Technische Universität Wien

Dissertation

Modeling and Simulation of Negative Bias Temperature Instability

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften

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

Robert Entner

9725877 / E754

[redacted]

E–Mail: robert@entner.net

Wien, im April 2007

To Anna, Sarah, and Johanna.

Abstract

SEMICONDUCTOR process and device simulators are well established tools for the reduction of the development time for semiconductor devices. Numerical simulation can help the device engineer to perform dozens of design experiments within the time normally needed to process one single wafer lot. Nowadays simulation efforts go beyond solving the basic semiconductor device equations. Especially the modeling and simulation of aging processes has tremendously gained in importance. This thesis tries to give insight into the topic of semiconductor device simulation and focuses on the modeling of degradation mechanisms.

After a first introduction, giving an overview of the motivation and the structure of this work, the fundamental equations for numerical device simulation, the modeling of generation and recombination mechanisms, as well as the modeling of quantum confinement effects are presented.

The possible defects at the Si/SiO_2 interface are evaluated in the following chapter as this interface is of utmost importance in CMOS technology. An important topic is also the experimental characterization methods of the interface quality.

In the following a thorough investigation on degradation mechanisms affecting the gate dielectric is presented. These are the hot carrier degradation, dielectric wearout and breakdown, and the effect of quantum mechanical tunneling. Here, a new model is proposed for the numerical modeling of trap assisted tunneling currents across several electron traps.

The main part of this work concentrates on negative bias temperature instability (NBTI). NBTI causes degradation of MOS structures at elevated temperatures and negative gate voltages. An elaborate investigation of literature from the first report to the recent understanding of this degradation mechanism is presented. A comprehensive model is derived, combining research results from different groups and the coupling to the basic semiconductor device equations.

The new NBTI model is compared to measurement data and gives excellent results. Furthermore the susceptibility of CMOS circuits as the CMOS inverter, an SRAM cell, and a ring oscillator to NBTI is investigated in mixed mode device and circuit simulations. Stationary and transient simulations show the impact of degradation on real life circuits.

Kurzfassung

HALBLEITER Prozess- und Bauelementsimulatoren sind etablierte Werkzeuge um die Entwicklungszeiten von Halbleiterprozessen so kurz wie möglich zu halten. Unter Zuhilfenahme numerischer Simulation kann der Bauelementingenieur dutzende Experimente in der Zeit durchführen, in der auf herkömmliche Weise nur die Herstellung eines einzigen Siliziumloses möglich ist. Heutzutage werden nicht mehr nur die grundlegenden Halbleitergleichungen gelöst. Vor allem die Modellierung und Simulation von Alterungsprozessen hat stark an Bedeutung gewonnen. Diese Arbeit gibt Einblick in das Thema Halbleitersimulation, wobei der Schwerpunkt auf der Modellierung dieser Degradationsmechanismen liegt.

Nach einer Einleitung, welche Überblick über Motivation und Struktur dieser Arbeit vermittelt, werden die grundlegenden Gleichungen zur numerischen Bauelementsimulation, die Modellierung von Generations- und Rekombinationsmechanismen sowie der räumlichen Quantisierung präsentiert.

Nachfolgend werden mögliche Defekte an der Si/SiO_2 Grenzfläche beschrieben, da die Qualität dieses Überganges von höchster Bedeutung für die CMOS Technologie ist. Ebenso wird auf die experimentelle Charakterisierung eingegangen.

Im Weiteren folgt eine genaue Untersuchung von Alterungsprozessen welche das Gate Dielektrikum betreffen. Diese sind Hot Carrier Degradation, Dielectric Wearout und Dielectric Breakdown sowie das quantenmechanische Tunneln. Es wird ein neues Modell vorgeschlagen, welches zur numerischen Betrachtung von störstellenunterstütztem Tunneln über mehrere Elektronenfallen dient.

Der Hauptteil dieser Arbeit konzentriert sich auf den Alterungseffekt der Negative Bias Temperature Instability (NBTI). Er führt zur Degradation von MOS Strukturen bei erhöhten Temperaturen und negativen Gate Spannungen. Es werden wissenschaftliche Arbeiten zum Thema NBTI vom ersten Bericht bis zum heutigen Wissensstand präsentiert. Ein umfassendes Modell wird vorgeschlagen, welches Forschungsergebnisse verschiedener Gruppen zusammenfasst und das Degradationsmodell mit den Halbleitergleichungen kombiniert.

