bands 19
2.6 Seperated processes happening during emission of an electron to the conduction band 20
2.7 Adiabatic potential energy surfaces in the configuration coordinate diagram for two example states i and j 23
2.8 Band diagrams showing triangular and trapezoidal tunneling barriers and the parameters used to describe them. 25
2.9 Band diagrams showing the dependence of carrier concentration n on the slope of the fermi level for two special cases. 28
2.10 Configuration coordinate diagram for a thermal transition between stable and meta-stable states j and i 29

3.1 Test structure used in section 3.1 32
3.2 Adiabatic potential energy surfaces in the configuration coordinate diagram for the trap parameters $R_{12} = 0.6$ and $S_{12}\hbar\omega = 2.3\,\mathrm{eV}$ 33
3.3 Plot of the thermal and field-dependent parts of the rate over the electric field 33
3.4 Band diagrams of the test structure in voltage and field limited regions . 34
3.5 Plots of the capture and emission rates k'' over F 35
3.6 Plots of the capture and emission rates k'' and approximations made . . 42
3.7 Arrhenius plots showing the capture and emission rates 43

3.8 Apparent activation energies for emission and capture 44
3.9 Energy distribution of the integrand $P_T\lambda$ of the capture rate 45
3.10 WKB factors and probabilities of thermal excitation for the capture rate 46
3.11 Energy distribution of the integrand $P_T\lambda$ of the emission rate 47
3.12 WKB factors and probabilities of thermal excitation for the emission rate 48
3.13 Test structure used in section 3.2 49
3.14 Simulated capture and emission times. Trap at a fixed energy, 6 nm dielectric 51
3.15 Simulated capture and emission times. Trap at a fixed energy, 3 nm dielectric 52
3.16 Simulated capture and emission times. The trap has a fixed offset to the band, 6 nm dielectric 53
3.17 Simulated capture and emission times. The trap has a fixed offset to the band, 3 nm dielectric 54
3.18 Test structure used in section 3.3.1 56
3.19 Tunneling current through the barrier layer of a GaN/AlGaN heterostructure 57
3.20 Comparison of tunnel currents simulated with the NMP model and the Fowler-Nordheim model 58
3.21 Comparison of tunnel currents simulated with the NMP model and the Frenkel-Poole model 60
3.22 Plots of the leakage currents resulting from the NMP model, the Poole-Frenkel and the Fowler-Nordheim model 61
3.23 Geometry of the MOSFET used in section 3.3.2 62
3.24 Gate, source, and drain voltages V_g, V_s and V_d during MSM measurement. 63
3.25 Currents during stress 63
3.26 Currents during recovery 64
3.27 Currents during stress for multiple stress times 65
3.28 Currents during recovery for multiple stress times 65
3.29 Band diagram showing the carrier transport to the individual traps during stress 66
3.30 Band diagram showing the carrier transport from the individual traps during recovery 67
3.31 Band diagram showing traps contributing to the carrier transport 67
3.32 Energetic and spatial distribution of transport rates of individual traps . 68

List of Tables

3.1 Simulation parameters for the discussion of capture and emission rates in section 3.1. 32

3.2 Field dependences of rate k'' 41

3.3 Simulation parameters for the evaluation capture and emission times with and without local band interaction in section 3.2 50

3.4 Simulation parameters for the discussion of TAT in a HEMT in section 3.3.1. 56

3.5 Simulation parameters for the discussion of trap currents in a MOSFET in section 3.3.2 62

List of Acronyms

BTI	bias temperature instability
CMOS	complementary metal-oxide-semiconductor
CTMC	continous-time Markov chain
DTMC	discrete-time Markov chain
FET	field effect transistor
FN	Fowler-Nordheim
FSM	four-state model
HEMT	high electron mobility transistor
MIS	metal-insulator-semiconductor
MOS	metal-oxide-semiconductor
MOSCAP	metal-oxide-semiconductor capacitor
MOSFET	metal-oxide-semiconductor field effect transistor
MSM	measure-stress-measure
NBTI	negative bias temperature instability
nMOS	n-type field effect transistor
NMP	non-radiative multi phonon
OTF	on-the-fly
PBTI	positive bias temperature instability
pMOS	p-type field effect transistor
PF	Poole-Frenkel
RD	reaction-diffusion
SILC	stress induced leakage current

SRH	Shockley-Read-Hall
TAT	trap assisted tunneling
TSM	two-stage model
TDDS	time-dependent defect spectroscopy
WKB	Wentzel, Kramers, Brillouin

CHAPTER 1 Introduction

I want to begin this work with a short presentation of two important phenomena for semiconductor reliability, namely bias temperature instability (BTI) and trap assisted tunneling (TAT). BTI as a phenomenon, characterization of BTI and modeling of BTI will be summarized in section 1.1. TAT as the mechanism responsible for reverse bias leakage in high electron mobility transistors (HEMTs) and stress induced leakage current (SILC) will be presented in section 1.2.

This work will focus on the numerical simulation of defects in semiconductor devices, generally attributed to be the cause of these phenomena. Simulations will be done in the framework of the four-state non-radiative multi phonon (NMP) model. The model will be extended to allow for charge exchange between defects and the conduction or valence bands close to the defect and the calculation of currents resulting from the defects. The model and its extensions will be covered in the theory chapter, chapter 2 of this work.

The behaviour of the extension and its significance will be discussed in the first part of the results chapter, chapter 3, before simulations on realistic devices will be done. Simulations of a GaN/AlGaN HEMT will yield reverse bias leakage currents which will be compared to currents resulting from the Fowler-Nordheim and Frenkel-Poole models. A MOSFET will be simulated to show displacement and transport currents resulting from oxide traps.

1.1 Bias Temperature Instability

Bias temperature instability (BTI) as a phenomenon has been known for a long time and was first described in 1966 [1]. Only in recent years however, it has gained more interest. Reasons are, among other circumstances, the introduction of nitrogen into the gate oxide of MOS transistors [2] and the higher electric fields in the oxide due to miniaturization [3]. Lateral scaling of devices decreases the number of defects and increases the influence of single defects and the distribution of their properties on degradation. BTI is a major concern to the reliability of microelectronic devices.

Bias temperature instability refers to the change in the threshold voltage of a MOS device as it is biased in inversion. Its name indicates its sensitivity to temperature and bias voltage, both of which accelerate the degradation of the device. BTI can be broken down further into positive BTI (PBTI) and negative BTI (NBTI). NBTI refers to the instability observed at negative gate bias and PBTI to the instability at positive bias. NBTI is most prominent in p-channel MOSFETs operating in CMOS circuits and PBTI in high-k nMOS devices. While both effects cause an increase in the absolute value of the threshold voltage, NBTI further affects the subthreshold behaviour of the device and causes g_m degradation. This is attributed to different locations of charge trapping in n- and pMOS devices [2]. The defects responsible for BTI in pMOS devices are close to the channel, this leads to strong Coulomb scattering which degrades the mobility of the carriers there.

1.1.1 Characterization of BTI

A number of experimental methods to characterize BTI have been developed during recent years. Two of them, the commonly used measure-stress-measure (MSM) method and time dependent defect spectroscopy (TDDS) will be summarized in this section.

The Measure-Stress-Measure (MSM) method, also called the stress-and-sense method, is a frequently used characterization method for BTI and can be used with MOSFET and MOSCAP devices [2]. Before stressing, the device or a similar one is characterized. Then the first stress cycle starts and stress is applied for a defined time. After stress, the threshold voltage shift is measured and the next stress cycle is started. The delay between stressing the device and measuring should be as short as possible, because recovery of BTI starts immediately after stress is removed. This has led to the development of short measurement methods such as ultra short pulse I_d-V_g measurements [4] or spot C-V measurements (for MOSCAP structures). With these methods, instead of doing a full characterization of the device after stress, only a single measurement is taken at a reference voltage and compared to the reference curve of the device to evaluate the threshold voltage shift.

If the dynamics of BTI recovery is of interest, recovery measurements of the drain current as proposed in [5], with logarithmic distributed sampling times can be done between the stress cycles. This is illustrated in figure 1.1. A similar measurement will be simulated in **section 3.3.2** of the results chapter to evaluate displacement and transport currents occurring during stress and recovery.

To gain a better understanding of the mechanisms of NBTI, a method called Time-Dependent Defect Spectroscopy (TDDS) is used. The first measurement was reported in 1988 but only recently the technique has gained interest [2]. It allows to learn about the behaviour of individual defects and their position within a device.

Time-Dependent Defect Spectroscopy (TDDS) is a measurement method allowing to monitor charge capture and emission times for individual defects. For small area devices with a small number of defects it is possible to see discrete steps in the drain current each time a defect changes its charge state. The individual step height

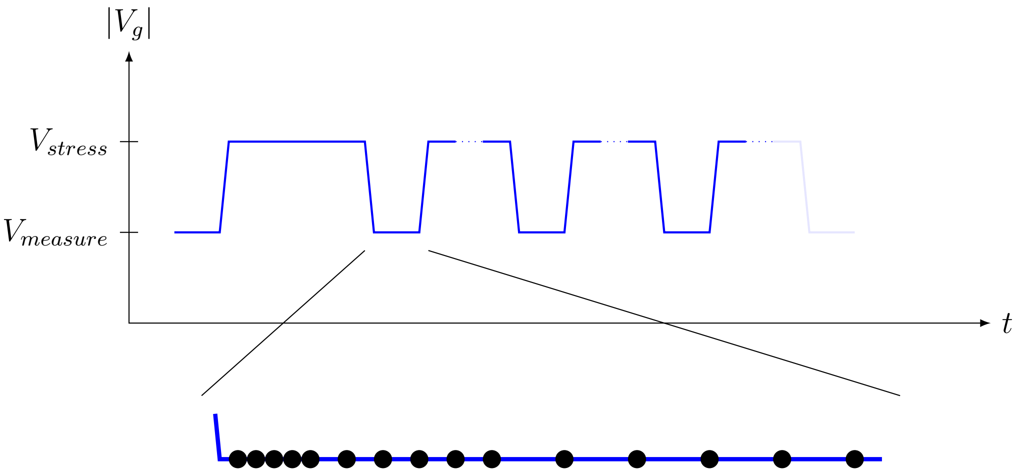


Figure 1.1: Gate voltage and measurement points for the extended MSM method. The sampling rate during measurement cycles is increased logarithmic to collect as much information on the recovery as possible in the measurement phase. Figure adopted from [5].

each defect shows depends on its position in the oxide. Together with its specific, bias and temperature dependent time constants this allow to map the observed events to individual defects.

To stimulate capture and emission events the device is repeatedly stressed and recovered. During recovery emission events can be directly observed in the drain current of the device. The capture events can not be observed directly because of high noise and limited resolution in the strong inversion regime where the device is stressed [2]. To evaluate the capture times the width of the stress cycle is varied and the capture events for each stress time counted. As the events are stochastic in nature, the measurements have to be repeated until the confidence intervals are small enough to extract distribution parameters.

The behaviour of capture and emission times observed for different stress voltages and temperatures provides information that have led to a better understanding of the physical nature of the defects.

1.1.2 Modeling of NBTI

In 1977, the first model for NBTI, the **reaction-diffusion (RD) model** [6] was developed. In the RD model, it is assumed that hydrogen atoms, introduced during manufacturing of the device are responsible for the degradation. The hydrogen atoms are introduced to passivate dangling silicon bonds at the interface between silicon and SiO_2, which otherwise create defect states at the interface. During stress, the hydrogen atoms dissociate from the silicon atoms (reaction) and then diffuse towards the oxide (diffusion). For low stress times, the amount of hydrogen atoms dissociated is assumed to be reaction limited and increases linear in this regime. At some point, the rate becomes diffusion limited and the de-passivated bonds increase with $t^{1/4}$. A schematic drawing of the processes happening in the RD model can be found in figure 1.2.

The RD model was modified as later experimental results did not agree with the initial power-law exponent. Recent publications questioned the validity of the model [7] and the suitability of reaction rate equations for the particle densities involved. It was since realized that the RD model can not explain the BTI, especially the recovery behaviour and the decorrelated capture and emission time constants of the defects. TDDS measurements have shown charging and discharging of individual pre-existing defects as the mechanism behind BTI.

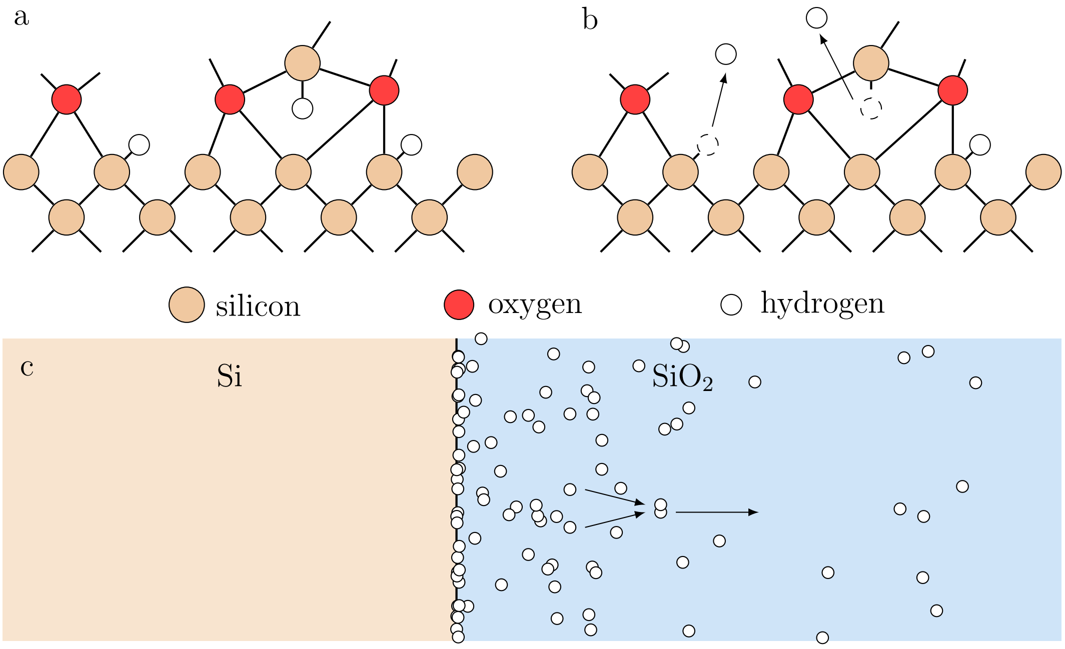


Figure 1.2: Schematic drawing of the processes happening in the reaction-diffusion model, adopted from [8]. Dangling bonds of silicon atoms at the Si-SiO_2 interface are passivated with hydrogen during manufaction (a). During stress, the hydrogen atoms dissociate from the silicon (reaction) and leave positively charged silicon ions (b). The hydrogen atoms diffuse away and some form H_2 (diffusion, c). Early during stress, the reaction is the rate-limiting factor, later the diffusion of hydrogen and subsequent conversion to H_2 limits the rates.

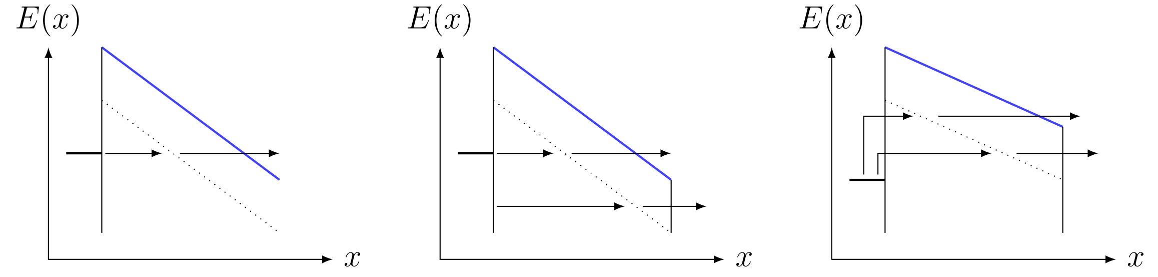
Realization that charge trapping in pre-existing oxide defects might be responsible for BTI have led to the development of new models. Early models were based on elastic tunneling [9, 10]. Later, modified versions of the Shockley-Read-Hall (SRH) model to account for tunneling [11] were used and phenomenologically extended with energy barriers by **Kirton and Uren** [12]. It was understood that non-radiative multi phonon transitions are responsible for charging and discharging of the traps. The **two-stage model (TSM)** [13] phenomenologically introduced a field enhancement factor, but the model was still not correctly reflecting the nature of microscopic trapping [14]. This led to the development of the **extended non-radiative multi phonon (eNMP) model** [14]. It has additional metastable states, necessary for the description of the observed weakly correlated capture and emission times and the bias dependence observed with TDDS. Its NMP charge transfer reaction is strictly derived from quantum mechanics.

The extended NMP model will be covered in the theory chapter, **chapter 2**, of this work. The model, currently describing interaction of the traps with the bands of the adjacent semiconductor or metal layers, will be extended to allow interactions with the conduction and valence band near the trap position in **section 2.3.4**. While these interactions are not relevant in SiO_2 due to its high band gap energy, they can become relevant for traps in wide band gap and high-k materials such as Gallium Nitride (GaN) and Silicon Nitride (Si_2N_3). An evaluation of the effect of these rates on the capture and emission times depending on the band gap and trap energies can be found in the results **section 3.2**.

1.2 Trap Assisted Tunneling

Trap Assisted Tunneling (TAT) describes the transport of charges between two points by capture and emission of a trap. It is regarded as the mechanism for leakage through nitrided and tunnel oxides, for stress-induced leakage currents (SILC) and at least partially for the reverse leakage of GaN/AlGaN high electron mobility transistors (HEMTs) [15].

A variety of analytical models have been developed for trap assisted tunneling. Fleischer and Lai [16] developed a model for TAT currents through triangular barriers, Houng et al. [17] a generalized model for triangular and trapezoid barriers. Sathaiya and Karmalkar [15] extended these models further by including thermally activated electrons. Illustrations of the tunneling paths for these models can be found in figure 1.3.



(a) Fleischer and Lai (b) Houng et al. (c) Sathaiya and Karmalkar

Figure 1.3: Band diagrams for trap assisted tunneling models, adopted from [15]. The blue line is the conduction band edge of the layer containing the traps, the horizontal black line on the left indicates the fermi niveau in the metal contact. The dotted line represents the traps and the arrows the charge transport.

For HEMTs, the measured leakage currents are often fitted with the popular Frenkel-Poole [18] and Fowler-Nordheim [19] models.

The Fowler-Nordheim model describes the current emitted from a metal through a triangular barrier and is usually written in the form

$$J_{FN} \propto F^2 e^{-\frac{A}{F}} \tag{1.1}$$

with a parameter A. It shows a low temperature dependence and is usually used to fit the leakage current at low temperatures and high fields.

The Frenkel-Poole model describes the current produced by carriers thermally emitted from a trap state to a band with the lowering of the thermal barrier by an electric field. The resulting current density can be written as

$$J_{PF} \propto F e^{-\frac{q\Phi_t - B\sqrt{F}}{k_B T}} \tag{1.2}$$

with parameters B and the trap depth Φ_t. Its temperature dependence follows an Arrhenius law and it is used to fit leakage currents at high temperatures and low fields. These formulas will be used in section 3.3.1.

In this work, TAT currents are calculated directly in the framework of the NMP model. This allows to calculate the transient displacement and static transport currents for each individual trap in the device. In **section 3.3** of the results chapter, simulations for two TAT scenarios, one of them also using the rate equations to the semiconductor bands, will be discussed.

1.2.1 Reverse Leakage in HEMTs

Some III-V semiconductor high electron mobility transistors, e.g. GaN/AlGaN HEMTs, show large gate leakage currents in reverse bias. Possible leakage paths are the vertical path from the gate to the channel and the lateral path from gate to drain or source. This is illustrated in figure 1.4. According to [20], the vertical part dominates over the lateral part in III-V nitride HEMTs. This is in contrast to other materials where the lateral path dominates.

While the exact physical mechanisms for the leakage current are not known, it does not seem to be direct tunneling. Unrealistic assumptions of the charge at the interface have to be made to fit the measurements [21] and the temperature dependence of the measured currents does not agree with direct tunneling. This leaves some form of trap assisted tunneling as a likely candidate for the leakage currents.

In **section 3.3.1** of the results chapter, simulated reverse leakage currents in a GaN/AlGaN device will be compared against the Frenkel-Poole and Fowler-Nordheim models

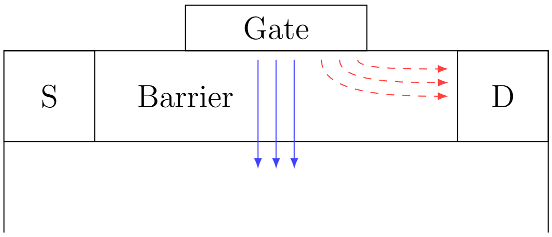


Figure 1.4: Schematic geometry of a HEMT and possible leakage paths. The vertical path is indicated by solid blue arrows, the lateral path with dashed red arrows. S and D denote Source and Drain.

often used to fit to the measured leakage currents in these devices. The simulation was done using the NMP model with the rates to the local conductance and valence bands derived in section 2.3.4.

1.2.2 Stress Induced Leakage Current

A phenomenon possibly linked to NBTI is Stress Induced Leakage Current (SILC). As the name suggests, SILC describes a leakage current through the oxide of a MOS device after high field stress. SILC is a concern for the reliability of devices with an oxide thickness smaller than about 10 nm, especially for non volatile memory devices [22]. The SILC has also been attributed to trap assisted tunneling [23].

In **section 3.3.2** of the results chapter, we will take a look at the transient and static trap assisted leakage currents simulated for a stressed MOSFET to gain understanding of the energetic and spatial distribution of defects causing TAT currents in thin oxides.

CHAPTER 2

Theory

In this chapter the four-state NMP model used in this work will be described and extended. First, in section 2.1 a short introduction to Markov chains will be given, as a principal understanding will be necessary later. In section 2.2 the basic principle of the four-state model will be outlined, without detailed explanation of the transition rates between states. It will be shown how transport and charging currents can be calculated in the framework of the model. Later, in section 2.3, the physical theory behind the charge trapping, the non-radiative multi phonon (NMP) theory and the derivation of the charge transfer rates used in the four-state model will be summarized. The model will be extended to cover field-assisted charge transfer to the local conduction and valence bands. At the end of the chapter, in section 2.4, thermal transitions, used in the four-state model for transitions without charge transfer, will be covered.

2.1 Markov Chains

Before advancing to the trap model in the next section, a short introduction to Markov chains, commonly used to describe traps in semiconductors, should be given.

Markov chains can be split in discrete-time Markov chains (DTMC) and continous-time Markov chains (CTMC). As there is no time discretization in the physical systems investigated in this work, we will focus on the continous-time type.

A Markov process is a random process that retains no memory of its past. This is called the Markov property. It means that future states of the process are only influenced by its current state and not by the past. A Markov chain is a Markov process with a finite set of states.

Mathematically, a continuous time Markov chain has a state space $\mathbf{I}$ with a finite set of states i and a transition matrix $\mathbf{K}$ with a dimension equal to the number of states and row sums of zero. The process $(X_t)_{t\geq 0}$ is then defined (infinitesimal definition) [24] for $h \to 0$ as

$$\mathbb{P}(X_{t+h} = j | X_t = i) = \delta_{ij} + k_{ij} h + o(h). \tag{2.1}$$

In the following, the occupancies of the states will be treated probabilistically. The probabilities $\vec{P}$ of the system occupying one of the states i can then be written as a master equation, a set of differential equations.

$$\frac{d\vec{P}(t)}{dt} = \mathbf{K}\vec{P}(t) \tag{2.2}$$

2.2 The Four-State Model

Traps in semiconductor devices are commonly described using two-state Markov chains with a neutral and a charged state. Time dependent defect spectroscopy (TDDS) measurements in oxides [25] have shown trapping behaviour which can not be described by the dynamics of a two-state model. The four-state model has an additional metastable state in the Markov chain for each charge state of the trap. With these additional states, many observed experimental features can be explained. Visual representations of the two- and four-state Markov chains can be found in figure 2.1.

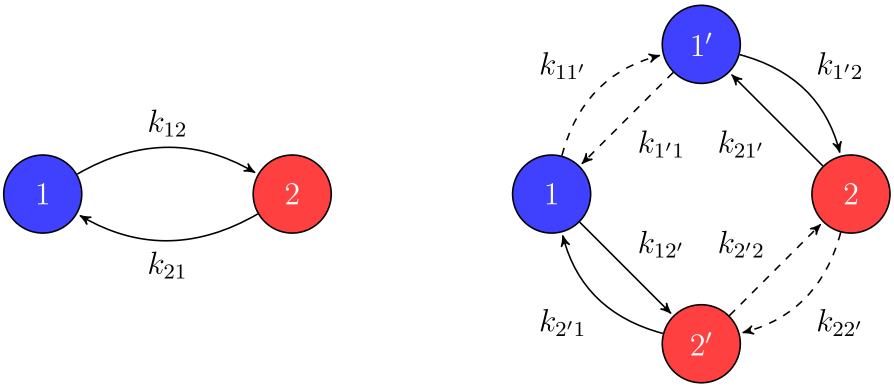


Figure 2.1: Markov chains of a two-state model (left) and the four-state model (right). States denoted as 1 and 2 are the charged and uncharged states respectively, metastable states are dashed. Transitions involving a charge transfer are drawn solid (NMP transitions), those without (thermal transitions) dashed.

The stable states in the four-state model will in the following be denoted as 1 and 2 for the charged and uncharged states respectively, metastable states will be marked with dash ($'$). Depending on the type of trap, the charged states 2 and $2'$ in this model have a different physical meaning. For hole traps, the charged states indicate the more positive state of the trap, while for electron traps it represents the more negative one.

Transitions involving a change transfer, 1-$2'$ and $1'$-2, are modeled as non-radiative multi phonon transitions (see section 2.3), transitions without charge transfer, 1-$1'$ and 2-$2'$, as pure thermal transitions (section 2.4). The physical meaning of transitions between the states change with trap type as the carrier involved in the transition is either an electron or a hole.

To calculate the transition rates between states, the trap model has a set of parameters describing a parabolic adiabatic potential energy surface in the configuration coordinate diagram for each configuration of the trap. This will also be explained in section 2.3.