Das neue Modell wird mit Messdaten verglichen und liefert exzellente Übereinstimmung. Weiters wird die Anfälligkeit von CMOS Schaltungen, wie dem CMOS Inverter, einer SRAM Zelle und eines Ringoszillators, in kombinierten Bauelement- und Schaltungssimulationen untersucht. Stationäre sowie transiente Simulationen zeigen den Einfluss von NBTI auf diese Schaltungen.

Acknowledgment

I WOULD like to express my deep gratitude to a number of people who supported me during my work in technical, financial, and personal form.

First of all I want to thank Prof. Tibor Grasser for giving me the opportunity to work in the Christian Doppler Laboratory for TCAD in Microelectronics at the Institute for Microelectronics. He was the supervisor of this thesis and without his scientific support the high quality of this work would not have been possible. The financial support by the Christian Doppler Gesellschaft allowed me to fully concentrate on the scientific work in combination with an industrial background, thanks to our industry partners Austriamicrosystems and Infineon. A big thank you goes to Prof. Magerl who agreed to take part in the examination committee. I would like to thank Gilbert Declerck, Guido Groeseneken, Robin Degraeve, and Ben Kaczer who gave me the opportunity to work at IMEC research center in Leuven, Belgium, during a two month internship.

I want to thank Prof. Siegfried Selberherr and Prof. Erasmus Langer who are responsible for the excellent working conditions at the Institute for Microelectronics. Both, the technical infrastructure and the amicable climate at the institute assisted me with my task. Also the support of Manfred Katterbauer and Ewald Haslinger in administrative and technical aspects is highly appreciated.

Of highest importance during my time at the Institute for Microelectronics was the endless support of my colleagues and friends. First and foremost I want to name Enzo Ungersböck. We shared our office and had literally uncountable discussions about computer problems, our fatherhood, where to go for lunch, or just daily business. In him I have found a very good friend. The proof-readers of my thesis, Oliver Triebl, Wolfgang Gös, and Philipp Hehenberger with their numerous suggestions greatly improved this work. Furthermore I want to thank Alireza Sheikholeslami, Andreas Gehring, Andreas Hössinger, Christian Hollauer, Gerhard Karlowatz, Gregor Meller, Johann Cervenka, Markus Karner, Martin Vasicek, Martin Wagner, Paul Wagner, Philipp Schwaha, Robert Wittmann, Siddhartha Dhar, Stefan Holzer, Stephan Wagner, Viktor Sverdlov, and Wilfried Wessner, who let me share their knowledge on semiconductors, programming, or operating system related topics.

Above all I want to thank my family. My loving wife Anna supported and encouraged me whenever possible. She brought me through the hard times when work seemed to find no end. My parents, who gave me the opportunity to attend university and who took the financial burden, especially my mother, who raised and guided me. My mother-in-law was supporting my wife and me with our children whenever possible.

And, of course, there are Sarah and Johanna. My daughters are most certainly the best achievement of my life ...

Contents

List of Abbreviations and Acronyms

BJT	...	Bipolar junction transistor
BTI	...	Bias temperature instability
CETD	...	Critical electron trap density
CHE	...	Channel hot electron
CMOS	...	Complementary MOS
CV	...	Capacitance-voltage
DAHC	...	Drain avalanche hot carrier
DCIV	...	Direct current current voltage
DOS	...	Density of states
DRAM	...	Dynamic random access memory
DUT	...	Device under test
ESD	...	Electrostatic discharge
ESR	...	Electron spin resonance
GPV	...	Gate pulsed voltage
HBM	...	Human body model
IV	...	Current-voltage
MOS	...	Metal-oxide-semiconductor
MOSFET	...	MOS field-effect transistor
NBT	...	Negative bias temperature
NBTI	...	Negative bias temperature instability
nMOSFET	...	n-channel MOSFET
pMOSFET	...	p-channel MOSFET
SDR	...	Spin dependent recombination
SGHE	...	Secondary generated hot electron
SHE	...	Substrate hot electron
SILC	...	Stress-induced leakage current
SNM	...	Static noise margin
SRAM	...	Static random access memory
SRH	...	Shockley-Read-Hall
TAT	...	Trap assisted tunneling
TCAD	...	Technology computer-aided design
TDDB	...	Time dependent dielectric breakdown
VTC	...	Voltage transfer characteristic
VLSI	...	Very large scale integration