Unless stated otherwise, in the following sections electron traps will be assumed.

2.2.1 Trap State Occupancies and Charge

With the definition of the Markov chain and the transition rates, expressions for the trap occupancies, trap charge, capture/emission times and charge transport can be found. All these quantities depend on time, but time dependence is not stated explicitly, unless necessary to shorten formulas and improve readability, e.g. $p_n = p_n(t)$.

As the Markov chain has to be always in one of the states i, the occupation probabilities of the states have to add up to one at any time.

$$\sum_m p_m = 1 \tag{2.3}$$

Initial conditions must be chosen accordingly. Usually one of the stable states is chosen as the initial state and its occupation probability set to one, the occupation of the other states to zero.

The change in occupancy of a trap state n can then be calculated by adding the rates to and from connected states m multiplied by the respective occupation probability p as in equation 2.2.

$$\frac{dp_n}{dt} = \sum_{m \neq n} (p_m k_{mn} - p_n k_{nm}) \tag{2.4}$$

Here, k_{nm} are the rates from state n to m. For the Markov chain in the four-state model the resulting equation system is:

$$\dot{p}_1 = p_{1'} k_{1'1} + p_{2'} k_{2'1} - p_1 (k_{11'} + k_{12'}) \tag{2.5a}$$
$$\dot{p}_{1'} = p_1 k_{11'} + p_2 k_{21'} - p_{1'} (k_{1'1} + k_{1'2}) \tag{2.5b}$$
$$\dot{p}_2 = p_{1'} k_{1'2} + p_{2'} k_{2'2} - p_2 (k_{21'} + k_{22'}) \tag{2.5c}$$
$$\dot{p}_{2'} = p_1 k_{12'} + p_2 k_{22'} - p_{2'} (k_{2'1} + k_{2'2}) \tag{2.5d}$$

The dot above the occupancies p_n denotes time derivatives here, i.e. $\dot{p}_n = \frac{dp_n}{dt}$.

The occupancies in equilibrium can be calculated by setting the all time derivatives to 0 i.e.

$$\dot{p}_1 = \dot{p}_{1'} = \dot{p}_2 = \dot{p}_{2'} = 0 \tag{2.6}$$

and solving the resulting equation system.

The trap charge can be found by summing the occupancy of the charged states m and multiplying them by the charge of the carriers involved $q_c = \pm q_0$.

$$q = q_c \sum_m p_m \tag{2.7}$$

A more in detail description of the four-state model, including the capture and emission times can be found in [26].

2.2.2 Carrier Transport, Trap Assisted Tunneling

A trap in the four-state model is generally not limited to a single reservoir within the device, but can exchange charge with all of the surrounding materials. While doing so, it not only changes its own charge, but also transports charge between the reservoirs. The rates k_{nm} that govern the NMP transitions in the Markov chain are then sums of individual exchange rates $k_{nm,u}$ between the trap and the corresponding reservoir u.

$$k_{nm} = \sum_u k_{nm,u} \tag{2.8}$$

To illustrate this, a sample configuration is shown in figure 2.2. The rate $k_{12'}$ (charge trapping to the metastable charged state) for example, would be composed of a number of individual rates:

$$k_{12'} = k_{12',contact} + k_{12',cb,barrier} + k_{12',vb,barrier} + k_{12',cb,bulk} + k_{12',vb,bulk}.$$

In this example, the bulk and contact rates are NMP rates to adjacent materials from the existing NMP model. Rates $k_{12',cb,barrier}$ and $k_{12',barrier}$ are NMP rates to the local conduction and valence bands, which will be derived as part of this work in section 2.3.4.

The resulting sum rates k_{nm} are used to calculate the occupancies of the trap states as explained in section 2.2.1. With the occupancies of the states, the rate of carrier transport r_u to each exchange point u can be calculated.

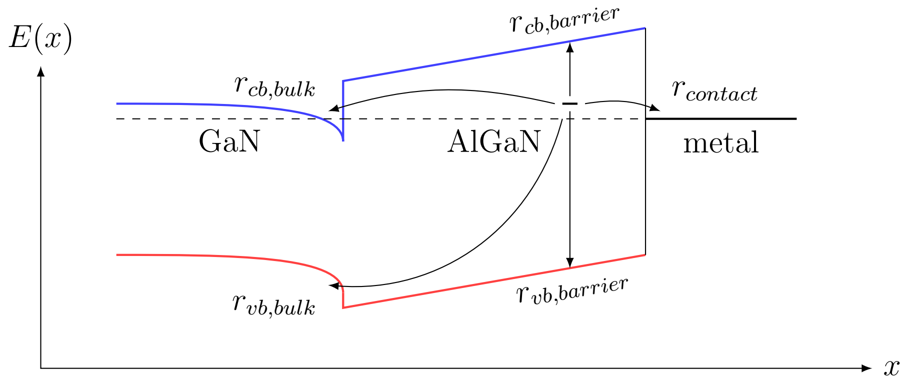


Figure 2.2: Band diagram of an example GaN/AlGaN heterostructure. The arrows show the possible paths of charge exchange with each reservoir. The trap inside the semiconducting barrier can exchange charge with the local conduction (blue) and valence bands (red), the bands of the bulk material and the metal contact.

$$r_u = \sum_{nm} (p_n k_{nm,u} - p_m k_{mn,u}) \tag{2.9}$$

For the four-state model, indices nm refer to the transitions from the charged states $n = 2$ and $2'$ to the uncharged states $m = 1$ and $1'$. The rate r_u to the reservoir u can be expressed as a current i_u to the reservoir simply by multiplying it by the charge of the involved carriers q_c.

$$i_u = q_c r_u \tag{2.10}$$

The currents must obey Kirchhoffs law, i.e. the sum of the currents to the trap must be equal to zero in equilibrium. More general, the sum of the currents must be equal to the change in trap charge with time. This can be shown by doing the sum over all points of exchange and substituting back equations 2.4, 2.7, 2.8 and 2.9.

$$\begin{aligned}
\sum_u i_u &= q_c \sum_u r_u \\
&= q_c \sum_u \sum_{nm} (p_n k_{nm,u} - p_m k_{mn,u}) \\
&= q_c \sum_{nm} (p_n k_{nm} - p_m k_{mn}) \\
&= -q_c \sum_n \dot{p}_n \\
\sum_u i_u &= -\dot{q}
\end{aligned}$$

2.3 Non-Radiative Multi Phonon Transitions

In the previous section we described the Markov chain used to model the behaviour of a trap and showed expressions for its charge, occupancies and carrier transport. The key processes that govern the behaviour of the Markov chain however, the transition rates k_{nm} between the Markov states, are still left to explain. In the next two sections, expressions for these transitions will be given. This section will focus on the NMP transitions between the charged and uncharged states of the trap.

The first part of this section will focus on the derivation of transfer rates of electrons between a single trapped and an untrapped state. In the second part this rate will be extended for transitions between the trapped state and a remote band of states, to describe interaction of the trap with adjacent materials. Then, in the third part, transitions between the trapped state and a local band of states will be derived for traps in semiconducting materials or dielectrics with reachable band states. Finally, the rates obtained will be simplified to yield analytical expressions suitable for use in device simulation.

2.3.1 NMP Transitions

To describe the transfer of charge to or from a trap, commonly called the charge transfer reaction, non-radiative multi phonon theory is used in the four-state NMP model. It can be derived directly from quantum mechanics.

To start the derivation, the behaviour of the system consisting of the trap and all involved atoms is expressed by their Schrödinger equation. The Schrödinger equation is then split into separate equations for electrons and nuclei, using the Huang-Born approximation, which uncouples electronic from nuclei states (adiabatic approximation) and thus allows for the separation [2].

$$\left(\hat{T}_e + \hat{V}_{ee}(\mathbf{r}) + \hat{V}_{en}(\mathbf{r};\mathbf{R}) + \hat{V}_{nn}(\mathbf{R})\right)\varphi_i(\mathbf{r};\mathbf{R}) = V_i(\mathbf{R})\varphi_i(\mathbf{r};\mathbf{R}) \tag{2.11}$$

$$\left(\hat{T}_n + \hat{V}_i(\mathbf{R})\right)\eta_{i\alpha}(\mathbf{R}) = E_{i\alpha}\eta_{i\alpha}(\mathbf{R}) \tag{2.12}$$

The terms in the Hamiltonians are interaction ($\hat{V}$) and kinetic contributions ($\hat{T}$), indices e and n indicate electrons and nuclei respectively. The separated wave functions are termed $\varphi_i(\mathbf{r};\mathbf{R})$ for the electronic system and $\eta_{i\alpha}(\mathbf{R})$ for the system of nuclei (vibrational system). $V_i(\mathbf{R})$ is the solution of the electron system, commonly termed adiabatic potential energy, and is dependent on the configuration of the nuclei. $E_{i\alpha}$ is the energy of the nuclear system. Indices i and $i\alpha$ indicate the quantum states of the electronic and vibrational systems.

The coordinates $\mathbf{R}$ of all nuclei involved in the transition can be reduced to a single configuration coordinate q, a generalized coordinate describing the positions in their trajectories during the reaction. The adiabatic potential energy $V(q)$ plotted over the configuration coordinate q is called configuration coordinate diagram. For sufficiently small changes in the position of the involved atoms, the adiabatic potential can be approximated as parabolic. Approximated sample adiabatic potential energy surfaces with parameters V, c, and q in the configuration coordinate diagram are shown in figure 2.3. The relative position and the shapes of the parabolas in the configuration coordinate diagram depend on intrinsic properties of the atomic structure, their relative energetic position is modified by the carrier energy E of the reservoir state [27].

The parabolic shape of the approximated system in the configuration coordinate diagram allows to model it as a quantum harmonic oscillator. To derive the transition probability from an electronic state i to j, perturbation theory along with the Frank-Condon approximation is used [2, 28, 29].

$$k_{ij} = A_{ij} f_{ij} \tag{2.13}$$

The transition rate k_{ij} from state i to j can be expressed as the product of an electronic matrix-element A and a lineshape function f.

$$A_{ij} = \frac{2\pi}{\hbar} |\langle \varphi_i | V' | \varphi_j \rangle|^2 \tag{2.14}$$

$$f_{ij} = \underset{\alpha}{ave} \sum_{\beta} |\langle \eta_{i\alpha} | \eta_{j\beta} \rangle|^2 \tag{2.15}$$

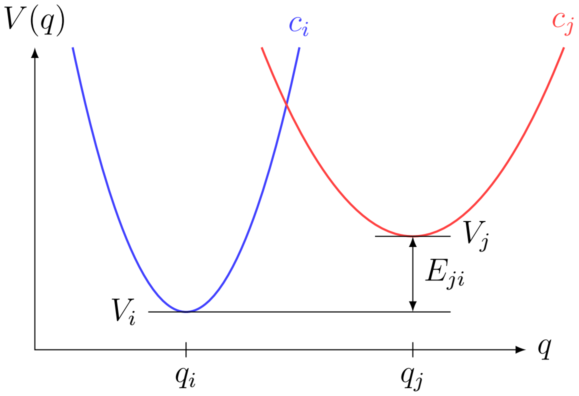


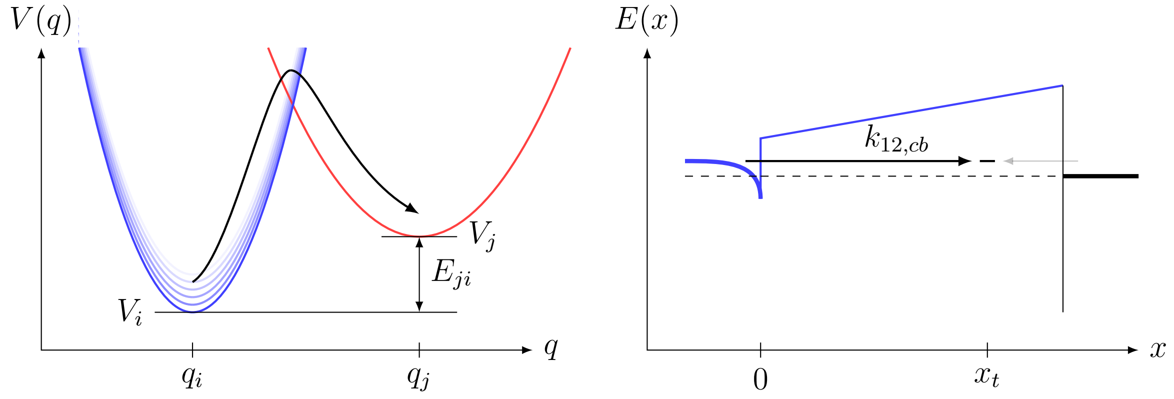
Figure 2.3: Adiabatic potential energy surfaces in the configuration coordinate diagram for example states i and j. V_i and V_j are the energy minimas of the parabolas, q_i and q_j the configuration coordinates at the minima. c_i and c_j are the curvatures of the parabolas. The lineshape function A_{ij} will later be calculated from these parameters and the energy difference $E_{ji} = V_j - V_i$.

The lineshape function $f(\hbar\omega)$, evaluated at $\hbar\omega = 0$ for non-radiative transitions, is the thermal average (denoted ave) over all probabilities for starting in a vibrational state $i\alpha$ and ending in any vibrational state $j\beta$. A_{ij} is the matrix element of the electron wave functions of the undisturbed eigenstates i and j with the pertubation V'. It is a measure for the electronic transition probability. A and f will later be approximated by simple analytical expressions for implementation in device simulators.

In the following, f will be written parametrized as a function of the energy difference between the final and the starting position of the electron, i.e. $f_{ij}(E_j - E_i)$. The curvatures and positions of the parabolas are not stated explicitly. They are described by the starting and end state i and j. A will be written as a function of the electron energy and either the tunnel distance or the electric field, $A_{ij}(E, x_t)$ or $A_{ij}(E, F)$.

2.3.2 Transitions to a Band of States

Equation 2.13 gives the transition rates from any state i to j. To calculate transitions between a trap state and a whole band of states, as shown in figure 2.4a, the rates are integrated over a continuum of band states (equation 2.16a). For simplicity, the following equations are given for two-state electron traps. Starting configurations for the lineshape function and the matrix element will always be denoted as i, final states as j. This section and the following section on the band edge approximation are based on [26] and [2]. All equations are for electron traps, indices 1 and 2 change for hole traps.



(a) Adiabatic potential energy surfaces in the configuration coordinate diagram for an example state j and a band of states i.

(b) Band diagram showing the capture of an electron from the conduction band of a neighbouring layer and the resulting trapezoidal tunneling barrier.

Figure 2.4: Transition k_{12} from a band of states to a trap in the configuration coordinate diagram (a) and the band diagram (b).

Equation 2.16a gives the rate of electron capture from the conduction band to the trap. D_n is the density of states in the conduction band and f_n is the electron occupancy, the factor $D_n f_n$ is the effective density of states at a specific energy.

$$k_{12,cb} = \int_{E_c}^{\infty} D_n(E) f_n(E) A_{ij}(E, x_t) f_{ij}(E_t - E)\, dE \tag{2.16a}$$

$$k_{21,cb} = \int_{E_c}^{\infty} D_n(E)(1 - f_n(E)) A_{ij}(E, x_t) f_{ij}(E - E_t)\, dE \tag{2.16b}$$

$$k_{12,vb} = \int_{-\infty}^{E_v} D_p(E)(1 - f_p(E)) A_{ij}(E, x_t) f_{ij}(E_t - E)\, dE \tag{2.16c}$$

$$k_{21,vb} = \int_{-\infty}^{E_v} D_p(E) f_p(E) A_{ij}(E, x_t) f_{ij}(E - E_t)\, dE. \tag{2.16d}$$

The rates for electron emission use the inverse of the electron occupancy $1 - f_n$. For interactions with the valence-band occurrences of the electron occupancy are replaced with the hole occupancy $f_n = 1 - f_p$.

2.3.3 Band Edge Approximation

To simplify the equations above, the energy of the carriers in the bands can be approximated with the band edge energy. This allows to move the lineshape function and the electronic matrix-element out of the integration.

The inverse fermi-dirac distributions $1 - f_n$ and $1 - f_p$ are expressed as

$$1 - f_n(E) = f_p(E) = 1 - \frac{1}{1 + e^{\beta(E - E_f)}} = e^{\beta(E - E_f)} f_n \tag{2.17}$$

$$1 - f_p(E) = f_n(E) = 1 - \frac{1}{1 + e^{\beta(E_f - E)}} = e^{\beta(E_f - E)} f_p \tag{2.18}$$

with $\beta = \frac{1}{k_B T}$, the Boltzmann constant k_B and the temperature T. The rate equations can then be written as

$$k_{12,cb} = A_{ij}(E_c, x_t) f_{ij}(E_t - E_c) \int_{E_c}^{\infty} D_n(E) f_n(E)\, dE \tag{2.19a}$$

$$k_{21,cb} = A_{ij}(E_c, x_t) f_{ij}(E_c - E_t) \int_{E_c}^{\infty} D_n(E) f_n(E) e^{-\beta(E_f - E_c)}\, dE \tag{2.19b}$$

$$k_{12,vb} = A_{ij}(E_v, x_t) f_{ij}(E_t - E_v) \int_{-\infty}^{E_v} D_p(E) f_p(E) e^{-\beta(E_v - E_f)}\, dE \tag{2.19c}$$

$$k_{21,vb} = A_{ij}(E_v, x_t) f_{ij}(E_v - E_t) \int_{-\infty}^{E_v} D_p(E) f_p(E)\, dE. \tag{2.19d}$$

The integral terms can be identified as the equations for electron and hole densities:

$$n = \int_{E_c}^{\infty} D_n(E) f_n(E) \, dE \tag{2.20}$$

$$p = \int_{-\infty}^{E_v} D_p(E) f_p(E) \, dE, \tag{2.21}$$

to yield

$$k_{12,cb} = A_{ij}(E_c, x_t) f_{ij}(E_t - E_c) n \tag{2.22a}$$

$$k_{21,cb} = A_{ij}(E_c, x_t) f_{ij}(E_c - E_t) n e^{-\beta(E_f - E_c)} \tag{2.22b}$$

$$k_{12,vb} = A_{ij}(E_v, x_t) f_{ij}(E_t - E_v) p e^{-\beta(E_v - E_f)} \tag{2.22c}$$

$$k_{21,vb} = A_{ij}(E_v, x_t) f_{ij}(E_v - E_t) p \tag{2.22d}$$

Analytical expressions for the lineshape function and the matrix element will be given in section 2.3.5.

2.3.4 Field-assisted Transitions

In the previous section, transitions to a band of states were covered. This allows for example to calculate the transition rates for oxide traps in MOS transistors interacting with the adjacent gate and bulk layers. The trap is assumed to interact with the closest points in these reservoirs.

If the trap is located sufficiently close to its local band edges, the bands will also act as reservoirs for the trap. But unlike the reservoirs distant to the trap, no single point of interaction can be defined for them. Additionally, the reservoir for these transitions does not shift relative to the trap with the electric field as it does for interactions with distant reservoirs. No implicit field dependence of the rates exists if only local transitions to the bands are considered.

The idea behind field-assisted transitions is that a carrier moving from a trap to a state in the band can gain an energetic advantage by doing the transition non-local, i.e. by tunneling to a position where the band edge is at a more favourable level.

We will start by looking at the system at a more abstract level, expressing the rates as integrals over the probabilities of separated processes, as done in [30]. Using the earlier results from chapter 2.3.2, we will find detailed expressions for capture and emission rates. At the end we will simplify those expressions for implementation in a device simulator.

To evaluate the influence of the electric field on the capture and emission rates of a trap, a situation as depicted in figure 2.5 is used. A trap is located inside a semiconducting material with conduction and valence bands at energies E_c and E_v, and a homogeneous electric field F is assumed along the device.

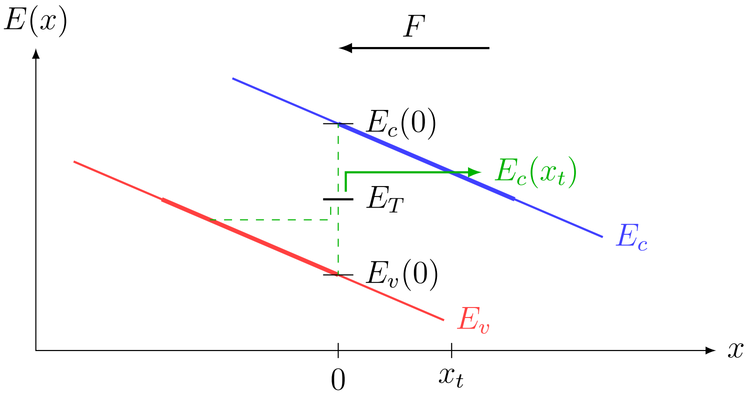


Figure 2.5: Possible paths of interaction between trap and bands. A trap is located at energy E_T inside a semiconducting material. Band edges $E_c(x)$ and $E_v(x)$ are sloped due to a constant electric field F.

Like earlier in this chapter, using the Huang-Born approximation to split the quantum mechanical system into an electronic and vibrational system, the emission or capture of an electron can be expressed as a combination of three seperate processes.

Taking for example the emission of an electron the processes would be:

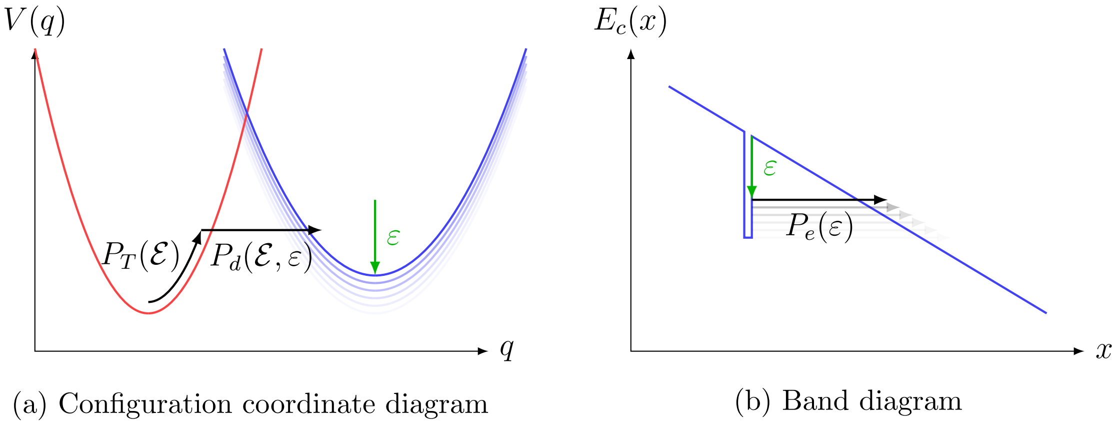
- The energy of the system in the bound state rises to a higher potential due to thermal excitation of the vibrational system.
- The vibrational system migrates from its charged configuration to an uncharged one.
- The electron system migrates from a state with the electron located at the trap position to a state with the electron located in the band.

Expressing each of these processes as a probability yields an expression for the transition rate. A visualization of the separate processes in the configuration coordinate diagram and the band diagram of the device can be seen in figure 2.6.

$$r \propto P_e(\varepsilon) P_d(\mathcal{E}, \varepsilon) P_T(\mathcal{E}) \tag{2.23}$$

Here, parameter ε is the change in potential of the band edge due to tunneling, and $\mathcal{E}$ is the energy of thermal excitation of the vibrational system.

- $P_T(\mathcal{E})$, the thermal excitation, is the probability of the bound state vibrational system being excited to a state $E > E_T$.
- $P_e(\varepsilon)$ is the probability of electron tunneling, i.e. the probability of the electronic system changing from state i to j. It depends on the distance the electron has to tunnel through the barrier and thus, for a given energy on the electric field.



(a) Configuration coordinate diagram (b) Band diagram

Figure 2.6: Seperated processes happening during emission of an electron to the conduction band. P_T is the probability of thermal excitation, P_d the probability of the vibrational wave function migrating from a state $i\alpha$ to $j\beta$ and P_e is the probability of the electronic wave function changing from state i to j. $\mathcal{E}$ is the thermal energy and ε is the change in band edge energy relative to the one at the trap position.

- $P_d(\mathcal{E},\varepsilon)$ is the probability of the the vibrational system changing from state $i\alpha$ to $j\beta$, it depends on both the thermal excitation of system and the lowering of the adiabatic potential due to tunneling.

Comparing the rate equation at this point with equation 2.13 we can see that $P_T P_d$ can be identified with the lineshape function f_{ij}, and the electron tunneling propability P_e with the matrix element A_{ij}.

The total emission rate with the influence of the field can then be evaluated by integrating the probabilities over all valid energy combinations. For the emission of an electron to the conduction band, for example, this means a thermal excitation $\mathcal{E} > 0$ and a field dependent lowering of the band edge by $\varepsilon < 0$.

$$r(F) = \iint_{\mathcal{E},\varepsilon} P_e(\varepsilon, F) P_d(\mathcal{E},\varepsilon) P_T(\mathcal{E})\, d\varepsilon d\mathcal{E} \tag{2.24}$$

Now we will split this rate in a field dependent and an independent part.

$$r(F) = r' + r''(F) \tag{2.25}$$

If we dismiss electron tunneling for the moment, i.e. we only look at the case of no electric field, we get the pure thermal rate to or from the local band edge

$$r' = P_e(0) \int_{\mathcal{E}} P_d(\mathcal{E}) P_T(\mathcal{E}) \, d\mathcal{E}. \tag{2.26}$$

This equation is basically an abstract version of the equations 2.16a in section 2.3.2, except for the fact that the electron does not have to tunnel to another position in the device. We can assume that the electronic transition time is much lower due to the large overlap of the wave functions i with all the electron wave functions j in the conduction band at or close to the trap position. Using equations 2.22 and exchanging the tunneling length dependent electronic matrix element $A_{ij}(E_c, x_t)$ with a local version A_0, the resulting equations read

$$k'_{12,cb} = A_0 f_{ij}(E_t - E_c) n \tag{2.27a}$$
$$k'_{21,cb} = A_0 f_{ij}(E_c - E_t) n e^{-\beta(E_f - E_c)} \tag{2.27b}$$
$$k'_{12,vb} = A_0 f_{ij}(E_t - E_v) p e^{-\beta(E_v - E_f)} \tag{2.27c}$$
$$k'_{21,vb} = A_0 f_{ij}(E_v - E_t) p. \tag{2.27d}$$

Going back to the general case $F \neq 0$, the assumed homogeneous electric field across the device results in a triangular potential barrier seen by the electron. It has a height of ε and a field dependent length of $x_t = |\varepsilon / qF|$. Tunneling through this barrier allows to reach states at an energy below the band edge energy at the trap location and gives an additional, field dependent contribution to the rate.

$$r''(F) = \iint_{\varepsilon \neq 0, \mathcal{E}} P_e(\varepsilon, F) P_d(\mathcal{E}, \varepsilon) P_T(\mathcal{E}) \, d\varepsilon d\mathcal{E} \tag{2.28}$$

For each infinitesimal lowering of the barrier, an amount of new states

$$D(E_{c,0} + \varepsilon) f_n(E_{c,0} + \varepsilon) d\varepsilon = -\frac{\partial n(x_t)}{\partial \varepsilon} d\varepsilon \tag{2.29}$$
$$D(E_{c,0} + \varepsilon) f_p(E_{c,0} + \varepsilon) d\varepsilon = \frac{\partial p(x_t)}{\partial \varepsilon} d\varepsilon \tag{2.30}$$

can be reached. The increase in transition rate however, is limited by the falling transparency of the barrier given by the electronic matrix element. Outside the potential well of the trap, the energy of the electron is lower than the barrier energy and its wave function diminishes exponentially. The matrix element in this case is again much smaller than for local emission and dependent on the strength of the field.