Chapter 1

Introduction

THE invention of the transistor in 1947 has started the exponential growth of an industry which is now, some decades later, a several hundred billion dollar industry. The first bipolar transistor was announced in December 1947 by William Shockley, John Bardeen, and Walter Brattain at Bell Labs. In 1956 the three researchers were honored with the Nobel Price in physics for their invention. The first metal-oxide-semiconductor (MOS) transistor and the first integrated circuits were demonstrated in the early 1960s. From that time on the development in the field of microelectronics was impressive. The integration density grew exponentially. This exponential growth was already identified in 1965 by Gordon Moore [1] as follows

> *"The complexity for minimum component costs has increased at a rate of roughly a factor of two per year (...) Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years. That means by 1975, the number of components per integrated circuit for minimum cost will be 65,000. I believe that such a large circuit can be built on a single wafer."*

Today, in 2007, we have more than 1 billion transistors on one single processor die. This high integration density has to be accompanied by stringent efforts to increase the reliability of each transistor, of the interconnect structure, the packaging, and the whole die to a maximum possible. The failure of a single transistor can lead to complete failure of the whole system.

Not only the integration density has been steadily growing, also the pressure on the industry to deliver short time-to-marked and therefore minimal research and development times for new technology nodes is increasing. This has led to intense efforts in the field of numerical simulation of the semiconductor manufacturing process and the resulting device structure, so called technology computer aided design (TCAD). It can reduce the number of test cycles with real semiconductor devices and drastically increase the possibilities to vary process parameters as doping concentrations, device geometries, materials and their composition to a minimum. Here, TCAD gives the opportunity to analyze the effect of process variation within hours instead of weeks for real processing.

The main target of this thesis is to extend the capabilities of the TCAD framework by a very important degradation mechanism, the negative bias temperature instability (NBTI). This degradation mechanism especially affects p-channel MOS field effect transistors (MOSFETs) which are stressed with negative gate voltages at elevated temperatures. NBTI has gained tremendous scientific and industrial interest as it can lead to severe shifts of important transistor parameters, as the threshold voltage or the drain current, and seems to be accelerated in recent technologies which rely on nitrided oxides, high-k dielectrics, and other novel approaches. The exact physical background is still not completely understood but the general consensus is that the Si/SiO_2 interface is damaged and interface traps and/or interface charges and probably oxide charges cause the degradation.

The structure of this thesis is designed to guide the reader from the fundamental basics of semiconductor device simulation, over the most important degradation mechanisms leading to the thorough elaboration of the effect of NBTI. Presenting the state-of-the-art scientific understanding of this form of device degradation, a comprehensive model is proposed and evaluated by the simulation of a range of case studies.

Chapter 2 presents the fundamental equations for semiconductor device simulation and their derivation from Maxwell's equations and also analytic MOSFET approximations. Next, the most important generation and recombination mechanisms found in semiconductor devices are presented. Here, the main focus is laid upon the phonon transition, or Shockley-Read-Hall, mechanism. It is of special interest for modeling the carrier generation and recombination at silicon/dielectric interface traps which are caused by negative bias temperature instability. The third part of this chapter gives an insight into modeling of quantum confinement at the silicon/dielectric interface.

Chapter 3 focuses on the silicon/silicon-dioxide interface. It is of utmost importance for the proper operation of CMOS technology. NBTI affects exactly this interface and generates defects which are discussed in this chapter. Chapter 4 shows possible means of characterizing the interface using the charge pumping method, the DCIV method, and capacitance-voltage measurements.

Chapter 5 gives an overview of common degradation mechanisms affecting the dielectric layer in CMOS transistors. Sources of degradation outlined here are the damage caused by hot carriers, the dielectric wear-out and breakdown, and quantum mechanical tunneling. The problem of trap-assisted tunneling through highly degraded devices is described in detail. It causes not only additional gate leakage currents but also threshold voltage shifts due to trapped carriers in the oxide, comparable to NBTI.

The main part of this thesis is found in Chapter 6. Here the history of negative bias temperature instability is presented along with the current understanding of the physical mechanisms involved. Methods for measuring the level of degradation are examined and common models trying to reflect the physics of NBTI are presented and discussed. The last part of this chapter presents the new model which was implemented in a numerical device simulator as part of this work. This new model can achieve excellent agreement with measurement data and is especially valuable for the evaluation of pure SiO_2 dielectrics as used in power MOSFETs.

Chapter 7 gives case studies of the presented model which is compared to measurement data and achieves excellent agreement. The impact of NBTI on circuit performance is evaluated in transient and stationary numerical device simulations.