The product of states, electronic matrix element and lineshape function are then summed over the barrier lowering ε. Introducing expressions for the band energy of the lowered barrier $E_c' = E_{c,0} + \varepsilon$, and expressing n and p as a function of ε, the additional field dependent rates are

$$k''_{12,cb} = -\int_\varepsilon \frac{dn(\varepsilon)}{d\varepsilon} f_{ij}(E_t - E_c') A_{ij}(E_c', F)\, d\varepsilon \tag{2.31a}$$

$$k''_{21,cb} = -\int_\varepsilon \frac{dn(\varepsilon)}{d\varepsilon} f_{ij}(E_c' - E_t) A_{ij}(E_c', F) e^{-\beta(E_f - E_c')}\, d\varepsilon \tag{2.31b}$$

$$k''_{12,vb} = \int_\varepsilon \frac{dp(\varepsilon)}{d\varepsilon} f_{ij}(E_t - E_v') A_{ij}(E_v', F) e^{-\beta(E_v' - E_f)}\, d\varepsilon \tag{2.31c}$$

$$k''_{21,vb} = \int_\varepsilon \frac{dp(\varepsilon)}{d\varepsilon} f_{ij}(E_v' - E_t) A_{ij}(E_v', F)\, d\varepsilon. \tag{2.31d}$$

Together with equations 2.27, they yield the rate equations for the interaction with the local bands.

$$\begin{aligned} k_{12,cb} = A_0 f_{ij}(E_t - E_c) n & \\ - \int_\varepsilon \frac{dn(\varepsilon)}{d\varepsilon} f_{ij}(E_t - E_c') A_{ij}(E_c', F)\, d\varepsilon & \end{aligned} \tag{2.32a}$$

$$\begin{aligned} k_{21,cb} = A_0 f_{ij}(E_c - E_t) n e^{-\beta(E_f - E_c)} & \\ - \int_\varepsilon \frac{dn(\varepsilon)}{d\varepsilon} f_{ij}(E_c' - E_t) A_{ij}(E_c', F) e^{-\beta(E_f - E_c')}\, d\varepsilon & \end{aligned} \tag{2.32b}$$

$$\begin{aligned} k_{12,vb} = A_0 f_{ij}(E_t - E_v) p e^{-\beta(E_v - E_f)} & \\ + \int_\varepsilon \frac{dp(\varepsilon)}{d\varepsilon} f_{ij}(E_t - E_v') A_{ij}(E_v', F) e^{-\beta(E_v' - E_f)}\, d\varepsilon & \end{aligned} \tag{2.32c}$$

$$\begin{aligned} k_{21,vb} = A_0 f_{ij}(E_v - E_t) p & \\ + \int_\varepsilon \frac{dp(\varepsilon)}{d\varepsilon} f_{ij}(E_v' - E_t) A_{ij}(E_v', F)\, d\varepsilon. & \end{aligned} \tag{2.32d}$$

In the following chapter these equations will be simplified for implementation in a device simulator.

2.3.5 Practical Considerations For Implementation

To further simplify the equations allowing them to be implementated in device simulators, simplified analytical expressions for the electronic matrix element A and the lineshape function f are used. The derivatives of the carrier concentration will be approximated to allow the usage of local variables.

The Lineshape Function

The lineshape function f_{ij} has its main contribution at the intersection point and peaks there in the classical limit [27]. We will assume transitions in the vibrational system to happen only at the intersection point of the two parabolas in the configuration coordinate diagram. This reduces the propability of defect tunneling P_d to a dirac peak at the energy of intersection E_{IP}. The number of states available for reaction will therefore be taken as the states available at the intersection point $P_T(\varepsilon_{ij})$.

$$f_{ij}(E) \approx P_T(\varepsilon_{ij}(E)) \tag{2.33}$$

The probability that the vibrational system is excited by an energy $\mathcal{E}$ is given by Boltzmann statistics and is proportional to a factor $e^{-\beta\mathcal{E}}$.

$$P_T(\varepsilon_{ij}) \propto e^{-\beta\varepsilon_{ij}} \tag{2.34}$$

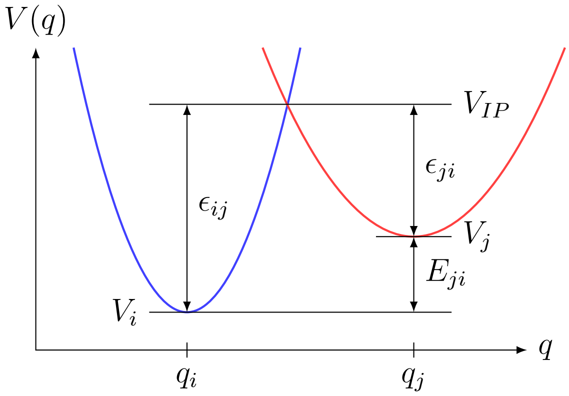


Figure 2.7: Adiabatic potential energy surfaces in the configuration coordinate diagram for two example states i and j. V_i and V_j are the base energies of the parabolas, V_{IP} their intersection point, ε_{ij} the NMP barrier from state i to j and ε_{ji} the NMP barrier in the reverse direction.

The height of the thermal barrier ε_{ij} is given by the intersection of the two parabolas in the configuration coordinate diagram as shown in figure 2.7. It can be calculated from the difference in ground state energies $E_{ji} = V_j - V_i$, the curvatures of the parabolas c_i, c_j and the distance $q_j - q_i$ of the parabolas in the configuration coordinate diagram [2].

$$\varepsilon_{ij} = \frac{c_i(q_j - q_i)^2}{\left(\frac{c_i}{c_j} - 1\right)^2} \left(1 \pm \sqrt{\frac{c_i}{c_j} + \frac{(E_{ji})(\frac{c_i}{c_j} - 1)}{c_j(q_j - q_i)^2}}\right)^2 \tag{2.35}$$

While two different parabolas generally have either two points of intersection or none at all, the smaller solution will be used as transition there is more likely. A different way of parametrizing the parabolas is using the Huang-Rhys factor S_{ij} and the ratio of curvatures R_{ij}.

$$S_{ij} = \frac{c_i(q_j - q_i)^2}{\hbar\omega} \tag{2.36}$$

$$R_{ij} = \sqrt{\frac{c_i}{c_j}} \tag{2.37}$$

Using these definitions, equation 2.35 can be rewritten as

$$\varepsilon_{ij} = \frac{S_{ij}\hbar\omega}{(R_{ij}^2 - 1)^2} \left(1 - R_{ij}\sqrt{\frac{S_{ij}\hbar\omega + E_{ji}(R_{ij}^2 - 1)}{S_{ij}\hbar\omega}}\right)^2 . \tag{2.38}$$

For the special case of $R_{ij} = 1$, there is, in general, only one intersection point and the formula above fails. The thermal barrier for this case can be calculated as

$$\varepsilon_{ij} = \frac{(S_{ij}\hbar\omega + Eji)^2}{4S_{ij}\hbar\omega}. \tag{2.39}$$

The Electronic Matrix Element

While it is possible to find analytical solutions for the wave functions in triangular potential wells (Airy functions) [31], insufficient knowledge of the exact shape of the potential well around the trap refrains us from calculating the matrix element A_{ij}. Thus it is usually approximated by an effective capture cross section σ_0 multiplied by the thermal velocity v_{th} of the carriers and a tunneling factor λ after Wentzel–Kramers–Brillouin (WKB).

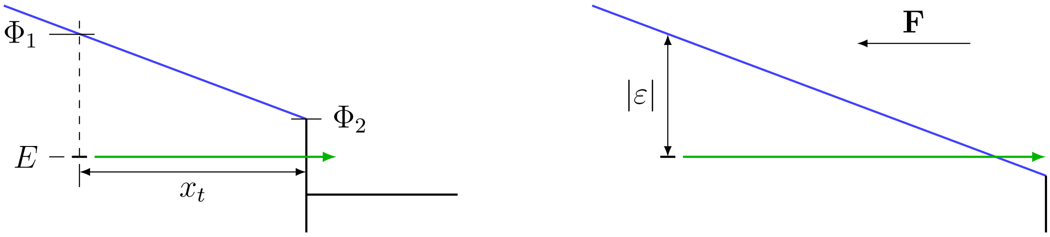
$$A_{ij}(E, x_t) \approx A_0 \lambda(E, x_t) = \sigma_0 v_{th} \lambda(E, x_t) \tag{2.40}$$

For a trapezoidal barrier as shown figure in 2.8a, the tunneling factor $\lambda(E, x)$ is given as [14, 19]

$$\lambda(E, x) = \exp\left(-\frac{4\sqrt{2m^*} x_t}{3\hbar q_0 (\phi_2 - \phi_1)} ((\phi_2 - E)^{3/2} - (\phi_1 - E)^{3/2}) \right). \tag{2.41}$$

Here, ϕ_1 and ϕ_2 are the potentials of the barrier and E the energy of the electron. The triangular barrier, shown in 2.8b, is a special case of the trapezoid barrier. Using the barrier height ε and the tunneling distance $x_t = |\varepsilon / qF|$, this equation simplifies to

$$\lambda(\varepsilon, F) = \exp\left(-\frac{4\sqrt{2m^*}}{3\hbar q_0 F} |\varepsilon|^{\frac{3}{2}} \right). \tag{2.42}$$



(a) Trapezoidal barrier (b) Triangular barrier

Figure 2.8: Band diagrams showing triangular and trapezoidal tunneling barriers and the parameters used to describe them.

It equals the asymptotic expansion of the Airy function [31].

The Additional States Reachable By Tunneling

To find a suitable expression for the number of states $\frac{\partial n}{\partial \varepsilon} d\varepsilon$, we start with the electron density $n(x_t)$ at position x_t.

$$n(x_t) = \int_{E_C(x_t)}^{\infty} D_n(E - E_C(x_t)) f_n(E - E_f)\, dE \tag{2.43}$$

Or, equivalent, expressed as a function of ε.

$$n(\varepsilon) = \int_{E_{C,0}+\varepsilon}^{\infty} D_n(E - (E_{C,0} + \varepsilon)) f_n(E - E_f)\, dE \tag{2.44}$$

Substituting $E = E' + \varepsilon$ and assuming a Fermi-Dirac distribution of the carriers yields

$$\begin{aligned} n(\varepsilon) &= \int_{E_{C,0}}^{\infty} D_n(E' - E_{C,0}) f_n(E' + \varepsilon - E_f)\, dE' &(2.45)\\ &= \int_{E_{C,0}}^{\infty} D_n(E' - E_{C,0}) \frac{1}{1 + e^{\beta(E'+\varepsilon-E_f)}}\, dE'. &(2.46)\end{aligned}$$

We can now derive $n(\varepsilon)$ and find

$$\begin{aligned} \frac{dn(\varepsilon)}{d\varepsilon} &= \int_{E_{C,0}}^{\infty} D_n(E' - E_{C,0}) \frac{1}{1 + e^{\beta(E'+\varepsilon-E_f)}} \frac{e^{\beta(E'+\varepsilon-E_f)}}{1 + e^{\beta(E'+\varepsilon-E_f)}} (-\beta)\, dE' &(2.47)\\ &= \int_{E_{C,0}}^{\infty} D_n(E' - E_{C,0}) f_n(\cdot)(1 - f_n(\cdot))(-\beta)\, dE' &(2.48)\\ &= \int_{E_{C,0}}^{\infty} D_n(E' - E_{C,0})(f_n(\cdot) - (f_n(\cdot))^2)(-\beta)\, dE'. &(2.49)\end{aligned}$$

If we assume the lowest point of integration to be still well above Fermi level i.e. $f_n << 1$, f_n^2 can be neglected and the result can be approximated as

$$\frac{dn(\varepsilon)}{d\varepsilon} \approx -\beta n(\varepsilon). \tag{2.50}$$

Doing the same calculation for p yields

$$\frac{dp(\varepsilon)}{d\varepsilon} \approx \beta p(\varepsilon). \tag{2.51}$$

Not surprisingly, the same factor ($\mp\beta$) is obtained directly if Boltzmann distribution of the carriers is assumed.

$$\begin{aligned} n(\varepsilon) &= \int_{E_{C,0}}^{\infty} D_n(E' - E_{C,0}) e^{-\beta(E'+\varepsilon-E_f)}\, dE' && (2.52) \\ \frac{dn(\varepsilon)}{d\varepsilon} &= -\beta \int_{E_{C,0}}^{\infty} D_n(E' - E_{C,0}) e^{-\beta(E'+\varepsilon-E_f)}\, dE' && (2.53) \end{aligned}$$

Using the simplifications for the lineshape function (equations 2.33, 2.34) and the matrix element (equation 2.40), equations 2.22 for interaction with a band of states can be written as

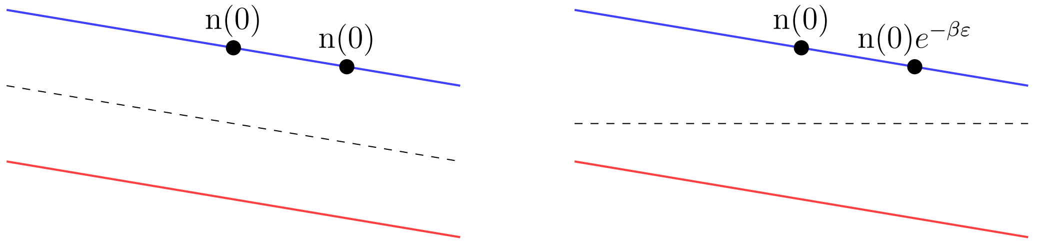
$$\begin{aligned} k_{12,cb} &= \sigma_0 v_{th} n e^{-\beta(\varepsilon_{ij}(E_t-E_c))} \lambda(E, x_t) && (2.54a) \\ k_{21,cb} &= \sigma_0 v_{th} n e^{-\beta(\varepsilon_{ij}(E_c-E_t))} e^{-\beta(E_f-E_c)} \lambda(E, x_t) && (2.54b) \\ k_{12,vb} &= \sigma_0 v_{th} p e^{-\beta(\varepsilon_{ij}(E_t-E_v))} e^{-\beta(E_v-E_f)} \lambda(E, x_t) && (2.54c) \\ k_{21,vb} &= \sigma_0 v_{th} p e^{-\beta(\varepsilon_{ij}(E_v-E_t))} \lambda(E, x_t). && (2.54d) \end{aligned}$$

With the same simplifications and the simplification for the additional states (equations 2.50 and 2.51), equations 2.32 for the interaction with local bands can be simplified to

$$\begin{aligned} k_{12,cb}(F) = \sigma_0 v_{th} \Bigg(& n(0) e^{-\beta(\varepsilon_{ij}(E_t-E_c))} \\ & + \beta \int_{\varepsilon} n(\varepsilon) \lambda(\varepsilon, F) e^{-\beta(\varepsilon_{ij}(E_t-E_c'))}\, d\varepsilon \Bigg) && (2.55a) \\ k_{21,cb}(F) = \sigma_0 v_{th} \Bigg(& n(0) e^{-\beta(\varepsilon_{ij}(E_c-E_t))} e^{-\beta(E_f-E_c)} \\ & + \beta \int_{\varepsilon} n(\varepsilon) \lambda(\varepsilon, F) e^{-\beta(\varepsilon_{ij}(E_c'-E_t))} e^{-\beta(E_f-E_c')}\, d\varepsilon \Bigg) && (2.55b) \\ k_{12,vb}(F) = \sigma_0 v_{th} \Bigg(& p(0) e^{-\beta(\varepsilon_{ij}(E_t-E_v))} e^{-\beta(E_v-E_f)} \\ & + \beta \int_{\varepsilon} p(\varepsilon) \lambda(\varepsilon, F) e^{-\beta(\varepsilon_{ij}(E_t-E_v'))} e^{-\beta(E_v'-E_f)}\, d\varepsilon \Bigg) && (2.55c) \\ k_{21,vb}(F) = \sigma_0 v_{th} \Bigg(& p(0) e^{-\beta(\varepsilon_{ij}(E_v-E_t))} \\ & + \beta \int_{\varepsilon} p(\varepsilon) \lambda(\varepsilon, F) e^{-\beta(\varepsilon_{ij}(E_v'-E_t))}\, d\varepsilon \Bigg). && (2.55d) \end{aligned}$$

Carrier Concentrations

Further simplifications which are useful for implementation and discussion of the model can be made for the carrier concentrations $n(\varepsilon)$ and $p(\varepsilon)$ for two special cases in the behaviour of the band edges and the Fermi level shown in figure 2.9.



(a) Fermi level with a constant offset to the band edges (b) Constant Fermi level, sloped conduction and valence band edges

Figure 2.9: Band diagrams showing the dependence of carrier concentration n on the slope of the fermi level for two special cases.

For the case of the Fermi level having a constant offset to the band, $n(\varepsilon)$ and $p(\varepsilon)$ are constant and can be replaced with $n(0)$ and $p(0)$. For the case of the Fermi level being constant, $n(\varepsilon)$ and $p(\varepsilon)$ in the integral can be replaced with:

$$n(\varepsilon) = n(0)e^{-\beta\varepsilon} \tag{2.56}$$
$$p(\varepsilon) = p(0)e^{\beta\varepsilon} \tag{2.57}$$

Polarization charges resulting from some material combinations can lead to such a situation at zero bias. Assuming the difference in slope between fermi level and band stays constant, the carrier densities can then be calculated as:

$$n(\varepsilon) = n(0)e^{-\beta\varepsilon F_\pi/F} \tag{2.58}$$
$$p(\varepsilon) = p(0)e^{\beta\varepsilon F_\pi/F} \tag{2.59}$$

With the polarization induced field F_π.

2.4 Thermal Transitions

Transitions without a transfer of charge, i.e. between meta-stable and stable states 1-1′ and 2-2′ are called thermal transitions. The two configurations of the trap are separated by an energy barrier in the configuration coordinate diagram. The exact shape of the barrier is not relevant, a visualization is given in figure 2.10. Transitions happen once an activation energy ε_{ij} is overcome, so the rate is proportional to a Boltzmann factor $e^{-\beta\varepsilon_{ij}}$. With the definition of an attempt frequency ν_0 the rate equation for thermal transitions is:

$$k_{ij} = \nu_0 e^{-\beta\varepsilon_{ij}} \tag{2.60}$$

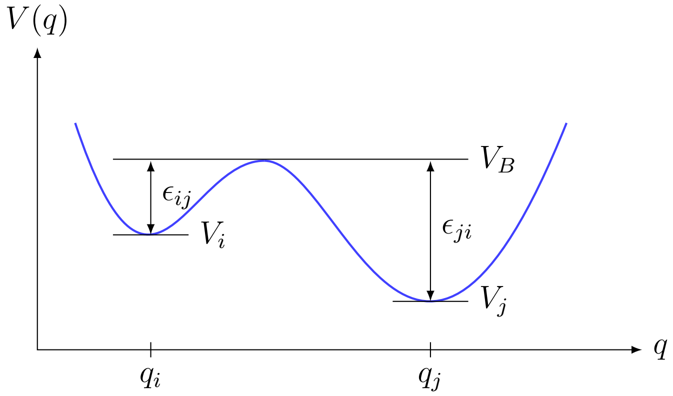


Figure 2.10: Configuration coordinate diagram for a thermal transition between stable and meta-stable states j and i. V_i and V_j are the energy minima of the states, V_B the energy of the thermal barrier, ε_{ij} the thermal barrier from state i to j and ε_{ji} the thermal barrier in the reverse direction. Graphic adopted from [26].

CHAPTER 3

Results

In the first part of this chapter the four-state model and especially the field dependent rates to the local bands will be discussed (section 3.1). A simple test structure will be used to show dependence of the rates on electric field (section 3.1.1) and temperature (section 3.1.2). Different domains of field dependence will be identified and the dependence of the apparent thermal energy barrier on the field will be evaluated.

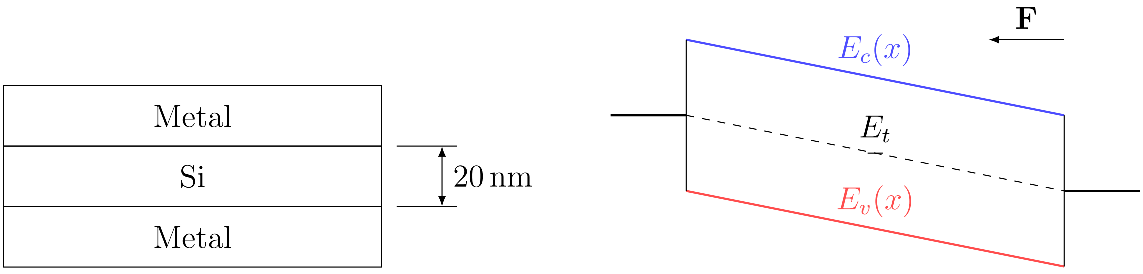
In the latter part of the chapter, realistic MOS or HEMT structures will be used for simulation. The influence of the local band interaction on capture and emission rates of the traps in a MOS structure in dependence of the band gap energy, trap level and device length will be discussed in section 3.2. Trap assisted tunneling currents in a GaN/AlGaN device calculated with the NMP model will be compared to the currents resulting from the popular Fowler-Nordheim and Frenkel-Poole models in section 3.3.1. Finally in section 3.3.2, transient charging and static tunnel currents resulting from the traps in a MOSFET in a MSM measurement setup will be discussed.

3.1 Capture and Emission Rates

To take a closer look on the resulting rates to the local bands, a simple test structure containing only a single trap will be used in this section. The test structure is shown in figure 3.1. For simplicity silicon was chosen as the semiconducting material under test and the metastable states in the four-state model are not used, making it an effective two-state model. If not specified otherwise, in the following the rates are to or from the conduction band, and the temperature is 300 K. The trap parameters can be found in table 3.1 and a plot of the resulting configuration coordinate diagram in figure 3.2.

As stated in section 2.3.4, the emission rates k_{nm} are composed of a pure thermal part $k'_{nm}(T)$ and a field dependent part $k''_{nm}(T,F)$.

$$k_{nm}(T,F) = k'_{nm}(T) + k''_{nm}(T,F) \tag{3.1}$$



(a) Geometry (b) Band diagram

Figure 3.1: Test structure used in section 3.1. A single trap is centered in a Si layer at 0 eV (midgap).

Table 3.1: Simulation parameters for the discussion of capture and emission rates in section 3.1.

Parameter	**Value**	**Description**
$E1$	0.0 eV	Trap level
$R12$	0.6	Ratio of curvatures for transitions 1-2
$S12$	2.3 eV	Huang-Rhys parameter S times $\hbar\omega$
σ_0	$1 \times 10^{-23}\,\mathrm{cm}^2$	Capture cross section

In figure 3.3a the total capture rate k_{12} from the conduction band, the pure thermal part k'_{12} and field dependent part k''_{12} of the capture rate of our test trap are shown. Figure 3.3b shows the same quantities for the emission rate. Additionally four general regions (**A,B,C,D**) have been defined, used in the following discussion.

At low fields the rates for capture and emission are very close. This is because the trap is positioned at the Fermi level and the system is in thermal equilibrium. At higher electric fields the rates for emission gain much more than the capture rates.

This is due to the behaviour of the NMP barrier and the resulting probabilities P_T of thermal excitation. It can be seen from figure 3.2 that the lowering of the band energy has a high influence on the NMP barrier from the trap to the band. The reverse NMP barrier changes less and increases with ε in for this set of trap parameters.

The regions shown in figure 3.3 are named:

- A - Low electric fields
- B - Voltage limited region
- C - Field limited region
- D - Saturation

While the widths of regions A and D are arbitrary, the transition between regions B and C is definite. The regions will be explained in the following.

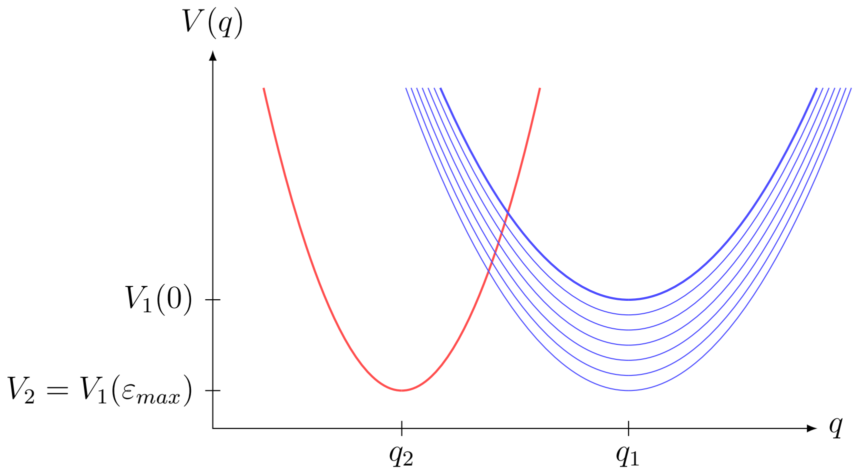
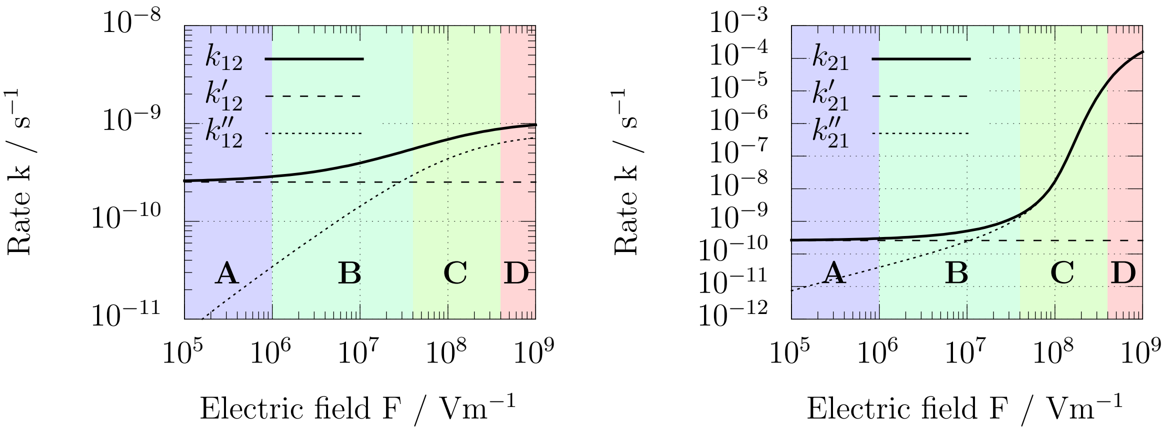


Figure 3.2: Adiabatic potential energy surfaces in the configuration coordinate diagram for the trap parameters $R_{12} = 0.6$ and $S_{12}\hbar\omega = 2.3\,\text{eV}$ over the range of ε. The topmost, thick parabola belongs to the uncharged (conduction band) state for $\varepsilon = 0.0\,\text{eV}$, the bottommost for $\varepsilon_{max} \approx 0.6\,\text{eV}$.



(a) Capture rates k_{12}

(b) Emission rates k_{21}

Figure 3.3: Plot of the total rates $k(F)$ over the electric field, its thermal and field dependent parts k' and $k''(F)$ respectively, for capture and emission.

A - Low electric fields: In this region, the rates are dominated by the local, pure thermal transitions. The field-dependent contribution is negligible.

$$k_{nm,A} \approx k'_{nm} \tag{3.2}$$

B - Voltage limited region: As illustrated in figure 3.4a, for low electric fields and thin layers, the band edge energy never reaches the trap energy. This limits the possible barrier lowering and in turn the states available for interaction. The field limit for this region is

$$F_B < \left| \frac{E_{c/v}(0) - E_t}{x_{max}} \right| \tag{3.3}$$

with the distance of the trap to the edge of the semiconducting layer x_{max}, the band edge energy at trap position $E_{c/v}(0)$ and trap energy E_t. In our test structure this happens until about 1.2 V across the 30 nm Si layer, or at a field strength of about 0.4 MV/cm.

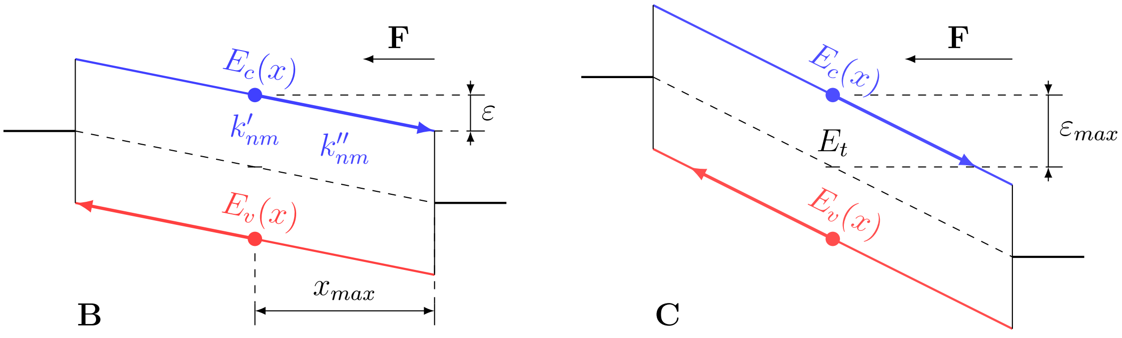
In this region two mechanisms are responsible for the increase of the rates. The tunneling distance to the already reachable states decreases and additional states become available for tunneling.

Approximate analytical expressions for the behaviour of the rates in regions B and C are given in section 3.1.1.

C - Field limited region: Illustrated in 3.4b. All thermal states of the trap contribute to the rates and the rates increase with the decreasing tunneling distance.

D - Saturation: At very high electric fields, the tunneling distances become very short and the rates saturate. They are limited by the thermal activation of the trap and the states available in the band.

$$k_{nm,D} \approx k'_{nm} + k''_{nm,sat} \tag{3.4}$$



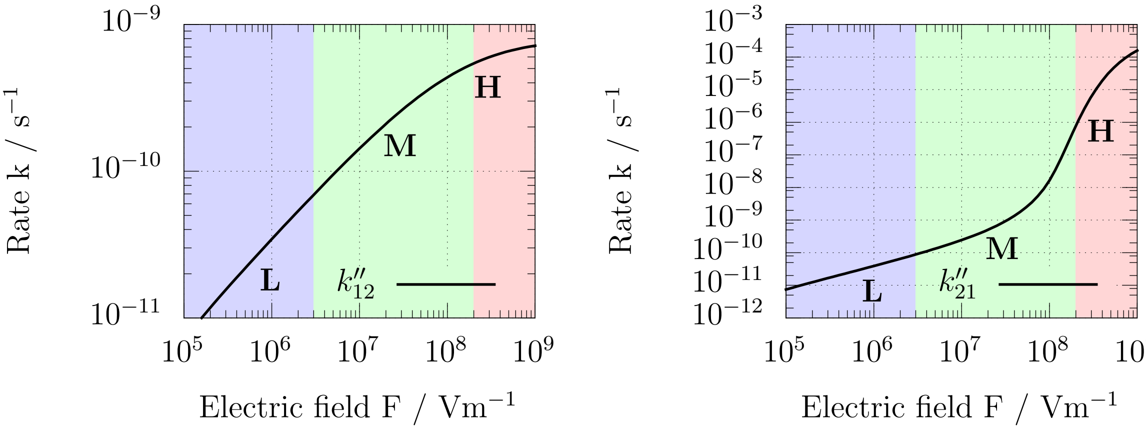
(a) Voltage limited region **B**. With higher electric fields more states in the conduction band become available for tunneling at lower thermal excitations of the trap.

(b) Field limited region **C**. All thermally excited states of the trap system contribute to the rates, higher fields reduce the tunneling distance.

Figure 3.4: Band diagrams of the test structure in voltage (**B**) and field (**C**) limited regions. In region **B**, the lowest energy of the conduction band edge limits the barrier lowering.

3.1.1 Field Dependence

In this section we will focus on the behaviour of the field dependent rates $k''(F)$. We will try to simplify the rate equations to yield analytical expressions for their field dependences. The field dependent rates k'' visible in figure 3.3 are plotted again in figure 3.5.



(a) Capture rate k''_{12}

(b) Emission rate k''_{21}

Figure 3.5: Plots of the capture and emission rates k'' over F with regions of different field dependence.

Looking at this plots we can define three regions of the field dependence, termed (**L,M,H**) for low, intermediate, and high fields. The rates show a growth obeying a power law in the L and exponential growth in the M region. In region H they saturate. For the capture rate k_{12} there is no exponential growth for this set of trap parameters.

As we have to assume the field dependence to differ between the voltage limited and field limited cases defined before, they will be treated separately.

Voltage Limited Region

To assess the field dependence in the voltage limited region, we first rewrite the integral in rate k''_{nm} from an integral over the energy to an integral over the tunneling distance to keep the integration borders constant.

$$k''_{nm} = -\sigma_0 v_{th} \int_0^{\varepsilon_{max}(F)} \frac{dn(\varepsilon)}{d\varepsilon} \lambda(\varepsilon, F) P_T(\varepsilon)\, d\varepsilon \tag{3.5}$$

$$= -\sigma_0 v_{th} \int_0^{x_{max}} \frac{dn(xF)}{d\varepsilon} \lambda(x, F) P_T(x) F\, dx \tag{3.6}$$

To simplify the expression for the probability of thermal excitation $P_T(\varepsilon)$, the dependence of the NMP barrier on ε is approximated by a constant B.

$$\frac{d\varepsilon_{nm}}{d\varepsilon} \approx B \tag{3.7}$$

Depending on the configuration of the parabolas, this can be a rather crude approximation but it allows to describe the different domains of the capture and emission rates qualitatively. The value of B depends on the electron-phonon coupling regime of the transitions. With equations 2.42 and 2.34, equation 3.6 reads:

$$k''_{nm,B} \approx -\sigma_0 v_{th} F \int_0^{x_{max}} \frac{dn(xF)}{d\varepsilon} e^{\left(-\frac{4\sqrt{2m^*}}{3\hbar q_0} x^{\frac{3}{2}}\right)\sqrt{F}} e^{-\beta(\varepsilon_{nm}(E_t-E_c)-BFx)}\, dx \tag{3.8}$$

The number of states $\frac{dn}{d\varepsilon}d\varepsilon = -\beta n(0) d\varepsilon$ are constant in our test structure, because the Fermi level has a constant offset to the band edges. A lower slope of the Fermi level would lead to a higher value for B.

$$k''_{nm,B} \approx \beta n \sigma_0 v_{th} F \int_0^{x_{max}} e^{\left(-\frac{4\sqrt{2m^*}}{3\hbar q_0} x^{\frac{3}{2}}\right)\sqrt{F}} e^{-\beta(\varepsilon_{nm}(E_t-E_c)-BFx)}\, dx \tag{3.9}$$

The constants and the electric field in the WKB factor are subsumed in an effective tunneling length $x_0(F)$ as a measure for the width of this factor in x.

$$x_0^{\frac{3}{2}} = \frac{1}{\frac{4\sqrt{2m^*}}{3\hbar q_0}\sqrt{F}} \tag{3.10}$$

$$x_0 = \left(\frac{3\hbar q_0}{4\sqrt{2m^*}}\right)^{2/3} F^{-1/3} \tag{3.11}$$

Substituting x_0 and factoring out the constant NMP barrier yields

$$k''_{nm,B} \approx \beta n \sigma_0 v_{th} e^{-\beta\varepsilon_{nm}(E_t-E_c)} F \int_0^{x_{max}} e^{-\left(\frac{x}{x_0(F)}\right)^{\frac{3}{2}}+\beta BFx}\, dx. \tag{3.12}$$

Or, with equation 2.27

$$k''_{nm,B} \approx k'_{nm}\, \beta F \int_0^{x_{max}} e^{-\left(\frac{x}{x_0(F)}\right)^{\frac{3}{2}}+\beta BFx}\, dx. \tag{3.13}$$

This equation is now used to discuss the field dependence for Low (L), Medium (M), and High (H) fields in the voltage limited region **B**.

L - Low fields: For low fields, the WKB factor dominates the behaviour of the integrand. The barrier lowering is negligible and the rates can be further approximated as

$$k''_{nm,B-L} \approx k'_{nm}\,\beta F \int_0^{x_{max}} e^{-\left(\frac{x}{x_0(F)}\right)^{\frac{3}{2}}}\,dx. \tag{3.14}$$

At very low fields, x_0 is much larger than x_{max} and the integral evaluates to x_{max}. The rates are proportional to F.

$$k''_{nm,B-L} \approx k'_{nm}\,\beta F x_{max} \tag{3.15}$$

At higher fields, x_0 becomes shorter than x_{max}. The integrand has its highest contribution at short distances and its effective width decreases with $F^{-1/3}$ due to the increasing height of the tunneling barrier with the field. Due to the fast decay of the integrand for $x > x_0$, the integral can be estimated with

$$\int_0^{\infty} e^{-\left(\frac{x}{x_0}\right)^{\frac{3}{2}}}\,dx = \Gamma(\frac{5}{3})x_0 \approx 0.9\,x_0. \tag{3.16}$$

The states available still increase proportional to F and thus the rates in the low field region are proportional to $F^{2/3}$.

$$k''_{nm,B-L} \approx k'_{nm}\,\beta\,0.9\left(\frac{3\hbar q_0}{4\sqrt{2m^*}}\right)^{2/3} F^{2/3} \tag{3.17}$$

M - Medium electric fields: As the triangular barrier gets steeper, the positive exponential term in equation 3.13, representing the lowering of the NMP barrier, starts dominating for short distances. The integrand will reach its maximum between 0 and x_{max}. To approximate the rates, we will calculate the location of the maxima, assume a gaussian shape of the integrand and do a series expansion around the maxima to get the amplitude and variance.

The exponential function is monotonous, so the location of its maximum can be found by finding the maximum of the exponent.

$$\frac{d}{dx_i}\left(-\left(\frac{x_i}{x_0}\right)^{\frac{3}{2}} + \beta BFx_i\right) \overset{!}{=} 0 \tag{3.18}$$

$$x_i = \left(\frac{2}{3}\beta BF\right)^2 x_0^3 \tag{3.19}$$

Doing the series expansion yields

$$k''_{nm,B-M} \approx k'_{nm}\,\beta F e^{-\left(\frac{x_i}{x_0}\right)^{\frac{3}{2}}+\beta BFx_i} \int_0^{x_{max}} e^{-\frac{(x-x_i)^2}{16/9\beta BFx_0^3}}\,dx. \tag{3.20}$$

Using the relation for the area under a Gauss curve

$$\int_{-\infty}^{\infty} e^{-\frac{(x+b)^2}{a}} = \sqrt{(a\pi)}, \tag{3.21}$$

the equation for simplifies to

$$k''_{nm,B-M} \approx k'_{nm}\,\beta F e^{-\frac{4}{27}(\beta BFx_0)^3}\frac{4}{3}\sqrt{\pi\beta BFx_0^3} \tag{3.22}$$

$$\approx k'_{nm}\,\beta\frac{4}{3}\sqrt{\pi\beta B}\left(\frac{3\hbar q_0}{4\sqrt{2m^*}}\right)Fe^{\frac{4}{27}(\beta B)^3\left(\frac{3\hbar q_0}{4\sqrt{2m^*}}\right)^2F^2}. \tag{3.23}$$

Subsuming all linear constants A and exponential constants in C shows the field dependence more clearly:

$$k''_{nm,B-M} \approx k'_{nm}\,AFe^{CF^2} \tag{3.24}$$

In this domain the effective energy barrier gets lower and the rates increase exponentially with F^2. If the dependence of the NMP barrier on ε is very low this domain will not be reached. For negative dependences of the barrier energy on ε, as in capture rate k''_{12} of the test structure used in this chapter, the exponential dependence leads to saturation.

For the case of negative B, there is no positive exponential term and the maximum is always at $x = 0$. To approximate this case, we drop the WKB factor from the integral and use its effective tunneling width x_0 as the upper integral bound.

$$k''_{nm,B-M} \approx k'_{nm}\,\beta F \int_0^{x_0} e^{\beta BFx}\,dx \tag{3.25}$$

This can now be integrated to

$$k''_{nm,B-M} \approx k'_{nm}\,\frac{1}{B}(e^{\beta BFx_0} - 1) \tag{3.26}$$

$$\approx k'_{nm}\,\frac{1}{B}(e^{\beta B\left(\frac{3\hbar q_0}{4\sqrt{2m^*}}\right)^{2/3}F^{2/3}} - 1). \tag{3.27}$$

H - High fields: To reach the high field region while simultaneously being voltage limited, the trap has to be positioned very close to the edge of the semiconducting layer. The rates show no saturation even at high fields due to the new states reachable. The WKB factor can be neglected and the rates results in

$$k''_{nm,B-H} \approx k'_{nm}\,\beta F \int_0^{x_{max}} e^{\beta BFx}\, dx \tag{3.28}$$

$$\approx k'_{nm}\,\frac{1}{B}(e^{\beta BFx_{max}} - 1). \tag{3.29}$$

Field Limited Region

Going back to an integration over energy, again to keep constant integral borders, and using the same approximations, the rates read

$$k''_{nm,C} \approx k'_{nm}\,\beta \int_0^{\varepsilon_{max}} e^{-\frac{4\sqrt{2m^*}}{3\hbar q_0}\frac{1}{F}\varepsilon^{\frac{3}{2}}+\beta B\varepsilon} d\varepsilon. \tag{3.30}$$

Similar to the effective tunneling length, we can define an effective tunneling barrier ε_0. It is a measure for the width in ε of the exponentially decaying WKB part of the integrand.

$$\varepsilon_0 = \left(\frac{3\hbar q_0}{4\sqrt{2m^*}}\right)^{\frac{2}{3}} F^{\frac{2}{3}} \tag{3.31}$$

With this definition, the rate can be written as

$$k''_{nm,C} \approx k'_{nm}\,\beta \int_0^{\varepsilon_{max}} e^{-\left(\frac{\varepsilon}{\varepsilon_0}\right)^{\frac{3}{2}}+\beta B\varepsilon} d\varepsilon. \tag{3.32}$$

L - Low fields: The integral has its highest contribution at high energies due to the WKB factor. This is analogous to the low field case in the voltage limited region.

Due to the low field, ε_0 is large and the dependence of the WKB factor on ε is much higher than that of the thermal probability. The NMP barrier stays effectively constant in the relevant energy range.

$$k''_{nm,C-L} \approx k'_{nm}\,\beta \int_0^{\varepsilon_{max}} e^{-\left(\frac{\varepsilon}{\varepsilon_0}\right)^{\frac{3}{2}}} d\varepsilon \tag{3.33}$$

The effective width of the integral increases with ε_0 and the approximated rate results in

$$\begin{aligned} k''_{nm,C-L} &\approx k'_{nm}\,\beta\,0.9\varepsilon_0 && (3.34)\\ &\approx k'_{nm}\,\beta\,0.9\left(\frac{3\hbar q_0}{4\sqrt{2m^*}}\right)^{2/3} F^{2/3} && (3.35)\end{aligned}$$

for low fields. This is the same result we obtained in equation 3.17, in the voltage limited region. The increase in the maximum tunneling height has no effect on the approximated rates because the integrand has its main contribution at low tunneling distances and heights.

M - Medium electric fields: At higher electric fields and $B > 0$, the positive exponential term in equation 3.32 starts dominating for low ε. The maxima of the integrand lies between 0 and ε_{max}. The rates can be approximates with a series expansion around this maximum as done in the voltage limited region.

$$\begin{aligned} \frac{d}{d\varepsilon_i}\left(-\left(\frac{\varepsilon_i}{\varepsilon_0}\right)^{\frac{3}{2}} + \beta B\varepsilon_i\right) &\overset{!}{=} 0 && (3.36)\\ \varepsilon_i &= \left(\frac{2}{3}\beta B\right)^2 \varepsilon_0^3 && (3.37)\end{aligned}$$

Doing the series expansion yields

$$\begin{aligned} k''_{nm,C-M} &\approx k'_{nm}\,\beta e^{-\left(\frac{\varepsilon_i}{\varepsilon_0}\right)^{\frac{3}{2}}+\beta B\varepsilon_i} \int_0^{\varepsilon_{max}} e^{-\frac{(\varepsilon-\varepsilon_i)^2}{16/9\beta B\varepsilon_0^3}}\,dx && (3.38)\\ &\approx k'_{nm}\,\beta e^{\frac{4}{27}(\beta B\varepsilon_0)^3}\frac{4}{3}\sqrt{\pi\beta B\varepsilon_0^3} && (3.39)\\ &\approx k'_{nm}\,\beta\frac{4}{3}\sqrt{\pi\beta B}\left(\frac{3\hbar q_0}{4\sqrt{2m^*}}\right) F e^{\frac{4}{27}(\beta B)^3\left(\frac{3\hbar q_0}{4\sqrt{2m^*}}\right)^2 F^2}. && (3.40)\end{aligned}$$

Again, the result stays the same as in the voltage limited region, as the maximum of the integrand in the medium field region is located between 0 and ε_{max}. This leads to the same exponential increase of the rates with F^2 for medium electric fields.

For $B < 0$, the maxima is always at $\varepsilon = 0$ and the rates can be approximated again by dropping the WKB factor from the integral changing the upper bound of the integration to ε_0.

$$k''_{nm,C-M} \approx k'_{nm}\,\frac{1}{B}\left(e^{\beta B\left(\frac{3\hbar q_0}{4\sqrt{2m^*}}\right)^{2/3} F^{2/3}} - 1\right) \qquad (3.41)$$

H - High fields: At high fields, the WKB factor becomes small for all energies and the effective energy barrier stops getting lower. The rate k'' saturates with $e^{-C/F}$. The total rate can be approximated as the product of all states and the WKB factor at the weighted average position of the states.

$$
\begin{aligned}
k''_{nm,C-H} &\approx k'_{nm}\,\beta e^{-\frac{4\sqrt{2m^*}}{3\hbar q_0}\frac{1}{F}\varepsilon_w^{\frac{3}{2}}} \int_0^{\varepsilon_{max}} e^{\beta B\varepsilon} d\varepsilon && (3.42)\\
&\approx k'_{nm}\left(e^{\beta B\varepsilon_{max}} - 1\right)\frac{1}{B} e^{-\frac{4\sqrt{2m^*}}{3\hbar q_0}\frac{1}{F}\varepsilon_w^{\frac{3}{2}}} && (3.43)
\end{aligned}
$$

With the weighted average of the states energy

$$
\begin{aligned}
\varepsilon_w &= \frac{\int_0^{\varepsilon_{max}} \varepsilon e^{\beta B\varepsilon} d\varepsilon}{\int_0^{\varepsilon_{max}} e^{\beta B\varepsilon} d\varepsilon} && (3.44)\\
&= \frac{e^{\beta B\varepsilon_{max}}}{e^{\beta B\varepsilon_{max}} - 1} - \frac{1}{\beta B}. && (3.45)
\end{aligned}
$$

In this region the rates increase exponentially with -1/F.

Conclusion

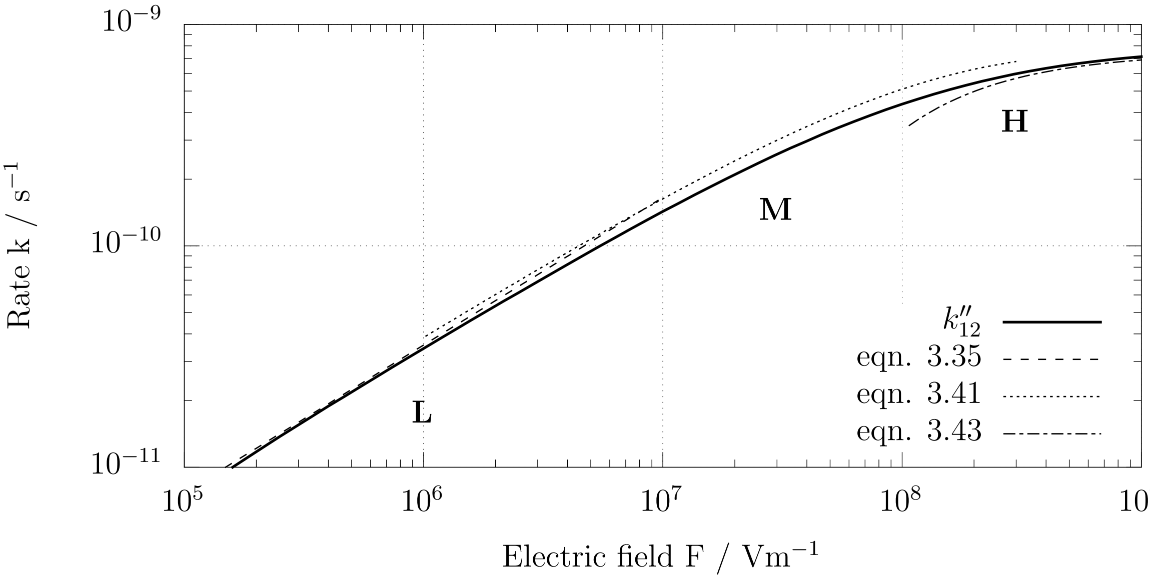
To conclude this section, the field dependences of the simplified rates k'' evaluated in this chapter are shown in table 3.2.

Table 3.2: Field dependences of rate k''

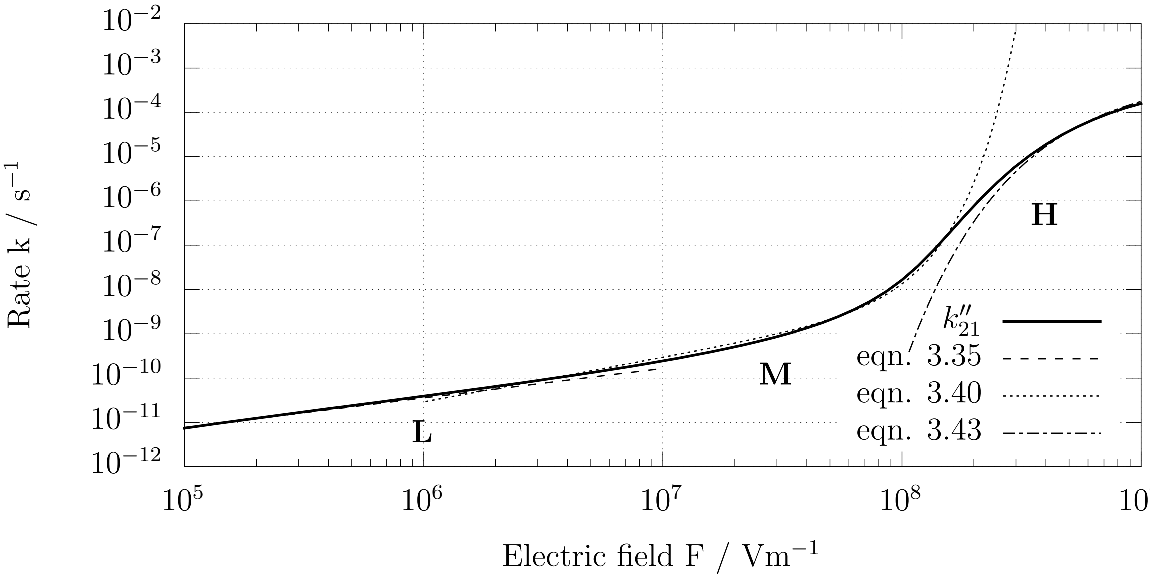
Region		**Dependence**	
L - Low Fields	-	linear	$F^{2/3}$
M - Medium Fields	$B > 0$	exponential	F^2
M - Medium Fields	$B < 0$	exponential	$-F^{2/3}$
H - High Fields	Voltage limited	exponential	BF
H - High Fields	Field limited	exponential	$-1/F$

In figure 3.6, plots for the field dependent rates with their simplifications are shown. They show good agreement indicating that the approximation of the NMP barrier lowering with a constant B (equation 3.7) works well for the trap tested.