Chapter 2

Simulation of Semiconductor Devices

OPTIMIZATION of a semiconductor manufacturing process can be a cumbersome task. It is often based on trial and error steps where different processing parameters such as the exact doping profile, geometries, and temperatures are evaluated. Since the possible combinations are literally unlimited, it is very important to have highly skilled device engineers working on this task. But still, as every new experiment takes up to weeks until the results are obtained and the expenses for fabrication are very high, alternatives or at least assistance in device development is of crucial importance. This is where semiconductor device simulation comes into play. Not only can different geometries and doping profiles be analyzed within hours, simulation also gives detailed insight into the device behavior.

Measurement data mainly concentrate on electrical characterization of the extrinsic data, semiconductor devices can deliver. These can be the voltage-current relations, the capacitance-voltage curves, or similar. Using simulation tools, also distributed quantities inside the semiconductor can be explored. This can be for example the carrier concentration, the electrostatic potential, the carrier temperature, or the current density. By using these software tools, the device engineer can gain additional information about how processing changes can alter the intrinsic properties of the semiconductor device.

2.1 Classical Semiconductor Device Equations

The semiconductor device equations can be used to describe the whole simulation domain of a semiconductor device. They are applied to the bulk semiconductor, the highly doped regions such as source and drain, and to dielectric regions such as the gate dielectric. In this section the classical semiconductor device equations are presented which are widely used for device simulation and their derivation.

2.1.1 Maxwell's Equations

The equations developed by James Clerk Maxwell describe the behavior of electric and magnetic fields and their interaction with matter. They were published by Maxwell in 1864 [2] and in its original form comprised of 20 equations in 20 unknowns. Later they were reformulated in vector notation to the following form

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}\,, \tag{2.1}$$

$$\vec{\nabla} \cdot \vec{B} = 0\,, \tag{2.2}$$

$$\vec{\nabla} \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}\,, \tag{2.3}$$

$$\vec{\nabla} \cdot \vec{D} = \rho\,. \tag{2.4}$$

Here, $\vec{E}$ is the electric field, $\vec{H}$ the magnetic field, $\vec{D}$ the displacement vector, and $\vec{B}$ the magnetic flux density vector. $\vec{J}$ denotes the conduction current density, ρ is the electric charge density, and $\partial/\partial t$ is the partial derivative with respect to time.

Equation (2.1) expresses the generation of an electric field due to a changing magnetic field (Faraday's law of induction), (2.2) predicts the absence of magnetic monopoles (magnetic sources or sinks), (2.3) reflects how an electric current and the change in the electric field produce a magnetic field (Ampere-Maxwell law), and finally (2.4) correlates the creation of an electric field due to the presence of electric charges (Gauss' law).

We are using the Maxwell's equations to derive parts of the semiconductor device equations, namely the Poisson equation and the continuity equations.

2.1.2 Poisson's Equation

Poisson's equation correlates the electrostatic potential Φ to a given charge distribution ρ. It can be derived from (2.4) using the relation between the electric displacement vector and the electric field vector,

$$\vec{D} = \hat{\varepsilon} \cdot \vec{E}\,, \tag{2.5}$$

where $\hat{\varepsilon}$ is the permittivity tensor. This relation is valid for materials with time independent permittivity. As materials used in semiconductor devices normally do not show significant anisotropy of the permittivity, $\hat{\varepsilon}$ can be considered as a scalar quantity ε in device simulation. The total permittivity is obtained from the relative ε_r and the vacuum permittivity ε_0 as $\varepsilon = \varepsilon_\mathrm{r}\varepsilon_0$. Table 2.1 gives an overview of relative permittivity constants for some materials commonly used in semiconductor devices.

As $\vec{\nabla} \times \vec{E} = 0$ for the stationary case $\vec{E}$ can be expressed as a gradient field of a scalar potential field

$$\vec{E} = -\vec{\nabla}\Phi\,. \tag{2.6}$$

Substituting (2.5) and (2.6) in (2.4) we get

$$\vec{\nabla} \cdot \hat{\varepsilon} \cdot \vec{\nabla}\Phi = -\rho\,. \tag{2.7}$$

Material	Relative Permittivity
Si	11.7
GaAs	12.5
Ge	16.1
SiO_2	3.9
HfO_2	~ 25
$HfSiO_4$	~ 15–18
ZrO_2	~ 20–25

Table 2.1: Relative permittivity constants for materials used in or considered for semiconductor devices.