More detailed plots of the energy distributions of the integrands in k''_{nm} can be found in section 3.1.3.



(a) Capture rate k''_{12}



(b) Emission rate k''_{21}

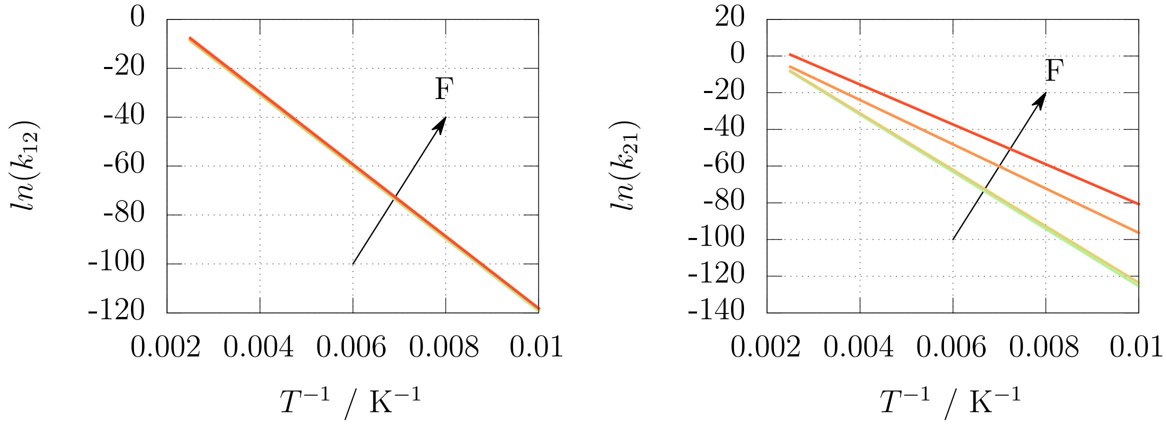
Figure 3.6: Plots of the capture and emission rates k'' and the approximations made in this section.

3.1.2 Temperature Dependence

Since both parts of the rates require thermal activation they must obey the Arrhenius equation [32]

$$k = A\, e^{-\frac{E_a}{k_B T}} \tag{3.46}$$

with the activation energy E_a, the Boltzmann constant k_B, the temperature T and a prefactor A. This can be shown by plotting the rates at different temperatures in an Arrhenius plot. In figure 3.7 this was done for the capture and emission rates of our sample trap for temperatures from 100 K to 400 K and electric field strengths of 10^6 V/m to 10^9 V/m.



(a) Capture rate k_{12}

(b) Emission rate k_{21}

Figure 3.7: Arrhenius plots showing the capture and emission rates k_{12} and k_{21} at field stengths 10^6, 10^7, 10^8 and 10^9 V/m.

For the lowest field strength plotted, the Arrhenius plots for capture and emission are very similar. At higher fields the curves for electron capture shift slightly upwards due to the higher number of states reachable but its slope stays effectively constant. The curves for emission on the other hand change both their slope and height, indicating an additional change in the activation energy. The slope of the curve changes most between 10^7 and 10^8 V/m. At higher fields it shifts mainly in magnitude, indicating that the increase in this region comes from the change in tunneling distance.

Using the Arrhenius equations, apparent activation energies E_a can be extracted from the slope of the rates. These activation energies are plotted in figure 3.8 for capture and emission.

$$E_A = k_B\, ln\left(\frac{k_2}{k_1}\right)\left(\frac{1}{T_1} - \frac{1}{T_2}\right)^{-1} \tag{3.47}$$

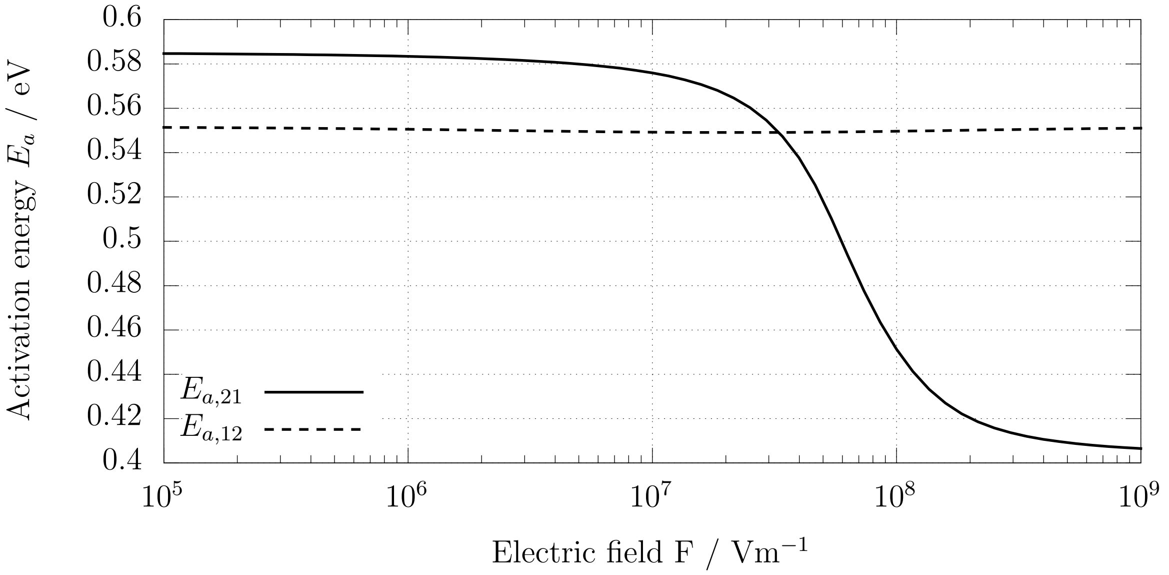


Figure 3.8: Apparent activation energies for emission and capture $E_{a,21}$ and $E_{a,12}$ over the electric field extracted from the temperature dependence of the rates k_{21} and k_{12}.

In figure 3.8 the apparent activation energy for electron emission starts declining at 10^7 V/m. At this point the lowering of the NMP barrier (proportional to ε) starts dominating over the WKB limitation (proportional to $\varepsilon^{3/2}$ and $1/F$) in a wider energy range. The lower energies gain more influence in the rate integral and thus the apparent activation energy falls until the field dependent exponent becomes negligible and the rates reach their thermal limit.

This does not happen for electron capture for this trap because both the NMP barrier and the tunneling barrier increase with higher $|\varepsilon|$. This is explained in more detail in the next section.

3.1.3 Energy Distribution

To get a more detailed understanding of the rates and their dependences on electric field evaluated in section 2.3.4, knowledge about the behaviour of the integrand in k''_{nm} is required.

In this section a closer look will be taken on the integrands $P_T(\varepsilon)$, the probability of thermal excitation and $\lambda(\varepsilon)$, the WKB factor.

Capture Rate k_{12}

Figure 3.9 shows integrand product $P_T(\varepsilon)\lambda(\varepsilon)$ over ε for different electric field strengths for the capture rate k_{12}.

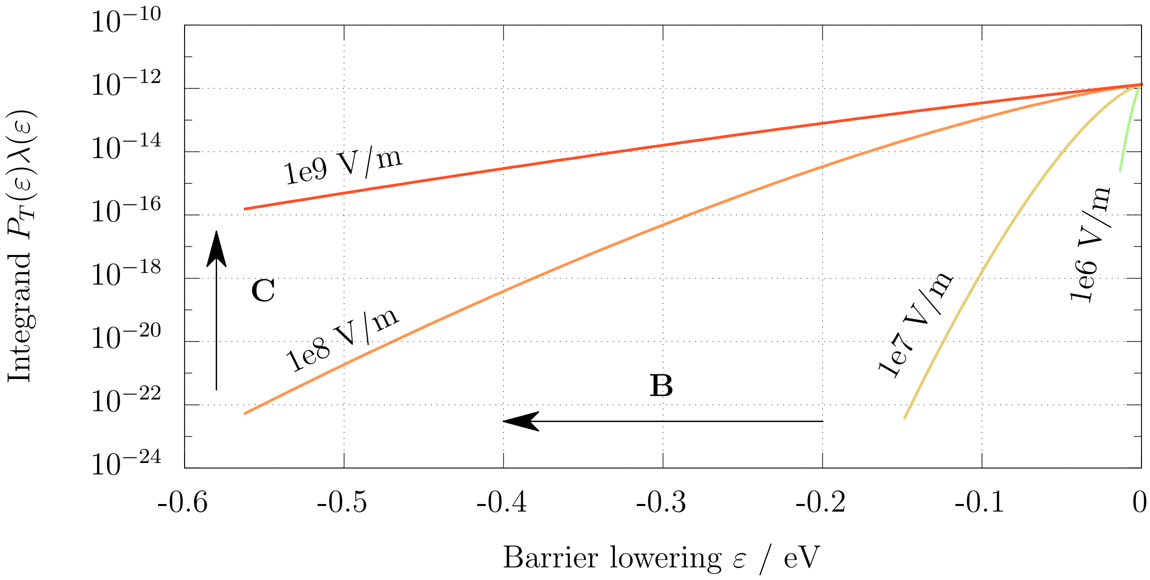


Figure 3.9: Energy distribution of the integrand $P_T\lambda$ of capture rate k''_{12} for electric field strengths 10^6,10^7,10^8 and 10^9 V/m. **B** and **C** refers to the voltage and field limited general regions, see figure 3.4.

For low electric fields, the possible barrier lowering ε is limited by the device length. In this region, termed voltage limited before, the width of the integral increases with the field. At higher fields, all thermal states can be reached and the rates are governed only by the increase of the WKB factor. As the tunneling distances decrease the integrand converges to its thermal limit. Tunneling close to the conduction band edge at the trap position always dominates the rates because with the chosen parameters the NMP barrier height for electron capture increases with lower band energies.

Detailed plots of the contributions of $\lambda(\varepsilon)$ and $P_T(\varepsilon)$ at different field strengths can be found in figure 3.10.

The top two plots show the assumption made in section 2.3.4, that the thermal barrier stays effectively constant for low fields, is justified. At higher electric fields, shown in the third plot, both the thermal probability and the WKB factor decrease the contributions at low energies. The lower right plot shows the situation at high electric fields, where the integral is essentially an integration over the thermal probability.

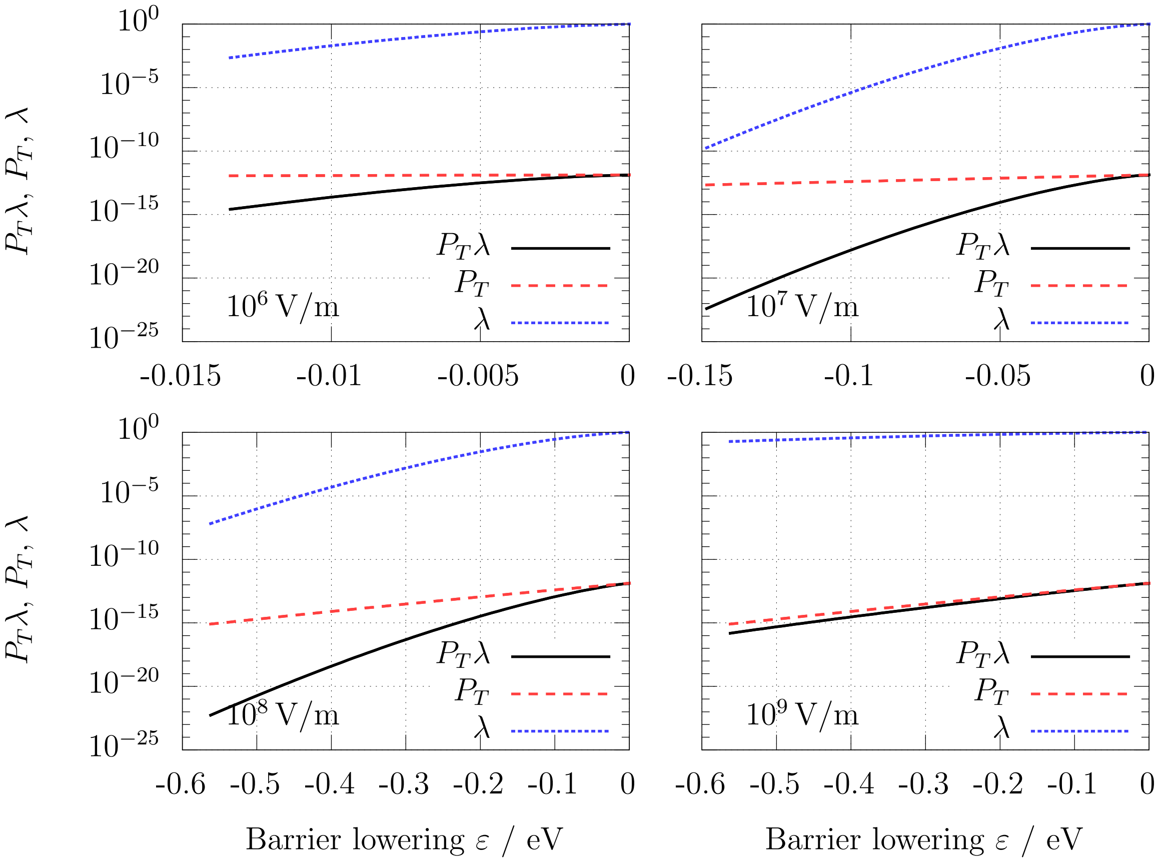


Figure 3.10: WKB factors $\lambda(\varepsilon)$, propability of thermal excitation $P_T(\varepsilon)$ and their product $P_T\lambda$ over energy for capture and field strengths 10^6 V/m to 10^9 V/m.

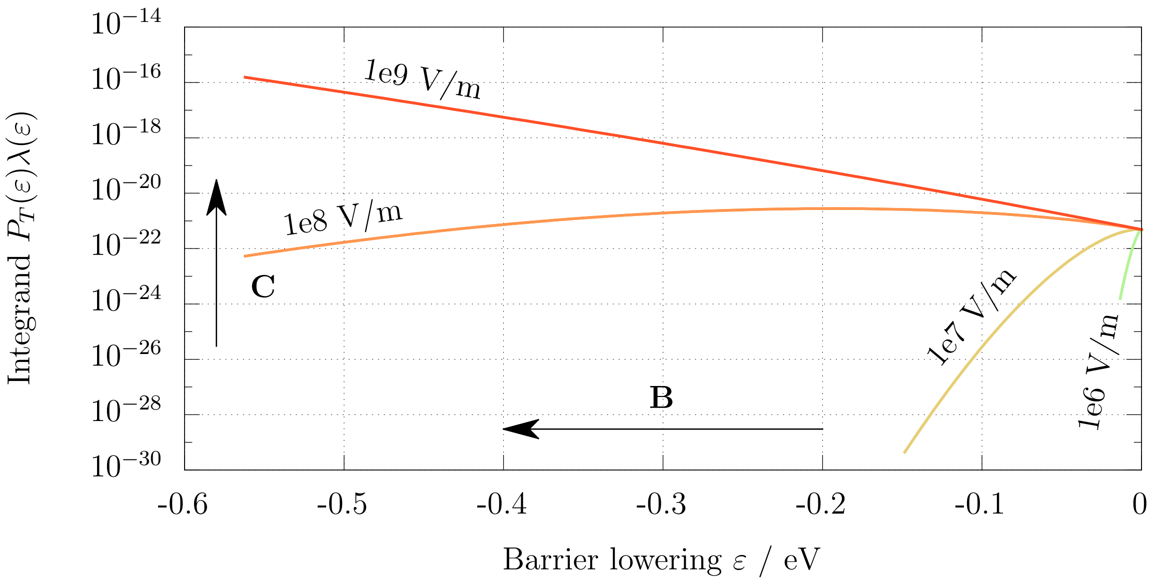


Figure 3.11: Energy distribution of the integrand $P_T\lambda$ of emission rate k''_{21} for electric field strengths 10^6,10^7,10^8 and 10^9 V/m. **B** and **C** refers to the voltage and field limited general regions, see figure 3.4.

Emission Rate k_{21}

Figure 3.11 shows integrand product $P_T(\varepsilon)\lambda(\varepsilon)$ over ε for different electric field strengths for the the emission rate k_{21}.

Again, for low electric fields, the possibility of emitting to lower energies is limited by the device length and the width of the integral increases with the field. At higher fields, all thermal states can be reached and the rates are governed by the increase of the WKB factor. As the tunneling distances decrease, the maximum of the distribution shifts to lower energies, which is what caused the apparent activation energy in section 2.4 to decrease.

Figure 3.12 shows the detailed plots with the contributions of $\lambda(\varepsilon)$ and $P_T(\varepsilon)$ for the emission rate.

As before, the plot for 10^6 V/m and 10^7 V/m show than the assumption made in section 2.3.4 is justified. The third plot for 10^8 V/m shows the integrand in region C, with the thermal factor dominating for high energies and the WKB factor dominating for low energies. At 10^9 V/m, the thermal factor dominates the integral for all energies and the WKB factor merely controls saturation.

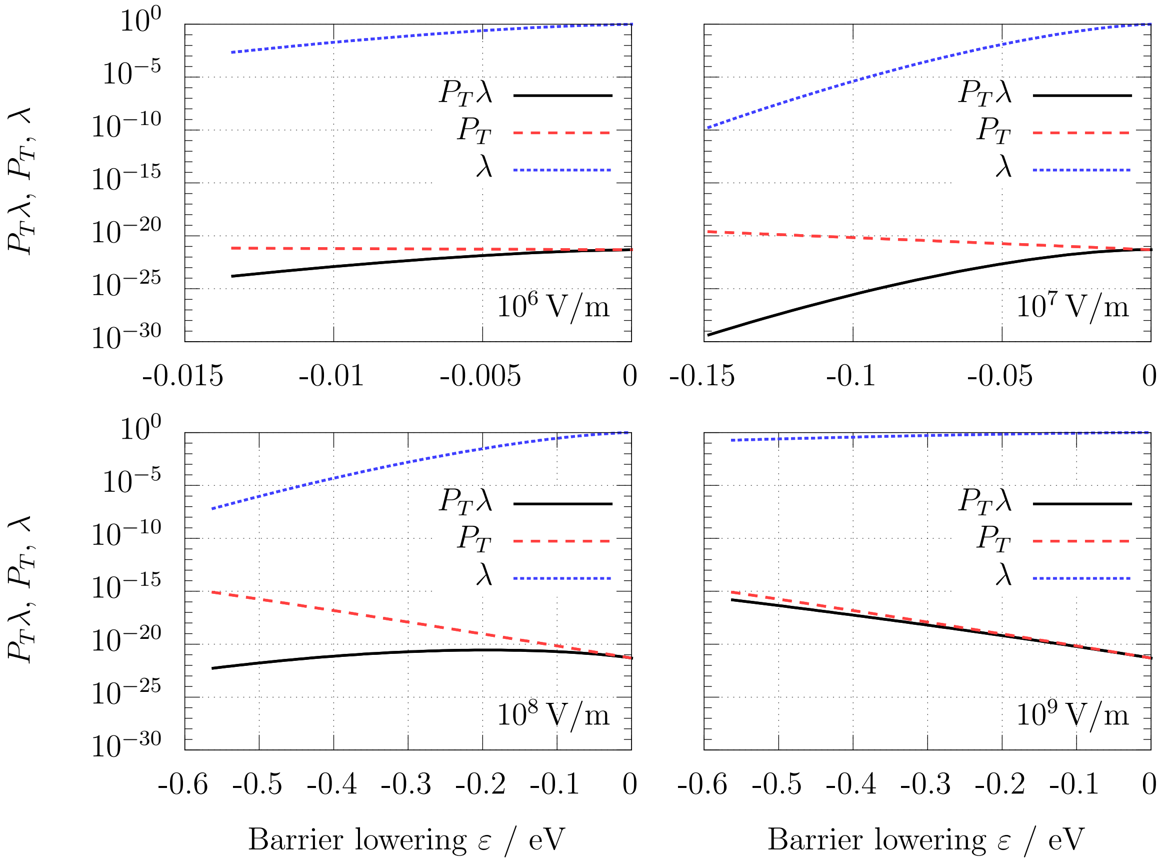


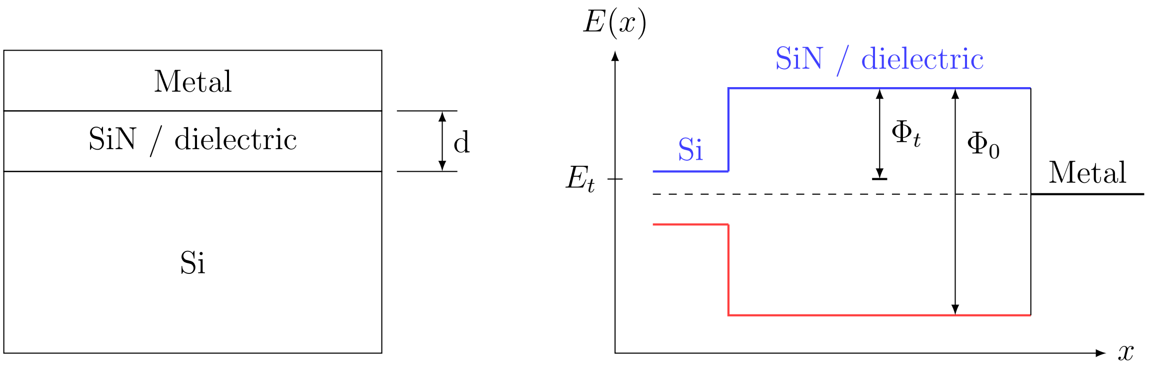
Figure 3.12: WKB factors $\lambda(\varepsilon)$, propability of thermal excitation $P_T(\varepsilon)$ and their product $P_T\lambda$ over energy for emission and field strengths 10^6 V/m to 10^9 V/m.

3.2 Significance of the Local Band Interactions

Now that we discussed the behaviour of the NMP rates to the local conduction and valence bands, we take a look at their effect on the behaviour of traps in a more practical setup. In this section we will compare capture and emission times calculated with and without the rates to the local bands in the semi-insulating layer of a MIS device.

The capture and emission times τ_c and τ_e are the expected time it takes a trap to capture or emit a carrier after the last emission or capture event.

The structure tested in this section is a MIS structure, consisting of a silicon bulk, a dielectric layer and a metal contact on top. An illustration of the simulation setup can be seen in figure 3.13.



(a) Geometry (b) Band diagram

Figure 3.13: Test structure used in section 3.2. A metal contact is insulated from the silicon bulk by a dielectric layer. For the simulations SiN was used as material for the dielectric.

The dielectric layer has a thickness d of either 3 nm or 6 nm and the trap is always located in the middle, at d/2. The energetic position of the trap is characterized either by its absolute energetic position in the band diagram E_t or by its energetic offset to the conduction band edge Φ_t. The band gap energy of the dielectric layer is defined by Φ_0. Other parameters of the trap and the dielectric layer are shown in table 3.3. The values for the effective mass and permittivity were taken from SiN, however their exact values does not change the qualitative behaviour of the results discussed.

The trap in this device has four reservoirs to interact with: silicon, gate and the SiN conduction and valence bands. For a given set of trap parameters and constant gate voltage, we can expect the interaction with the silicon and metal reservoirs to be mainly governed by the distance from the trap to the reservoir and their relative energetic position, while the interaction with the conduction and valence bands will depend heavily on the trap depth Φ_t.

Table 3.3: Simulation parameters for the evaluation capture and emission times with and without local band interaction in section 3.2

Parameter	**Value**	**Description**
$R12p$	0.6	Ratio of curvatures for transition 1-2'
$S12p$	$2.3\,\mathrm{eV}$	Huang-Rhys parameter S times $\hbar\omega$ for transition 1-2'
$R1p2$	0.6	Ratio of curvatures for transition 1'-2
$S1p2$	$0.7\,\mathrm{eV}$	Huang-Rhys parameter S times $\hbar\omega$ for transition 1'-2
ε_{T2}	$0.7\,\mathrm{eV}$	Energy difference between states 2' and 2
ε_{1p1}	$0.4\,\mathrm{eV}$	Thermal barrier between states 1 and 1'
ε_{2p2}	$0.4\,\mathrm{eV}$	Thermal barrier between states 2 and 2'
σ	$1 \times 10^{-23}\,\mathrm{cm}^2$	Capture cross section
ε_r	7.0	Relative permittivity of the dielectric layer
m^*	$0.321\ m_0$	Effective mass of the dielectric layer

With this in mind, we will compare the capture and emission times over the band gap energy for two different device lengths and the cases of a deep trap with constant trap level E_t relative to the silicon and metal reservoirs and a more shallow trap with a constant trap level Φ_t relative to the energies of conduction and valence bands.

Constant E_t, 6 nm dielectric

We will start by taking a look at simulation results for the longer, 6 nm device and constant trap parameters $E_1 = 0.4\,\mathrm{eV}$ and $E_{1'} = 0.5\,\mathrm{eV}$ for the energies of the stable and metastable trap states. The voltage applied at the metal is 2.0 V which results in an electric field in the dielectric of about $2.3 \times 10^8\,\mathrm{V/m}$. The constant trap level and the constant voltage in this simulation leads to a constant energetic offset between the trap states and the silicon and metal reservoirs.

Figure 3.14 shows the capture and emission times τ_c and τ_e over the band gap energy Φ_0.

The first thing noticeable is that while the emission times shows a significant lowering with band interaction enabled, the capture rates stay effectively the same. This is due to the very low amount of carriers available for capturing in the conduction band of the dielectric.

The capture and emission times calculated without band interaction both show a similar increase with the band gap energy in the logarithmic plot. As we keep the energetic offset between the trap and these reservoirs constant, this increase can only be attributed to the higher tunneling barrier between the trap and the reservoirs.

For electron emission with band interaction enabled, up to around 5 eV bandgap, the emission to the local band causes a significant decrease in τ_e. The emission time in this

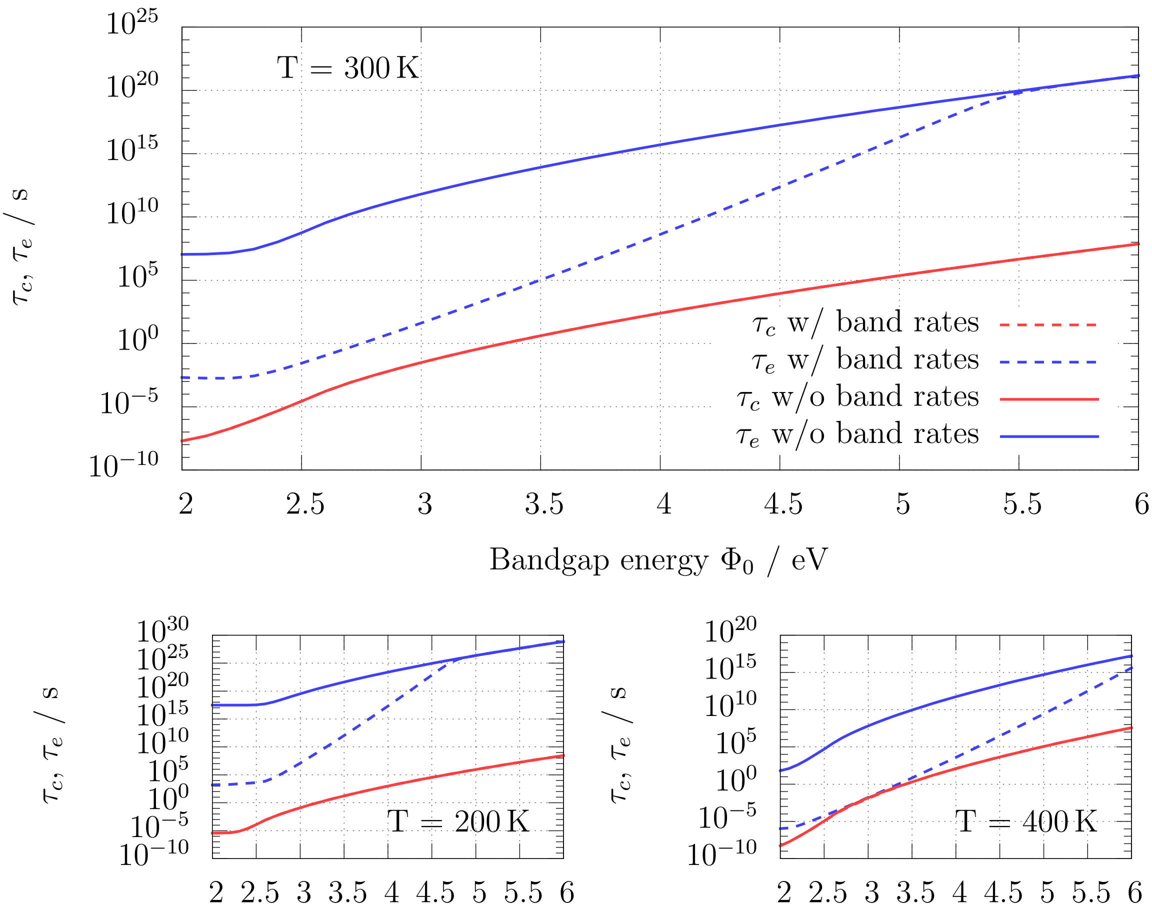


Figure 3.14: Simulated capture and emission times with and without enabled band rates over the band gap energy. The trap is at a fixed energy in the middle of a 6 nm dielectric.

area increases exponentially with the band gap energy, as it changes the energetic offset between trap and conduction band. With higher temperature, the slope of the emission time gets lower as this barrier is easier to overcome.

Constant E_t, 3 nm dielectric

For the shorter device, a lower voltage of 1.5 V was applied to the device, resulting in an electric field strength of about 2.9×10^8 V/m. The resulting capture and emission times can be seen in figure 3.15.

With the thickness of the dielectric layer $d = 3$ nm the tunneling distances to the silicon and metal reservoirs decrease to 1.5 nm. This has a strong effect on the capture and emission times and decreases them significantly compared to the results for the longer dielectric.

At low temperatures, the effect of the rates to the band becomes completely neglegible even at low band gap energies. For higher temperatures again a lowering of the emission times can be seen at band gap energies of up to 3 eV.

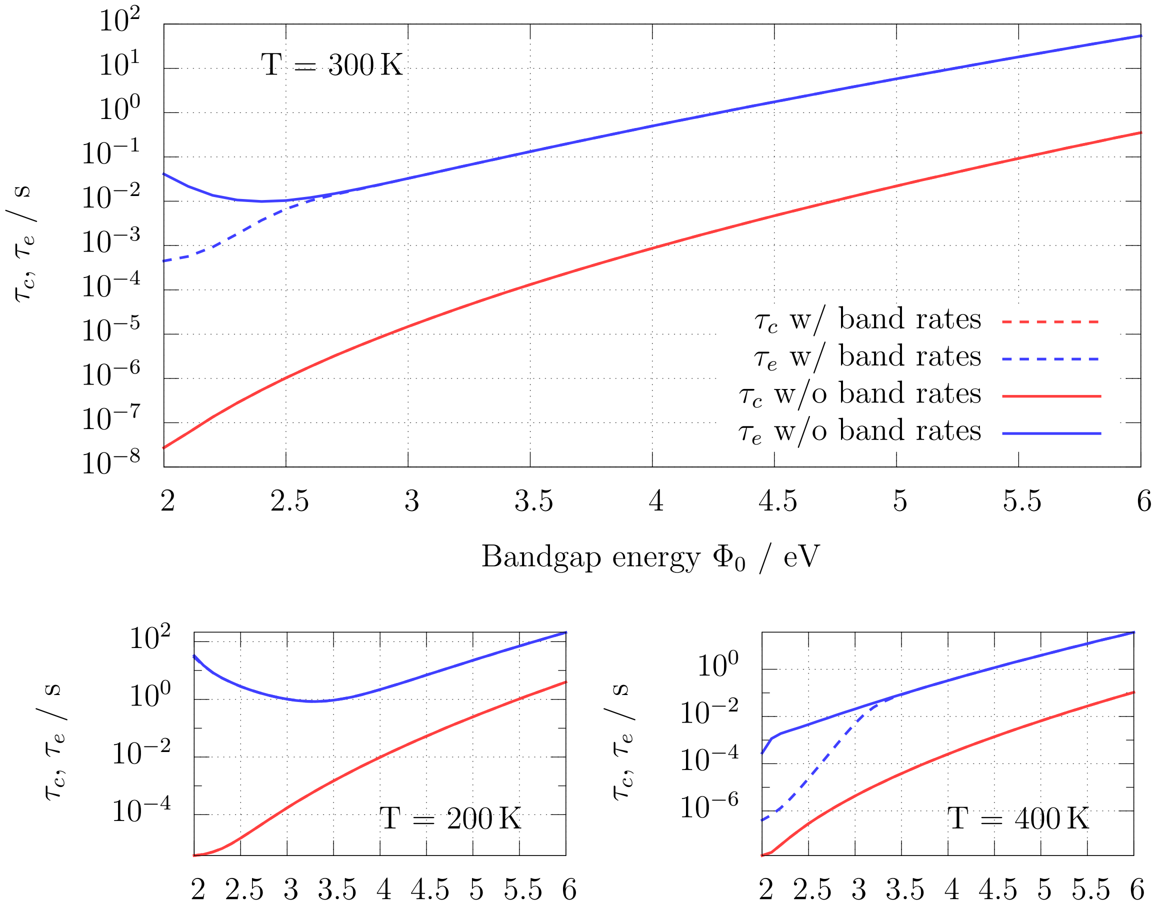


Figure 3.15: Simulated capture and emission times with and without enabled band rates over the band gap energy. The trap is at a fixed energy in the middle of a 3 nm dielectric.

Constant Φ_t, 6 nm dielectric

Now, instead of keeping the level of the trap constant with respect to the silicon and metal reservoirs, we keep the trap level at a constant offset to the conduction band edge. This will shift the interaction of the trap with the silicon and metal reservoirs into focus. It also means that the trap will now be shallow at high band gaps compared to the deep trap in the previous part.

To fix Φ_t at 0.8 eV, the trap parameters were defined as $E_1 = E_{1'} = \Phi_0/2 - 0.8\,\mathrm{eV}$. The resulting plots of the capture and emission times for the longer device with $d = 6\,\mathrm{nm}$ can be seen in figure 3.16.

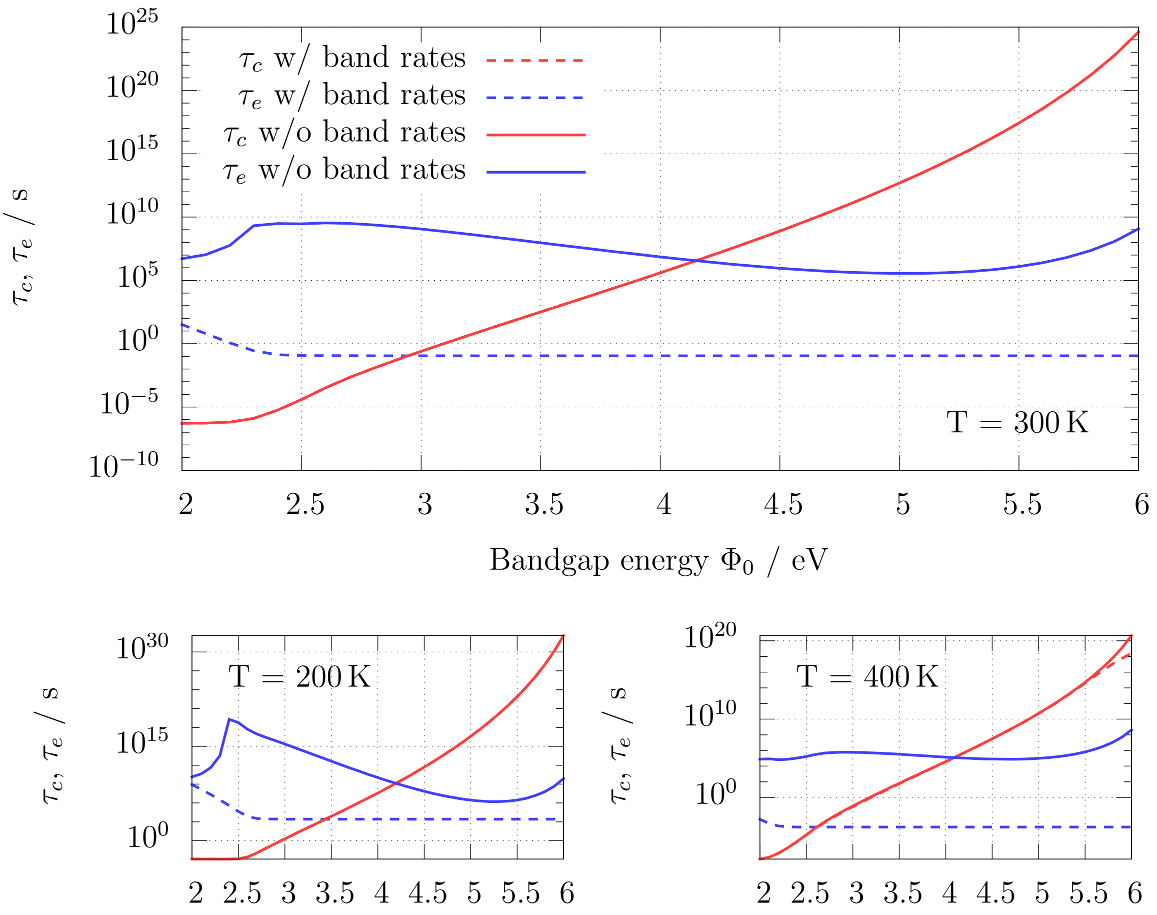


Figure 3.16: Simulated capture and emission times with and without enabled band rates over the band gap energy. The trap is at an energy of $E_c - 0.8\,\mathrm{eV}$ and located in the middle of a 6 nm dielectric.

Due to the relatively low and constant energetic distance of the trap to the conduction band, the emission times now stay effectively constant for all band gap energies if the interaction with the bands is enabled.

With the band rates disabled, the emission and capture times now show the behaviour of the NMP barriers to and from the metal and silicon reservoirs. Their height depends on the intersection of the parabolas representing the adiabatic potential energy surface in the configuration coordinate diagram. At higher temperatures the NMP barriers are easier to overcome and their effect on the capture and emission times decrease.

Again, the plots show no effect of the band interaction on the capture rate due to the very low amount of free carriers in the conduction band of the semi-insulating layer.

Constant Φ_t, 3 nm dielectric

Finally, the plots for constant Φ_t and the short device are shown in figure 3.17.

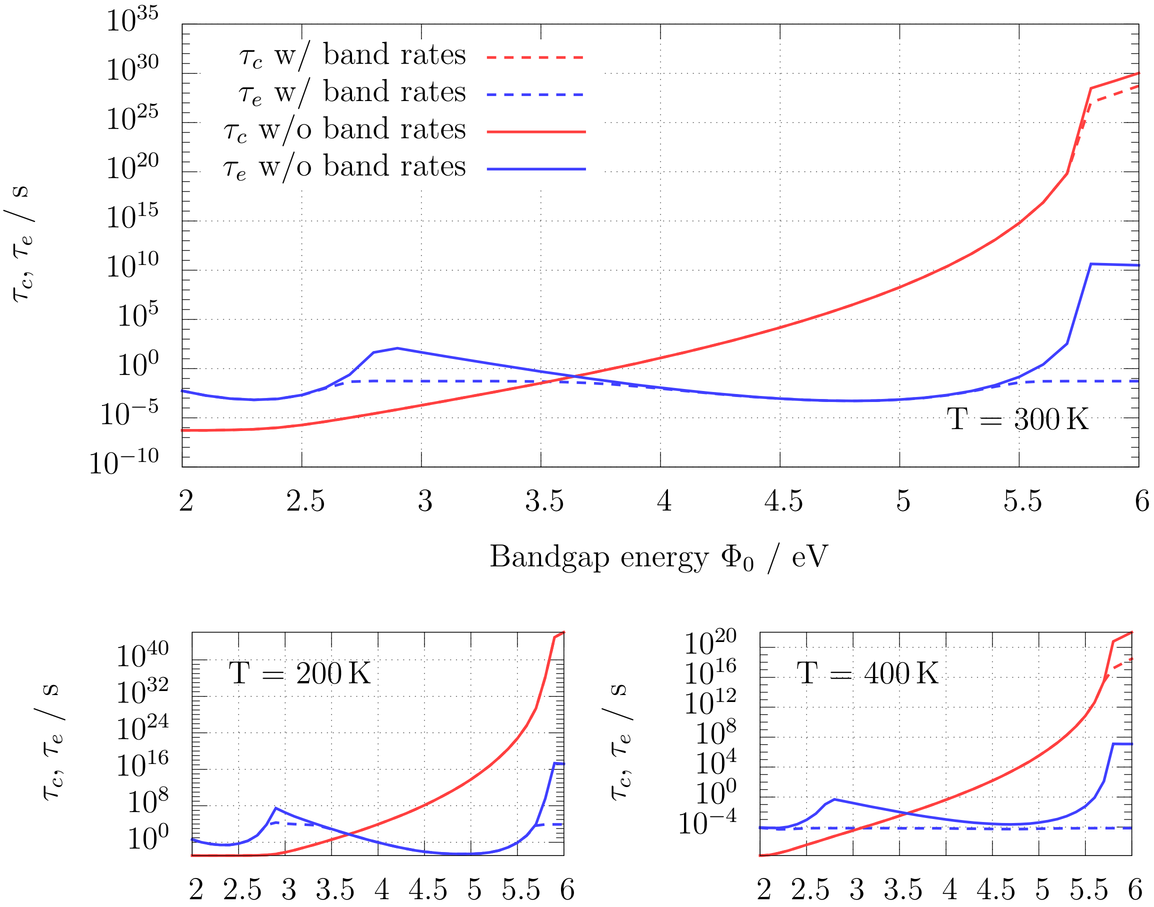


Figure 3.17: Simulated capture and emission times with and without enabled band rates over the band gap energy. The trap is at an energy of $E_c - 0.8$ eV and located in the middle of a 3 nm dielectric.

Due to the shorter distance to the silicon and metal reservoirs, the times calculated without the band rates are now a few orders of magnitude smaller than for the longer device. Their qualitative behaviour stays the same, which reassures the earlier statement that their behaviour is governed by the NMP barriers. The plots with the band interaction enabled show no clear domination of either mechanism. For high temperatures the emission to the bands dominates more often while for low temperature the emission to silicon and metal shows is dominant.

To conclude this section, it can be said that capturing from the band can be neglected for dielectric layers with low doping and few free carriers. The significance of emission to the band depends on temperature, trap depth and device length. It increases for higher temperatures, longer devices and, in general, shallower traps. Emission should never be neglected for shallow traps with $\Phi_t < 1$ eV. For deep traps it may be neglected for high band gap energies above about 6 eV or thin layers under about 3 nm.

3.3 Trap Assisted Tunneling

As discussed in section 2.2.2, a trap with more than one reservoir will in general transport charge between the reservoirs if it is not in thermal equilibrium.

This tunneling current resulting from traps is responsible, at least partially, for various leakage phenomena in semiconductor devices. Two scenarios will be shown in this section. In the first part, a simulation of trap induced leakage currents in a high electron mobility transistor (HEMT) using the NMP model will be discussed and compared to the analytical Frenkel-Poole model and the Fowler-Nordheim tunneling mechanisms. In the second part, stress, recovery and leakage currents resulting from oxide traps in a 2D MOS structure will be evaluated and discussed.

3.3.1 Reverse Leakage Currents in a HEMT

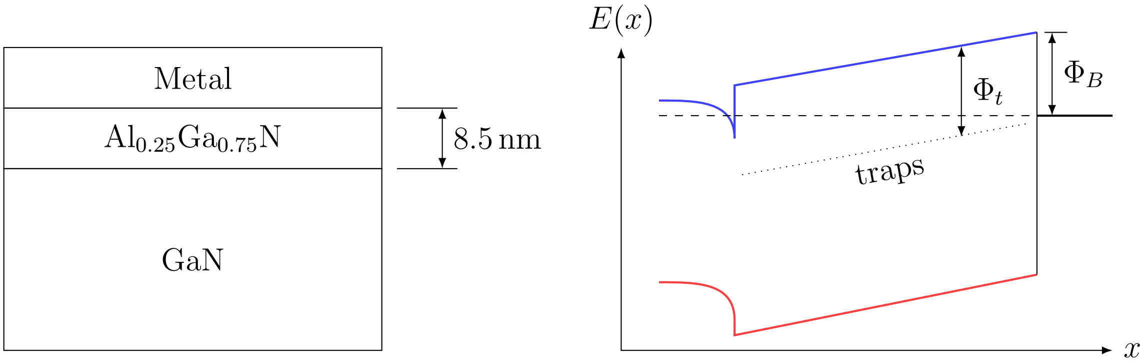
A phenomenon where traps located in a semiconducting material are a key factor are the reverse leakage currents found in GaN/AlGaN high electron mobility transistors (HEMTs).

Measurements [33, 34] have found leakage currents which show Frenkel-Poole like behaviour at low bias and high temperatures and Fowler-Nordheim like behaviour at high bias and at low temperatures. In this section we will simulate a HEMT with the NMP model and compare the resulting currents with the Frenkel-Poole and the Fowler-Nordheim models. The simulation of the HEMT was done one dimensional, which means that source and drain contacts were not simulated. This is reasonable as, like stated in chapter 1 and according to [20], the leakage current from gate to channel dominates for AlGaN/GaN HEMTs.

To evaluate the currents, a test structure as depicted in figure 3.18 is used. The test structure is a GaN/AlGaN heterostructure with a Schottky contact on top of an AlGaN barrier layer and a GaN bulk. The $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{N}$ layer has a thickness of 8.5 nm and an Al content of x = 0.25.

The traps are distributed uniformly across the barrier layer and their energetic position is slightly below the Fermi level at the Schottky contact. The trap and simulation parameters can be found in table 3.4.

This simulation again uses the extension of the NMP rates to the local bands derived in section 2.3.2. This allows for two trap assisted leakage paths through the barrier. The first path is traps capturing from the metal and emitting directly to the GaN conduction band. The second is traps capturing from the metal and emitting to the AlGaN conduction band. Judging from the results of the previous section 3.2, the emission to the band should dominate for all traps close to the metal as their distance to the GaN conduction band is large. For traps close to the channel, emission to the channel should dominate over emission to the local conduction band, but these traps



(a) Geometry (b) Band diagram

Figure 3.18: Test structure used in section 3.3.1. A metal Schottky contact is on top of an AlGaN barrier and a GaN bulk.

Table 3.4: Simulation parameters for the discussion of TAT in a HEMT in section 3.3.1.

Parameter	**Value**	**Variance**	**Description**
$E1$	2.8 eV	-	Trap level
$R12$	0.5	0.1	Ratio of curvatures for transitions 1-2
$S12$	0.6 eV	0.1 eV	Huang-Rhys parameter S times $\hbar\omega$
σ_0	1×10^{-23} cm^2	-	Capture cross section
Φ_B	0.8 eV	-	Schottky barrier
ε_r	8.5	-	Relative permittivity of the barrier layer
m^*	0.2 m_0	-	Effective mass in the barrier layer

generally have large capture times from the metal, again due to the large distance, which limits their influence on the leakage current.

The resulting tunneling currents over the electric field and for temperatures ranging from 200 K to 350 K can be found in figure 3.19.

Comparison With Other Models

In the papers that inspired these simulations, the measured leakage currents are compared to currents resulting from the Fowler-Nordheim and the Frenkel-Poole models.

In [33], Zhang et al. measured an AlGaN/GaN Schottky device over a wide range of temperatures and found the leakage current to show Frenkel-Poole like behaviour at high temperatures and Fowler-Nordheim like behaviour at very low temperatures.

Ganguly et al. measured the leakage current of an InAlN/AlN/GaN HEMT [34] and found it to resemble Fowler-Nordheim emission at high electric fields. To analytically

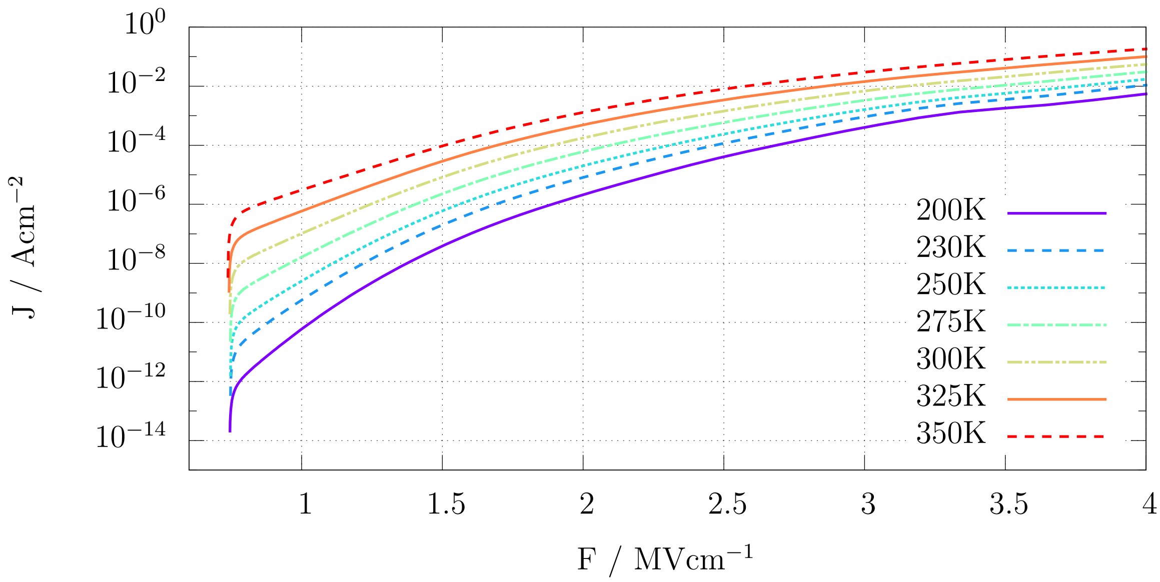


Figure 3.19: Tunneling current through the barrier layer of a GaN/AlGaN heterostructure resulting from the NMP traps in the barrier. Different colours show the current at temperatures ranging from 200 K to 350 K.

describe the current at low fields, they used a modification of the Frenkel-Poole model to account for the polarization fields found in III-Nitride devices.

Fowler-Norheim model

The Fowler-Nordheim model describes the current emitted from a metal through a triangular barrier and is usually written in the form

$$J_{FN} = CF^2 e^{-\frac{A}{F}} \tag{3.48}$$

with a pre-factor C, and

$$A = \frac{4\sqrt{2m_T^*(q\Phi_B)^3}}{3q\hbar}. \tag{3.49}$$

Here, $\hbar$ is the reduced Planck constant, m_T^* the effective tunneling mass and Φ_B the barrier height in eV. For the parameters $m_T^* = 0.2\,m_0$ and $\Phi_B = 0.8\,\mathrm{eV}$, A evaluates to $2.19 \times 10^9\,\mathrm{m/V}$.

Current data believed to follow a Fowler-Nordheim behaviour are commonly plotted as $log(J/F^2)$ on the vertical axis against $1/F$ on the horizontal axis. This is called a Fowler-Nordheim plot and the resulting curve should be a straight line. The slope of the line is defined by the parameters Φ_B and m^* in the exponent of the Fowler-Nordheim equation.