As we consider the permittivity a scalar, which is constant in homogeneous materials, we obtain Poisson's equation

$$\vec{\nabla} \cdot \vec{\nabla}\Phi = -\frac{\rho}{\varepsilon}\,. \tag{2.8}$$

The space charge density ρ consists of

$$\rho = \mathrm{q}(p - n + C)\,, \tag{2.9}$$

where q is the elementary charge, p and n the hole and electron concentrations, respectively, and C the concentration of additional, typically fixed, charges. These fixed charges can originate from charged impurities of donor (N_D) and acceptor (N_A) type and from trapped holes (ρ_p) and electrons (ρ_n),

$$C = N_\mathrm{D} - N_\mathrm{A} + \rho_p - \rho_n\,. \tag{2.10}$$

The inclusion of trapped carriers is important for the simulation of the impact of degradation on the device performance (Section 2.2, Section 5.3.2, Chapter 6).

Together (2.8) and (2.9) lead to the form of Poisson's equation commonly used for semiconductor device simulation

$$\vec{\nabla} \cdot \vec{\nabla}\Phi = -\frac{\mathrm{q}}{\varepsilon}(p - n + N_\mathrm{D} - N_\mathrm{A} + \rho_p - \rho_n)\,. \tag{2.11}$$

2.1.3 Continuity Equations

The continuity equations can be derived from (2.3) by applying the divergence operator, $\vec{\nabla}\cdot$, to the equation and considering that the divergence of the curl of any vector field equals zero

$$\vec{\nabla} \cdot \vec{\nabla} \times \vec{H} = \vec{\nabla} \cdot \vec{J} + \vec{\nabla} \cdot \frac{\partial \vec{D}}{\partial t} = 0\,. \tag{2.12}$$

Separating the total current density $\vec{J}$ into hole and electron current densities, $\vec{J} = \vec{J}_p + \vec{J}_n$, and using (2.4) and (2.9) gives

$$\vec{\nabla} \cdot \vec{J}_n + \vec{\nabla} \cdot \vec{J}_p + \mathrm{q}\left(\frac{\partial p}{\partial t} - \frac{\partial n}{\partial t} + \frac{\partial N_\mathrm{D}}{\partial t} - \frac{\partial N_\mathrm{A}}{\partial t} + \frac{\partial \rho_p}{\partial t} - \frac{\partial \rho_n}{\partial t}\right) = 0\,. \tag{2.13}$$

When we consider the charged impurities as time invariant and introduce a quantity R to split up (2.13) into separate equations for electrons and holes, we get

$$\vec{\nabla} \cdot \vec{J}_n - \mathrm{q}\frac{\partial n}{\partial t} - \mathrm{q}\frac{\partial \rho_n}{\partial t} = \mathrm{q}R\,, \tag{2.14}$$

$$\vec{\nabla} \cdot \vec{J}_p + \mathrm{q}\frac{\partial p}{\partial t} + \mathrm{q}\frac{\partial \rho_p}{\partial t} = -\mathrm{q}R\,. \tag{2.15}$$

The quantity R gives the net recombination rate for electrons and holes. A positive value means recombination, a negative value means generation of carriers. Models for R are presented in Section 2.3.

2.1.4 Current Relations

There are two major effects which lead to current flow in silicon. First, the drift of charged carriers due to the influence of an electric field, and second, the diffusion current due to a concentration gradient of the carriers.

2.1.4.1 Carrier Drift

Charged carriers in a semiconductor subjected to an electric field are accelerated and acquire a certain drift velocity. The orientation depends on the charge state, holes are accelerated in direction of the electric field and electrons in opposite direction. The magnitude of the drift velocity depends on the statistical probability of scattering events. At low impurity concentration, the carriers mainly collide with the crystal lattice. Is the impurity concentration higher the collision probability with the charged dopants through Coulomb interaction becomes more and more likely, thus reducing the drift velocity with increasing doping concentration.

For low electric fields, the drift component of the electric current can be expressed in terms of Ohm's law as

$$\vec{J}_n^{\mathrm{drift}} = \sigma_n \vec{E}\,, \tag{2.16}$$

$$\vec{J}_p^{\mathrm{drift}} = \sigma_p \vec{E}\,. \tag{2.17}$$

Here, σ denotes the conductivity and can be expressed in terms of carrier mobilities for electrons and holes, μ_n and μ_p, as

$$\sigma_n = \mathrm{q}n\mu_n\,, \tag{2.18}$$

$$\sigma_p = \mathrm{q}p\mu_p\,. \tag{2.19}$$

The mobility for electrons is, due to their lower effective mass, about three times higher than for holes.

2.1.4.2 Carrier Diffusion

A concentration gradient of carriers leads to carrier diffusion. This is because of their random thermal motion which is more probable in the direction of the lower concentration. The current

contribution due to the concentration gradient is written as

$$\vec{J}_n^{\text{diff}} = \mathrm{q} D_n \vec{\nabla} n \,, \tag{2.20}$$

$$\vec{J}_p^{\text{diff}} = -\mathrm{q} D_p \vec{\nabla} p \,. \tag{2.21}$$

Here, D_n and D_p are the diffusion coefficients for electrons and holes, which can, in thermal equilibrium for non-degenerate semiconductors, be expressed in terms of the mobility using the Einstein relation

$$D_n = \frac{\mathrm{k_B} T}{\mathrm{q}} \mu_n \,, \tag{2.22}$$

$$D_p = \frac{\mathrm{k_B} T}{\mathrm{q}} \mu_p \,. \tag{2.23}$$

2.1.4.3 Drift-Diffusion Current Relations

Combining the current contributions of the drift and the diffusion effect we get

$$\vec{J}_n = \mathrm{q} n \mu_n \vec{E} + \mathrm{q} D_n \vec{\nabla} n \,, \tag{2.24}$$

$$\vec{J}_p = \mathrm{q} p \mu_p \vec{E} - \mathrm{q} D_p \vec{\nabla} p \,. \tag{2.25}$$

2.1.5 The Semiconductor Equations

With the Poisson equation (2.11), the continuity equations for electrons and holes (2.14) (2.15), and the drift-diffusion current relations for electron- and hole-current (2.24) (2.25) we now have a complete set of equations which can be seen as fundamental for the simulation of semiconductor devices:

$$\vec{\nabla} \cdot \vec{\nabla} \Phi = -\frac{\mathrm{q}}{\varepsilon} (p - n + C) \,, \tag{2.26}$$

$$\vec{\nabla} \cdot \vec{J}_n - \mathrm{q} \frac{\partial n}{\partial t} - \mathrm{q} \frac{\partial \rho_n}{\partial t} = \mathrm{q} R \,, \tag{2.27}$$

$$\vec{\nabla} \cdot \vec{J}_p + \mathrm{q} \frac{\partial p}{\partial t} + \mathrm{q} \frac{\partial \rho_p}{\partial t} = -\mathrm{q} R \,, \tag{2.28}$$

$$\vec{J}_n = \mathrm{q} n \mu_n \vec{E} + \mathrm{q} D_n \vec{\nabla} n \,, \tag{2.29}$$

$$\vec{J}_p = \mathrm{q} p \mu_p \vec{E} - \mathrm{q} D_p \vec{\nabla} p \,. \tag{2.30}$$

These equations, not including the charge contribution of trapped carriers, have first been published by VanRoosbroeck in 1950 [3].

This set of equations is widely used in numerical device simulators and provides only the basics for device simulation. In modern simulators they are accompanied by higher order current relation equations like the hydrodynamic, six-, or eight-moments models. There are models for the carrier mobility, the carrier generation and recombination (Section 2.3), for quantum effects like quantum mechanical tunneling (Section 5.3) or quantum confinement (Section 2.4.1) and of course for modeling of device degradation, as negative bias temperature instability (Chapter 6).

2.2 Analytic MOSFET Approximations

For the simulation of electronic circuits it is often of interest to have compact models for the involved devices. In comparison to solving the semiconductor device equations using compact models reduces the simulation time drastically. Compact models can also be a good guide on what effects the change of physical quantities has on the device characteristics.

The disadvantage is that for each device the right model has to be chosen and the according model parameters have to be extracted. Furthermore, it is not possible to investigate novel device geometries or new materials without processing the device and obtaining an adequate model first.

2.2.1 Interface and Oxide Charges

A change in the interface and oxide charges (ΔQ_{it} and ΔQ_{ox}) contributes to a threshold voltage shift ΔV_{th} as

$$\Delta V_{\mathrm{th}} = \frac{\Delta Q_{\mathrm{it}} + \Delta Q_{\mathrm{ox}}}{C'}\,, \tag{2.31}$$

where C' is the capacitance per unit area of the oxide.