A Fowler-Nordheim plot of the currents simulated with the NMP model at different temperatures and the analytically calculated current for the parameters m^* and ϕ_B are shown in figure 3.20 for electric field strengths from $2\,\mathrm{MVcm}^{-1}$ to $4\,\mathrm{MVcm}^{-1}$.

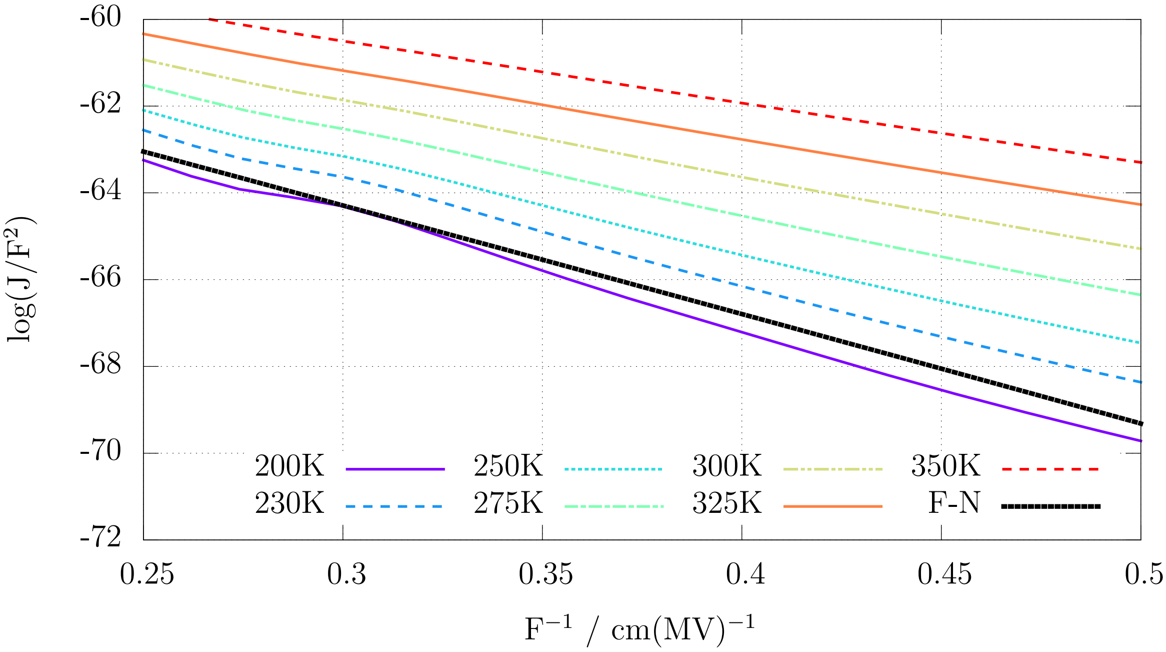


Figure 3.20: Fowler-Nordheim plots of the simulated NMP tunnel current and the analytical Fowler-Nordheim current for a barrier height of $\Phi_B = 0.8\,\mathrm{eV}$ and $m_T^* = 0.2\,m_0$.

The first thing that comes to mind when looking at this plot is the temperature dependence. The Fowler-Nordheim emission current calculated from the analytical formula shows no dependence on temperature while the NMP model generally does so. This is however not in contradiction with the measurement results obtained from Zhang et al. and Ganguly et al. who found the currents to follow Fowler-Nordheim behaviour best at low temperatures and high fields. It can be seen from the plots that the temperature spread decreases with increasing field strength. The slopes of the curves decrease with temperature which indicates a lowering of the effective barrier with temperature.

Frenkel-Poole model

The Frenkel-Poole model describes the current produced by carriers thermally emitted from a trap state to a band with the lowering of the thermal barrier by an electric field. The resulting current density can be written as

$$J_{PF} = CFe^{-\frac{q\Phi_t - A\sqrt{F}}{k_B T}} \tag{3.50}$$

with a constant prefactor C, the trap depth Φ_t, and

$$A = q_0 \sqrt{\frac{q_0}{\pi \varepsilon_0 \varepsilon_s}}. \tag{3.51}$$

Here, q_0 is the value of the electron charge, ε_0 the permittivity of free space, and ε_s the relative permittivity of the barrier. For $\varepsilon_s = 8.5$, A evaluates to 4.17×10^{-24} As(Vm)$^{0.5}$. Ganguly et al. corrected the model for electric fields induced by polarization charge in their work [34]. The modified equation for the current density reads

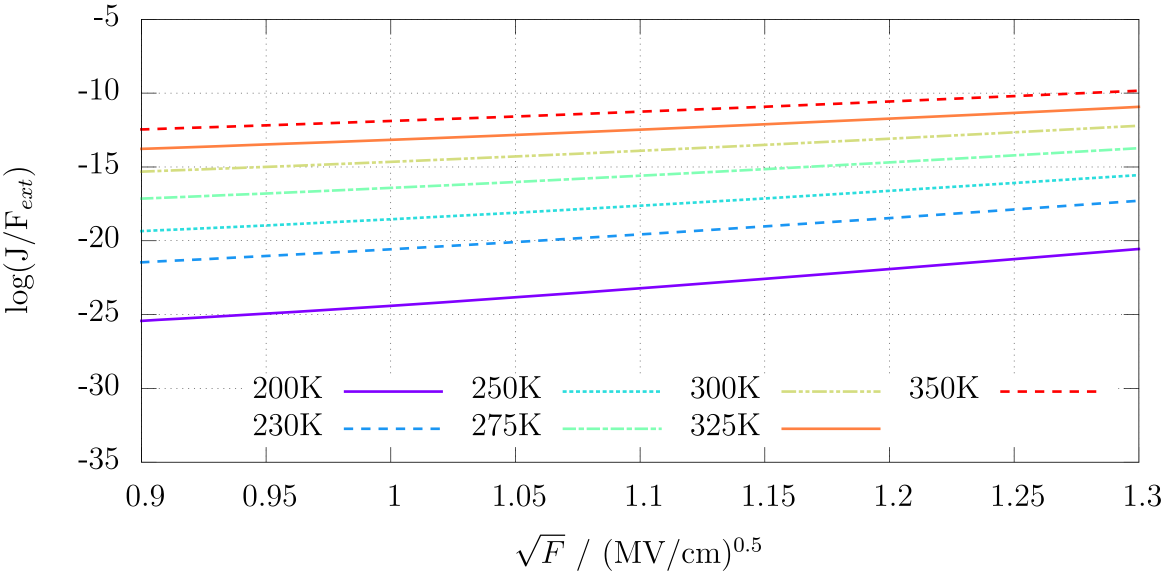
$$J_{PF} = Ce^{-\frac{q\Phi_t}{k_B T}} \left(Fe^{\frac{A\sqrt{F}}{k_B T}} - F_\pi e^{\frac{A\sqrt{F_\pi}}{k_B T}} \right) \tag{3.52}$$

with the electric field induced by polarization charges F_π.

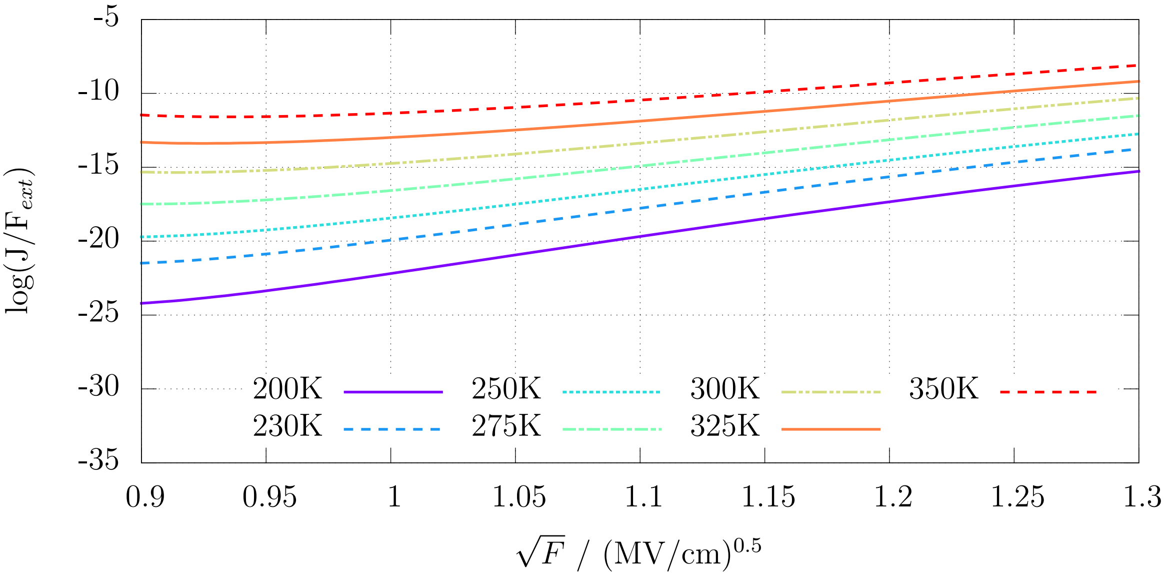
Similar to the Fowler-Nordheim plots before, currents following a Frenkel-Poole behaviour are commonly plotted with $log(J/F)$ on the vertical and $\sqrt{F}$ on the horizontal axis. Again, the resulting curves should be straight lines. For the modified equation, $log(J/F_{ext})$ with the applied field $F_{ext} = F - F_\pi$ is plotted on the vertical axis.

Plots for the currents resulting from the modified Frenkel-Poole equation and the NMP model are shown in figure 3.21 for a electric field range from 0.8 to 1.7 MV/cm.

The figure shows good agreement of the currents for high temperatures and low electric fields. For higher fields the temperature dependence of the NMP model decreases compared to the Frenkel-Poole model. Similar behaviour can also be found in the experimental data obtained by Ganguly et al.



(a) Poole-Frenkel



(b) NMP

Figure 3.21: Plots of the currents resulting from the modified Frenkel-Poole model (a) and the NMP model (b).

Conclusion

The current simulated using the NMP model follows both analytical models depending on both the temperature and the electric field. At high fields and low temperatures, the NMP current expresses behaviour similar to that of the Fowler-Nordheim model. At low fields the NMP currents show a temperature dependence similar to that of the Frenkel-Poole model. The model always shows a temperature dependence, although it decreases with increasing field. This behaviour does not stand in contradiction with the measurement results from Zhang et al. and Ganguly et al. and further allows to model the “crossover region” [34] where neither analytical model is usable. Plots showing the NMP current compared with both the Fowler-Nordheim and the Frenkel-Poole models can be found in figure 3.22.

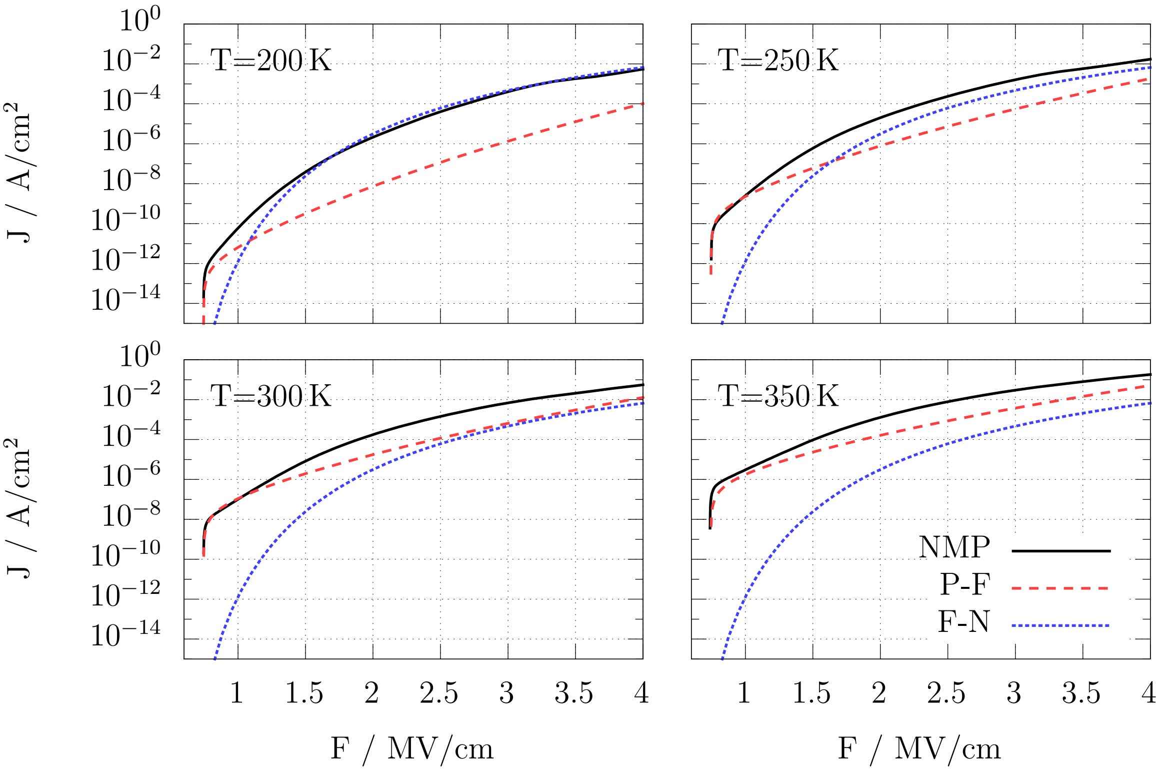


Figure 3.22: Plots of the leakage currents resulting from the NMP model, the Poole-Frenkel and the Fowler-Nordheim model. At low temperature, the NMP current behaves like the Fowler-Nordheim current for most of the field range, while higher temperatures the thermal activation dominates and tunneling dominated behaviour shifts to higher fields.

3.3.2 Stress and Recovery Currents in a MOS Transistor

In the last part of this work, we will focus on the currents resulting from oxide defects in a MOSFET. The simulation data including the trap set were thankfully received from Gerhard Rzepa, who calibrated them to fit measurements on a real device. A schematic drawing of the device can be found in figure 3.23.

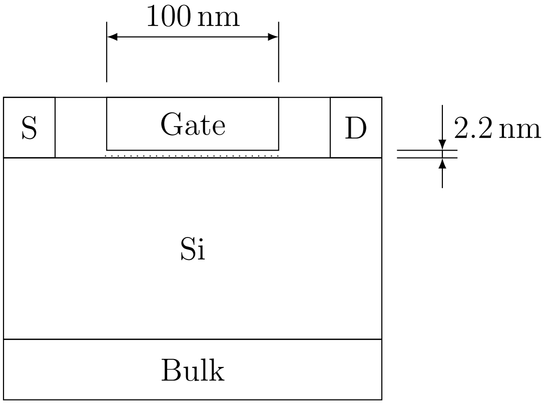


Figure 3.23: Geometry of the MOSFET used in section 3.3.2. The gate oxide is 2.2 nm thick, 100 nm wide and 1.5 μm long SiO_2. The dots in the oxide indicate the positon of the traps.

The device features a 2.2 nm thick, 100 nm wide and 1.5 μm long oxide. The trap density used for the simulation is $2 \times 10^{19}\,\mathrm{cm}^{-3}$ and the traps are distributed homogeneous in the lower half of the oxide. The remaining parameters used for simulation can be found in table 3.5.

Table 3.5: Simulation parameters for the discussion of trap currents in a MOSFET in section 3.3.2

Parameter	**Value**	**Variance**	**Description**
$E1$	$-1.400\,\mathrm{eV}$	$0.260\,\mathrm{eV}$	Trap level in state 1
$E1p$	$-0.395\,\mathrm{eV}$	$0.462\,\mathrm{eV}$	Trap level in state 1’
$R12p$	0.237	0.225	Ratio of curvatures for transition 1-2’
$S12p$	$1.392\,\mathrm{eV}$	$0.364\,\mathrm{eV}$	Huang-Rhys parameter S times $\hbar\omega$ (1-2’)
$R1p2$	0.376	0.332	Ratio of curvatures for transition 1’-2
$S1p2$	$1.063\,\mathrm{eV}$	$0.010\,\mathrm{eV}$	Huang-Rhys parameter S times $\hbar\omega$ (1’-2)
ε_{T2}	$0.343\,\mathrm{eV}$	$0.384\,\mathrm{eV}$	Energy difference between states 2’ and 2
ε_{1p1}	$0.233\,\mathrm{eV}$	$0.271\,\mathrm{eV}$	Thermal barrier between states 1 and 1’
ε_{2p2}	$0.429\,\mathrm{eV}$	$0.279\,\mathrm{eV}$	Thermal barrier between states 2 and 2’
σ_0	$1 \times 10^{-23}\,\mathrm{cm}^2$	-	Capture cross section
T	443 K	-	Temperature

The device was measured in a measure-stress-measure (MSM) setup, as introduced in section 1.1.1. The transient simulation reproduces this measurement, and with the expressions given in section 2.2.2 allows us to extract the resulting currents. The voltages applied to the device during stress and recovery phases are shown in figure 3.24. The gate voltage switches between −2.7 V for stress and −0.5 V for recovery, and the drain voltage between 0.005 V and −0.2 V.

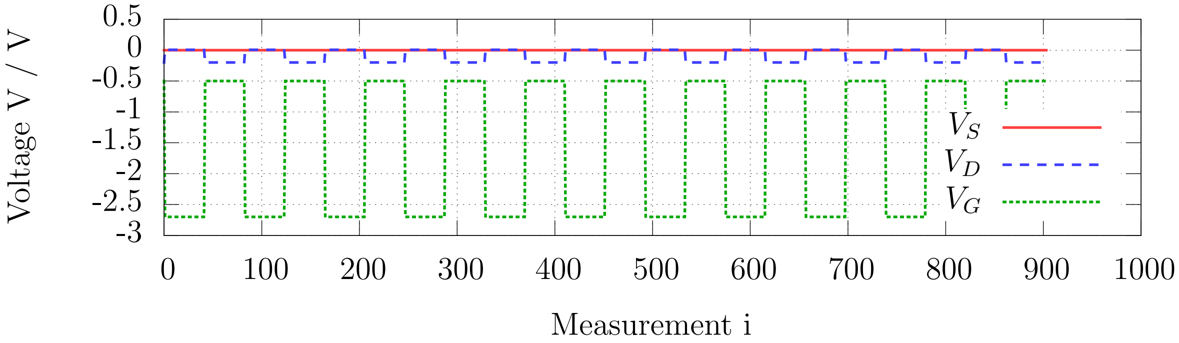


Figure 3.24: Gate, source, and drain voltages V_g, V_s and V_d during MSM measurement.

The sampling points are spread logarithmic across each stress and recovery phase. The durations of the stress phases increase from 1.1×10^{-6} s to 1.1×10^{4} s.

In figure 3.25 the currents from the gate and the channel to the traps are plotted for the fifth stress cycle with a stress time of 10^{-2} s.

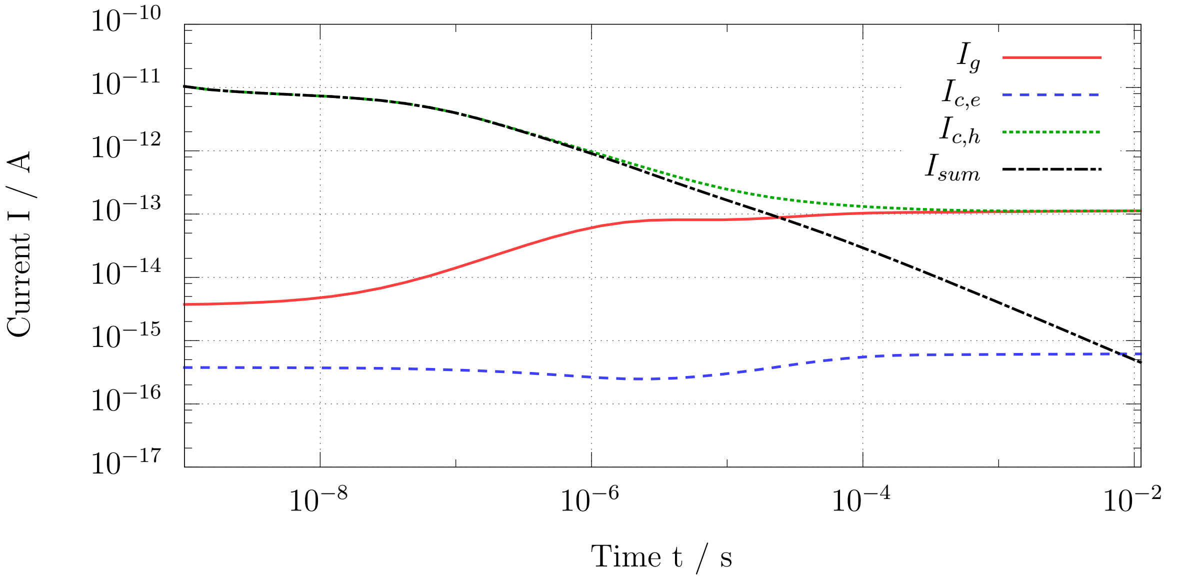


Figure 3.25: Currents during stress. A sum current I_{sum} results from charging of the traps. It is sourced by the hole-current from the channel $I_{c,h}$. The gate current I_g and the electron current from the channel $I_{c,e}$ play a minor role. Overlayed is a transport current from the channel to the gate.

In addition to the gate and channel currents, the sum current is shown. It is the current resulting from the change of the trap charges, sourced by the capture of holes from the channel. After about 10^{-5} s, a static transport current from the channel to the gate dominates.

Figure 3.26 shows the same currents during recovery, again for the fifth cycle.

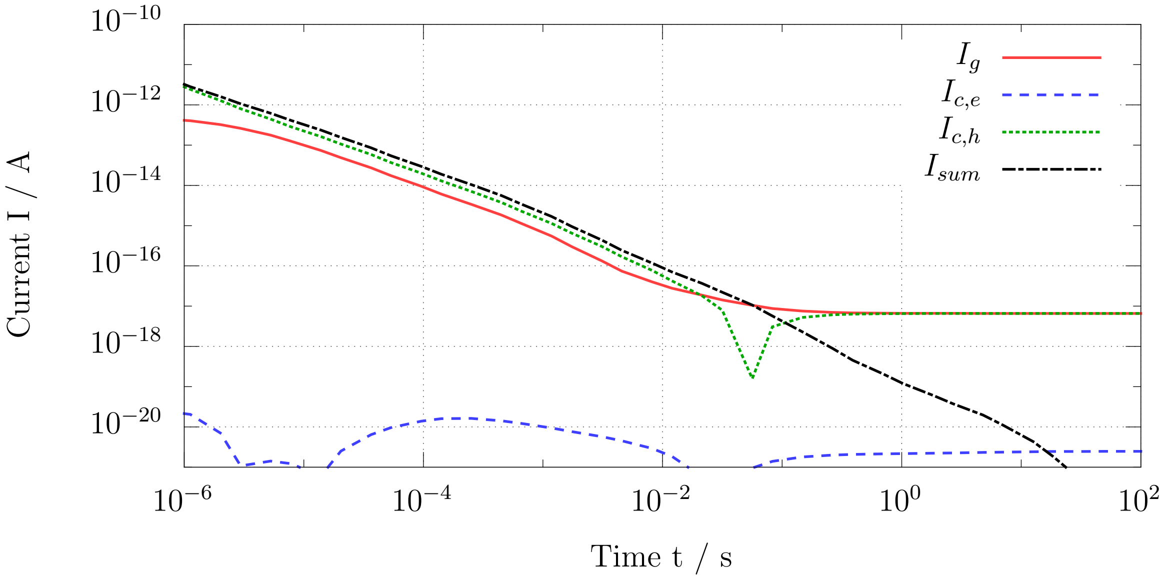


Figure 3.26: Currents during recovery. A sum current I_{sum} results from discharging of the traps. It is sourced by both the hole-current from the channel $I_{c,h}$ and the gate current I_g. The electron current from the channel $I_{c,e}$ plays a minor role. Overlayed is a transport current from the channel to the gate.

During recovery, the negative discharging current comes from both, channel and gate. After about 10^{-2} s, the small static transport current from the channel to the gate starts dominating and the current from the channel changes its sign to positive.

The stress and recovery currents for stress times ranging from 10^{-6} s to 10^{4} s can be found in figure 3.27 (stress) and figure 3.28 (recovery).

It can be seen that the decrease of the charging current is almost linear in the double logarithmic plot, except for the beginning of the stress cycle. The static transport currents do not change with stress time.

The recovery current plots show higher currents for longer stress times. The slope of the recovery current is lower for higher stress times which indicates that more traps with longer emission times were charged during stress.

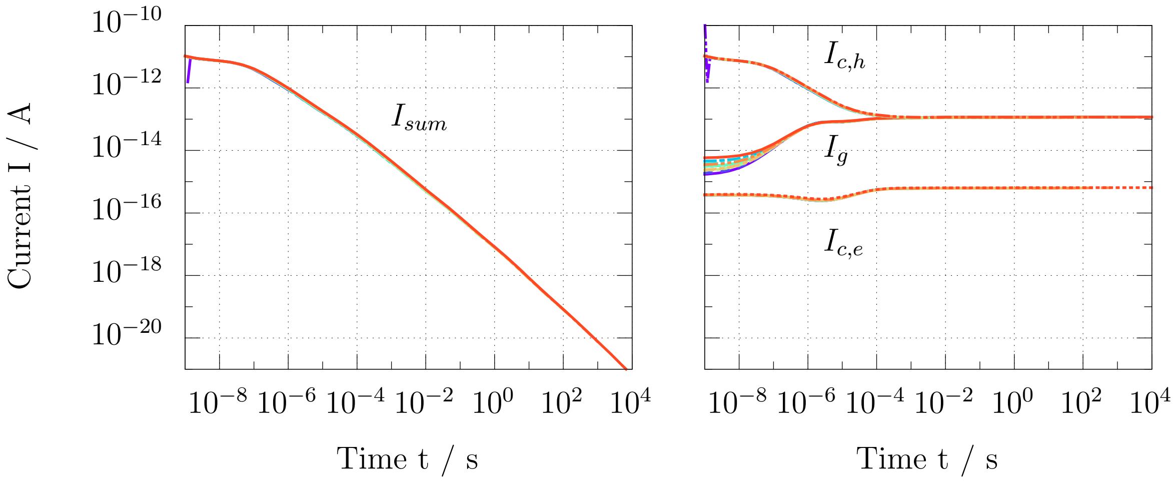


Figure 3.27: Currents during stress for stress times ranging from 1.1×10^{-6} s (blue) to 1.1×10^{4} s (red). The plots differ only slightly as the recovery phases were sufficiently long to discharge most of the traps.