In the context of NBTI interface charges are usually the result of charged interface defects D_{it}. The interface charge depends on the Fermi-level E_{F} and the trap occupancy f and can be calculated as [4]

$$Q_{\mathrm{it}} = \mathrm{q}_0 \int D_{\mathrm{it}}(E) f(E_{\mathrm{F}}, E) \mathrm{d}E\,. \tag{2.32}$$

Charged oxide traps D_{ox} contribute to the threshold voltage shift depending on their position in the dielectric. The resulting, effective, Q_{ox} can be evaluated as [5]

$$Q_{\mathrm{ox}} = \mathrm{q}_0 \int\int D_{\mathrm{ox}}(x, E) f(x, E)(1 - x/t_{\mathrm{ox}}) \mathrm{d}x \mathrm{d}E\,. \tag{2.33}$$

Here, t_{ox} is the oxide thickness.

2.2.2 The Basic Models

A very common compact model for the MOSFET is the Level 1 model implemented in the circuit simulator SPICE. There the threshold voltage is obtained as

$$V_{\mathrm{th}} = V_{\mathrm{T0}} + \gamma \left(\sqrt{2\phi_p - V_{\mathrm{bs}}} - \sqrt{2\phi_p} \right) + \frac{Q_{\mathrm{it}} + Q_{\mathrm{ox}}}{C'}\,, \tag{2.34}$$

where V_{T0} is the threshold voltage for $V_{\mathrm{bs}} = 0\,\mathrm{V}$ and γ the body-effect parameter, defined as

$$\gamma = \frac{\sqrt{2\varepsilon_{\mathrm{S}} \mathrm{q}_0 N_{\mathrm{A}}}}{C'}\,. \tag{2.35}$$

Here, ε_S is the permittivity of the silicon substrate and N_A the acceptor doping concentration. The potential in the neutral p-type region ϕ_p is evaluated as

$$\phi_p = \frac{k_B T}{q_0} \ln \frac{N_A}{n_i} . \tag{2.36}$$

During NBTI stress this threshold voltage is shifted due to trapped charges by ΔV_{th} obtained from (2.31).

In the linear regime, where

$$V_{gs} > V_{th} \quad \text{and} \quad V_{ds} < V_{gs} - V_{th} \tag{2.37}$$

the drain current is obtained as

$$I_{ds} = \mu C' \frac{W}{L_{eff}} \left(V_{gs} - V_{th} - \frac{V_{ds}}{2} \right) V_{ds} (1 + \lambda V_{ds}) . \tag{2.38}$$

Here, μ is the effective carrier mobility, W the device width, L_{eff} the effective gate length, and λ is an empirical parameter reflecting the channel length modulation.

In the saturation regime, where

$$V_{gs} > V_{th} \quad \text{and} \quad V_{ds} > V_{gs} - V_{th} \tag{2.39}$$

the drain current is modeled as

$$I_{ds} = \frac{\mu C'}{2} \frac{W}{L_{eff}} (V_{gs} - V_{th})^2 (1 + \lambda V_{ds}) . \tag{2.40}$$

2.2.3 NBTI Related Models

In NBTI investigations the degradation of the transconductance g_m is an important figure of merit. The transconductance is defined as the change of drain current as a result of a change in the gate voltage

$$g_m = \frac{\Delta I_d}{\Delta V_g} . \tag{2.41}$$

Devine *et al.* proposed a transconductance shift versus interface trap (N_{it}), relation as [6]

$$\Delta g_m = g_{m0} \frac{\alpha N_{it}}{1 + \alpha N_{it}} , \tag{2.42}$$

where α is a processing related parameter. As mobility model Devine proposed

$$\mu = \frac{\mu_0}{1 + \alpha N_{it}} . \tag{2.43}$$

With respect to modeling of NBTI these equations show very well how, for example, a reduction of the carrier mobility reduces the drain current, as does an increase of the threshold voltage.

2.3 Carrier Generation and Recombination

Carrier generation is a process where electron-hole pairs are created by exciting an electron from the valence band of the semiconductor to the conduction band, thereby creating a hole in the valence band. Recombination is the reverse process where electrons and holes from the conduction respectively valence band recombine and are annihilated. In semiconductors several different processes exist which lead to generation or recombination, the most important ones are:

- photon transition or optical generation/recombination,
- phonon transition or Shockley-Read-Hall generation/recombination,
- Auger generation/recombination or three particle transitions, and
- impact ionization.

In thermal equilibrium the generation and recombination processes are in dynamic equilibrium. When the system is supplied with additional energy, for example through the absorption of photons or the influence of temperature, additional carriers are generated. The most important generation/recombination processes for the simulation of semiconductor devices are summarized in the following.