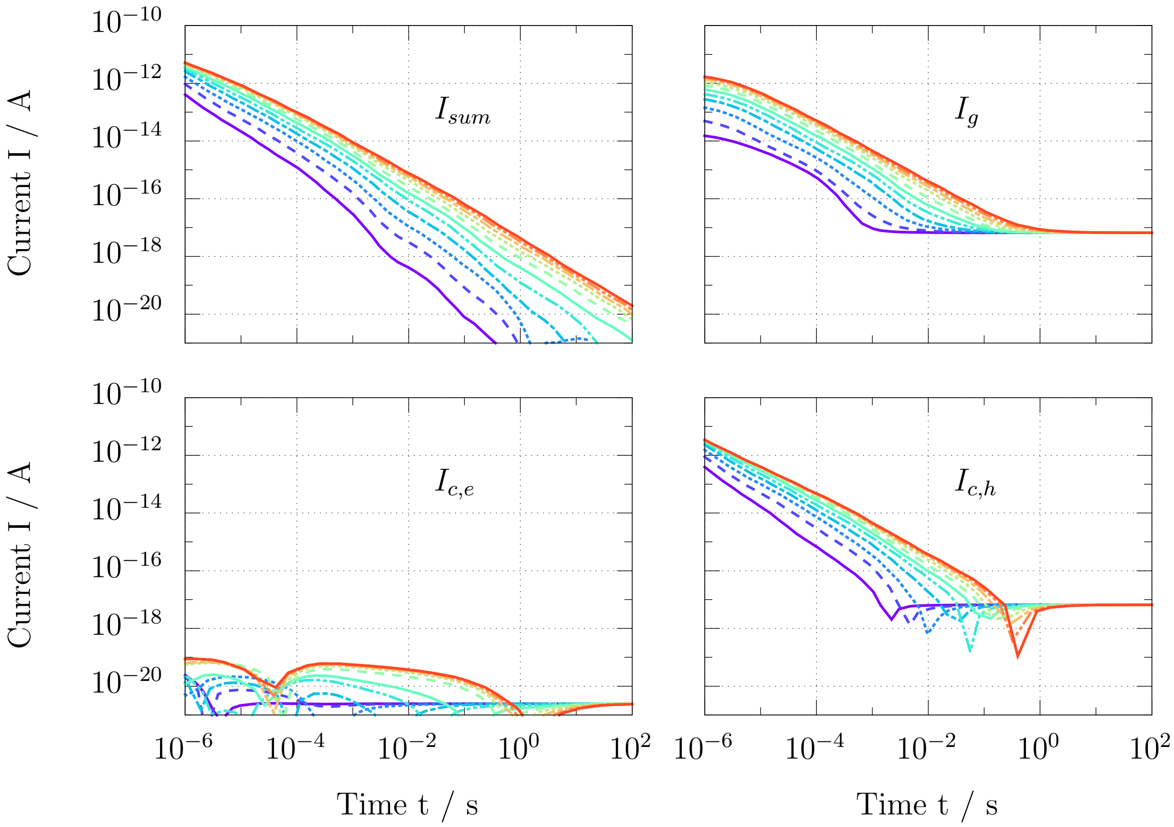


Figure 3.28: Currents during recovery for stress times ranging from 1.1×10^{-6} s (blue) to 1.1×10^{4} s (red). Notice the sign change in the channel current plots.

In figures 3.29, 3.30 and 3.31 the individual contributions of the traps to the charging, discharging and transport currents are shown in the band diagram of the device for the last stress and recovery cycle with a stress time of 10^4 seconds.

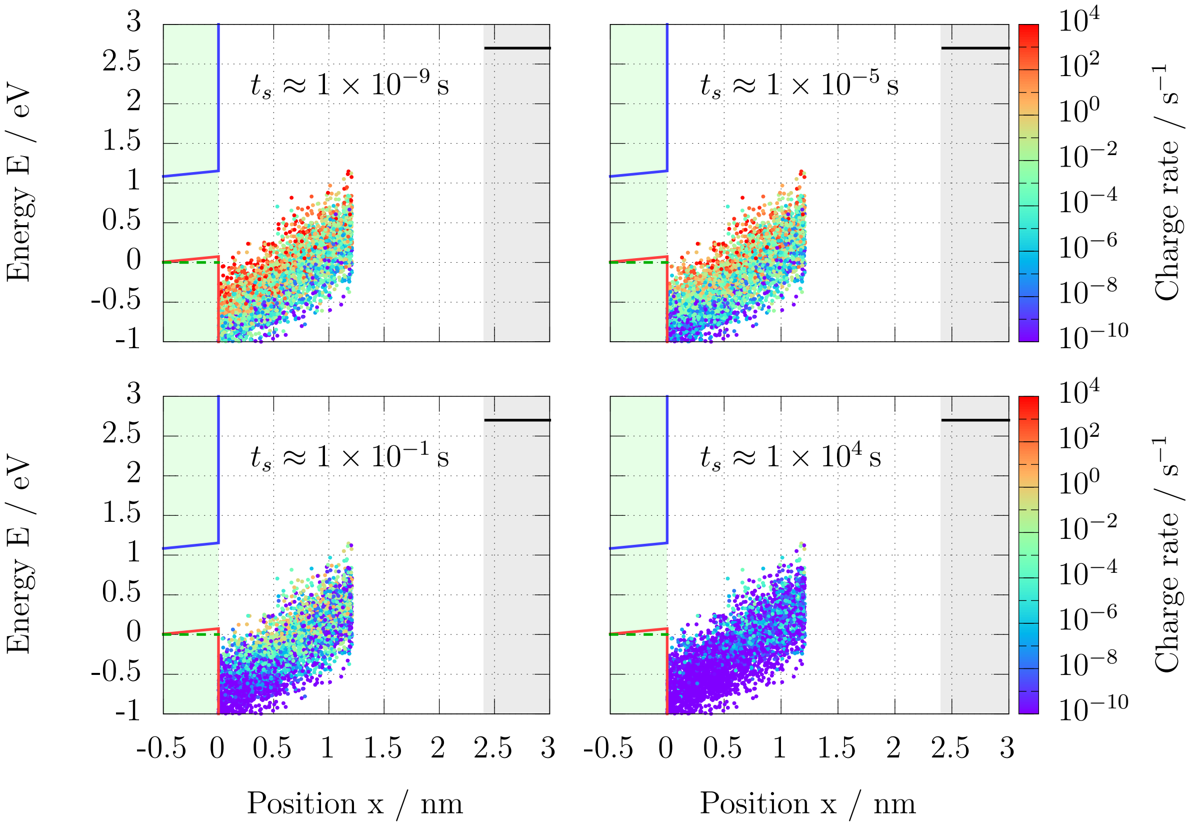


Figure 3.29: Band diagram showing the carrier transport to the individual traps during stress. The lightly green coloured area on the left is silicon, the white area the oxide and the gray area on the right the metal gate.

Figure 3.29, shows the charge currents during the stress cycle. Shortly after starting the stress cycle, traps higher in energy show the highest charging rates. With increasing stress time, the charging currents generally decrease and traps below the fermi level quickly become inactive as their equilibrium occupancy during stress differs only slightly from that during recovery.

After the switch to the recovery cycle, shown in figure 3.30, the traps that were above the fermi level during stress and low in energy now show the highest discharge current. The current decreases and its distribution shifts higher in energy.

At the end of the stress and recovery cycles, the static transport current from bulk to gate dominates the gate current, shown in figure 3.31. Both stress and recovery transport currents are mainly caused by traps close to the center of the oxide, both energetically and spatially.

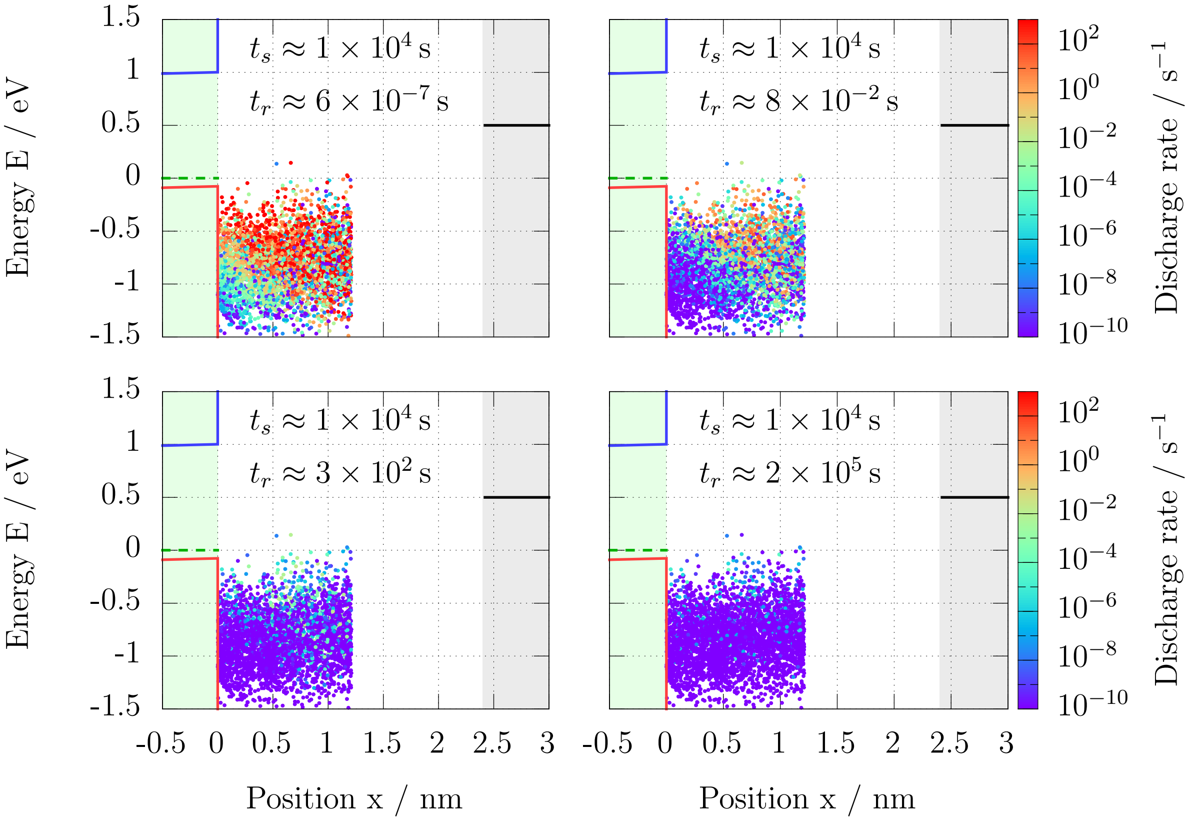


Figure 3.30: Band diagram showing the carrier transport from the individual traps during recovery. The lightly green coloured area on the left is silicon, the white area the oxide and the gray area on the right the metal gate.

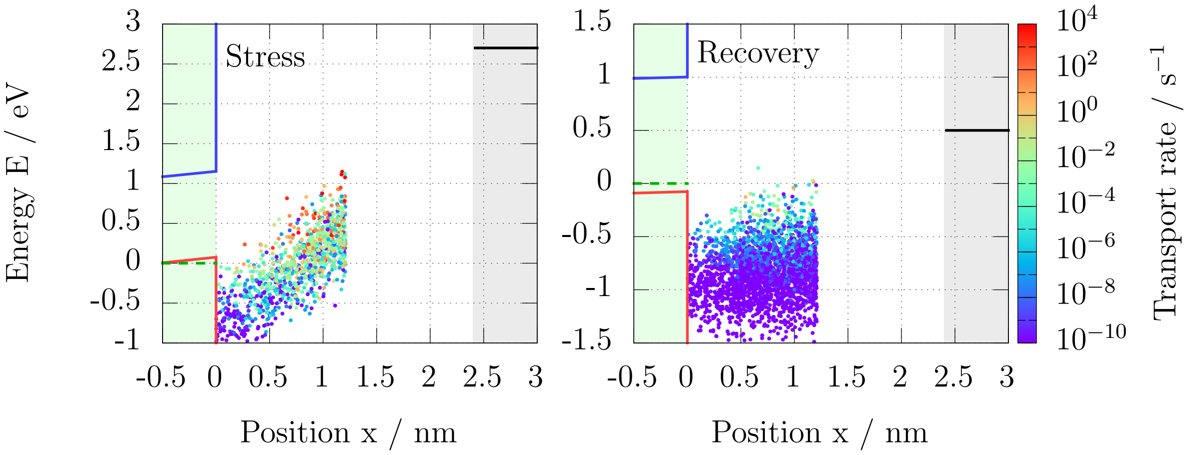


Figure 3.31: Band diagram showing the traps contributing to the transport current during stress and recovery. The lightly green coloured area on the left is silicon, the white area the oxide and the gray area on the right the metal gate.

Conclusion

The current resulting from the traps in the oxide generally consists of a transient charging current and a static transport current. Both the transient stress and recovery currents show approximately linear decrease when plotted in a double logarithmic plot.

The charging current is higher for traps higher in effective energy and decreases faster for traps below the channel Fermi level. The discharging current after stress comes mainly from traps in the active energy area, i.e. from the traps that were above or close to the channel Fermi level during stress.

Biggest contributors to the static transport current are traps located energetically between the channel and metal Fermi levels. The fraction of traps contributing to the transport current increases for positions closer to the center of the oxide. This is shown more clearly than in the earlier 2D plots in figure 3.32.

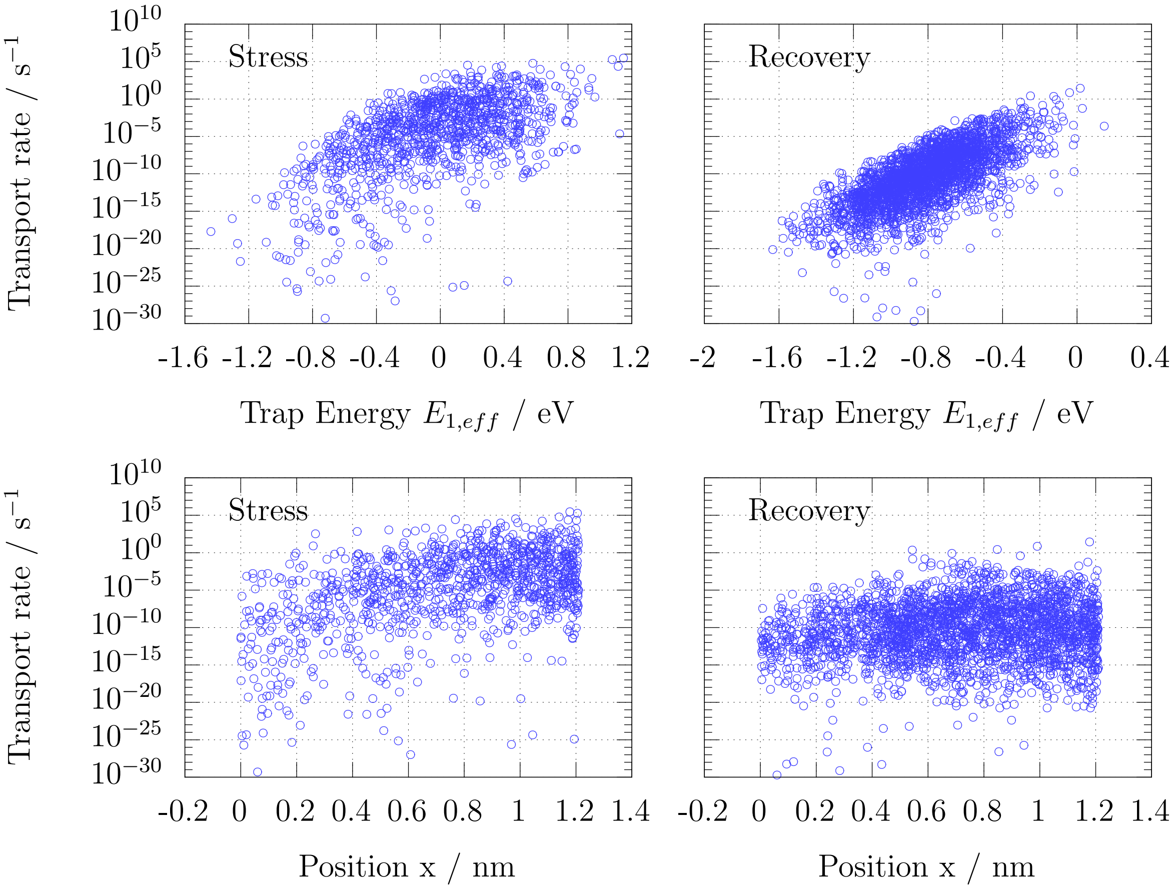


Figure 3.32: Energetic and spatial distribution of transport rates of individual traps. The transport rates of the traps decrease strongly below the channel Fermi level at 0 eV. Traps located closer to the center of the oxide are more likely to transport charge and higher transport rates are possible. It should be noted that the distribution of the traps in the device is homogenous in x and gaussian in E.

CHAPTER 4

Summary, Outlook

In the framework of the non-radiative multi phonon (NMP) model, rates describing the charge transfer between a trap in a semiconducting material and the conduction and valence bands of this material have been derived. This was done to supplement the rates to adjacent materials that already existed and allow simulation of traps in semiconductors. These rates were discussed to find their dependence on electric field and temperature. They were implemented in a device simulator together with equations for the currents to each exchange point to enable simulation of trap assisted tunneling (TAT).

Capture and emission times were simulated for a MIS device to discuss the importance of the additional charge transfer mechanism depending on the band gap and thickness of the semiconducting or semi insulating material used, and trap depth. Reverse leakage in a GaN/AlGaN HEMT was simulated and compared with the Fowler-Nordheim and Frenkel-Poole models often used to fit parts of the measured data. Depending on field and temperature, the simulations show behaviour comparable with both models depending on temperature and field and qualitative agreement with measurements in the literature. Finally, a MOSFET in a MSM setup has been simulated to evaluate transient charging and static transport currents during stress and recovery phases and the contribution of individual traps to these currents.

While it has been shown that the NMP model can be used to calculate TAT for individual traps, this is limited to short tunneling distances as direct charge transfer between traps is not yet accounted for. Thus hopping transport over longer distances is not possible at the moment. Implementation of trap-trap interaction is challenging as the number of interactions between traps increases quadratic with their number. Even worse, the necessary computation time for the solution of this equation system with its densely populated matrix would increase with n^3. A suitable metric will have to be found to limit the amount of traps each trap interacts with.

Bibliography

[1] Y. Miura and Y. Matukura, "Investigation of silicon-silicon dioxide interface using MOS structure," *Japanese Journal of Applied Physics*, no. 5, pp. 180–180, 1966.

[2] T. Grasser, ed., *Bias Temperature Instability for Devices and Circuits.* Springer, 2013.

[3] H. Reisinger, R.-P. Vollertsen, P.-J. Wagner, T. Huttner, A. Martin, S. Aresu, W. Gustin, T. Grasser, and C. Schlünder, "The effect of recovery on NBTI characterization of thick non-nitrided oxides," in *Integrated Reliability Workshop Final Report, 2008. IRW 2008. IEEE International*, pp. 1–6, IEEE, 2008.

[4] C. D. Young, Y. Zhao, M. Pendley, B. H. Lee, K. Matthews, J. H. Sim, R. Choi, G. A. Brown, R. W. Murto, and G. Bersuker, "Ultra-short pulse current–voltage characterization of the intrinsic characteristics of high-κ devices," *Japanese journal of applied physics*, vol. 44, no. 4S, p. 2437, 2005.

[5] B. Kaczer, T. Grasser, P. J. Roussel, J. Martin-Martinez, R. O'Connor, B. O'sullivan, and G. Groeseneken, "Ubiquitous relaxation in BTI stressing—new evaluation and insights," in *Reliability Physics Symposium, 2008. IRPS 2008. IEEE International*, pp. 20–27, IEEE, 2008.

[6] K. O. Jeppson and C. M. Svensson, "Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices," *Journal of Applied Physics*, vol. 48, no. 5, pp. 2004–2014, 1977.

[7] F. Schanovsky and T. Grasser, "On the microscopic limit of the modified reaction-diffusion model for the negative bias temperature instability," in *Reliability Physics Symposium (IRPS), 2012 IEEE International*, pp. XT–10, IEEE, 2012.

[8] M. A. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," *Microelectronics Reliability*, vol. 45, no. 1, pp. 71–81, 2005.

[9] T. L. Tewksbury III, *Relaxation Effects in MOS Devices due to Tunnel Exchange with Near-Interface Oxide Traps.* PhD thesis, Massachusetts Institute of Technology, 1992.

[10] V. Huard, C. Parthasarathy, N. Rallet, C. Guerin, M. Mammase, D. Barge, and C. Ouvrard, "New characterization and modeling approach for NBTI degradation from transistor to product level," in *2007 IEEE International Electron Devices Meeting*, 2007.

[11] A. McWhorter, "1/f noise and germanium surface properties," *Semiconductor surface physics*, pp. 207–228, 1957.

[12] M. Kirton and M. Uren, "Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/f) noise," *Advances in Physics*, vol. 38, no. 4, pp. 367–468, 1989.

[13] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel, "A two-stage model for negative bias temperature instability," in *Reliability Physics Symposium, 2009 IEEE International*, pp. 33–44, IEEE, 2009.

[14] W. Gös, *Hole trapping and the negative bias temperature instability.* PhD thesis, Technische Universität Wien, 2011.

[15] D. M. Sathaiya and S. Karmalkar, "Thermionic trap-assisted tunneling model and its application to leakage current in nitrided oxides and AlGaN/GaN high electron mobility transistors," *Journal of applied physics*, vol. 99, no. 9, p. 093701, 2006.

[16] S. Fleischer, P. Lai, and Y. Cheng, "Simplified closed-form trap-assisted tunneling model applied to nitrided oxide dielectric capacitors," *Journal of applied physics*, vol. 72, no. 12, pp. 5711–5715, 1992.

[17] M. P. Houng, Y. H. Wang, and W. J. Chang, "Current transport mechanism in trapped oxides: A generalized trap-assisted tunneling model," *Journal of applied physics*, vol. 86, no. 3, pp. 1488–1491, 1999.

[18] J. Frenkel, "On pre-breakdown phenomena in insulators and electronic semiconductors," *Physical Review*, vol. 54, no. 8, p. 647, 1938.

[19] R. H. Fowler and L. Nordheim, "Electron emission in intense electric fields," *Proceedings of the Royal Society of London*, 1928.

[20] E. Miller, X. Dang, and E. Yu, "Gate leakage current mechanisms in AlGaN/GaN heterostructure field-effect transistors," *Journal of applied Physics*, vol. 88, no. 10, pp. 5951–5958, 2000.

[21] S. Karmalkar, N. Satyan, and D. M. Sathaiya, "On the resolution of the mechanism for reverse gate leakage in AlGaN/GaN HEMTs," *Electron Device Letters, IEEE*, vol. 27, no. 2, pp. 87–89, 2006.

[22] B. Ricco, G. Gozzi, and M. Lanzoni, "Modeling and simulation of stress-induced leakage current in ultrathin SiO_2 films," *Electron Devices, IEEE Transactions on*, vol. 45, no. 7, pp. 1554–1560, 1998.

[23] L. Larcher, A. Paccagnella, and G. Ghidini, "A model of the stress induced leakage current in gate oxides," *Electron Devices, IEEE Transactions on*, vol. 48, no. 2, pp. 285–288, 2001.

[24] J. R. Norris, "Continuous-time Markov chains I," in *Markov Chains*, pp. 60–107, Cambridge University Press, 1997.

[25] T. Grasser, "Stochastic charge trapping in oxides: From random telegraph noise to bias temperature instabilities," *Microelectronics Reliability*, vol. 52, p. 39–70, Jan 2012.

[26] G. Rzepa, "Microscopic modeling of NBTI in MOS transistors," Master's thesis, Technische Universität Wien, 2013.

[27] F. Schanovsky, O. Baumgartner, V. Sverdlov, and T. Grasser, "A multi scale modeling approach to non-radiative multi phonon transitions at oxide defects in MOS structures," *Journal of Computational Electronics*, vol. 11, p. 218–224, May 2012.

[28] K. Huang and A. Rhys, "Theory of light absorption and non-radiative transitions in F-centres," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 204, pp. 406–423, Dec 1950.

[29] C. H. Henry and D. V. Lang, "Nonradiative capture and recombination by multiphonon emission in GaAs and GaP," *PHYSICAL REVIEW B*, vol. 15, no. 2, 1975.

[30] S. D. Ganichev, I. N. Yassievich, and W. Prettl, "Tunnelling ionization of deep centres in high-frequency electric fields," *Journal of Physics: Condensed Matter*, vol. 14, no. 50, p. R1263, 2002.

[31] G. Hurkx, F. O'Hara, and M. Knuvers, "Modelling forward-biased tunneling," in *Solid State Device Research Conference, 1989. ESSDERC '89. 19th European*, pp. 793–796, Sept 1989.

[32] K. J. Laidler, "The development of the Arrhenius equation," *Journal of Chemical Education*, vol. 61, no. 6, pp. 494–498, 1984.

[33] H. Zhang, E. J. Miller, and E. T. Yu, "Analysis of leakage current mechanisms in Schottky contacts to GaN and Al_0.25ga_0.75n/gan grown by molecular-beam epitaxy," *Journal of Applied Physics*, vol. 99, no. 2, p. 023703, 2006.

[34] S. Ganguly, A. Konar, Z. Hu, H. Xing, and D. Jena, "Polarization effects on gate leakage in InAlN/AlN/GaN high-electron-mobility transistors," *Applied Physics Letters*, vol. 101, no. 25, p. 253519, 2012.