2.3.1 Photon Transition

The photon transition is a direct, band-to-band, generation/recombination process. An electron from the conduction band falls back to the valence-band and releases its energy in the form of a photon (light). The reverse process, the generation of an electron-hole pair, is triggered by a sufficiently energetic photon which transfers its energy to a valence band electron which is excited to the conduction band leaving a hole behind. The photon energy for this process has to be at least of the magnitude of the band-gap energy E_g. Figure 2.1 gives an overview of this process. The initial electron/hole constellation is found in (a) while the constellation after the generation/recombination process is found in (b).

For these state changes in the semiconductor the energy *and* the momentum have to be conserved. The energy is emitted or absorbed via a photon with the energy

$$E = h\nu\,, \tag{2.44}$$

where h is Planck's constant and ν is the frequency of the emitted or absorbed photon. However, as the momentum of a photon is very small no momentum transfer is possible in the transition process. Therefore, only direct band-to-band transitions are possible, where no change in momentum is necessary. As silicon and germanium are indirect semiconductors and have their valence band maximum and their conduction band minimum on different positions in momentum space, direct transitions are very unlikely to occur and can in most cases be neglected for those materials. In direct semiconductors like GaAs, this effect is very important.

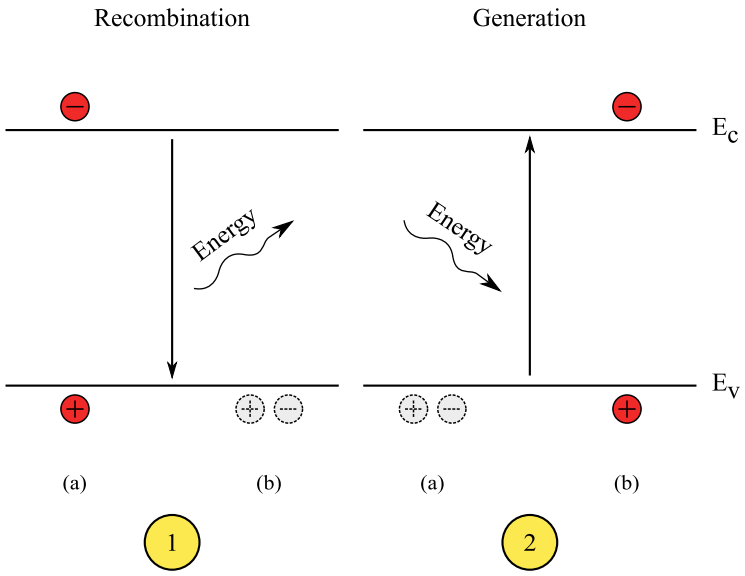


Figure 2.1: Direct generation/recombination process. During photon assisted recombination an electron from the conduction band re-combines with a hole in the valence band. The excess energy is transferred to a photon. The reverse process obtains its energy from radiation and generates an electron hole pair.

The process of carrier recombination is directly proportional to the amount of available electrons and holes. By assuming the capture and emission rates $C_{\mathrm{c}}^{\mathrm{OPT}}$ and $C_{\mathrm{e}}^{\mathrm{OPT}}$ the recombination and generation ($G = -R$) rates can be written as

$$R_{np}^{\mathrm{OPT}} = C_{\mathrm{c}}^{\mathrm{OPT}} np\,, \tag{2.45}$$

$$G_{np}^{\mathrm{OPT}} = C_{\mathrm{e}}^{\mathrm{OPT}}\,. \tag{2.46}$$

As the recombination and generation are balanced in thermal equilibrium, where $np = n_{\mathrm{i}}^2$

$$R_{np}^{\mathrm{OPT}} = G_{np}^{\mathrm{OPT}}\,, \tag{2.47}$$

$$C_{\mathrm{c}}^{\mathrm{OPT}} n_{\mathrm{i}}^2 = C_{\mathrm{e}}^{\mathrm{OPT}}\,, \tag{2.48}$$

the total band-to-band recombination is calculated as the deviation from the thermal equilibrium

$$R^{\mathrm{OPT}} = C_{\mathrm{c}}^{\mathrm{OPT}}(np - n_{\mathrm{i}}^2)\,. \tag{2.49}$$

This process always strives for thermal equilibrium. For an excess concentration of carriers $np - n_{\mathrm{i}}^2 > 0$ and carrier recombination dominates, while for low carrier densities $np - n_{\mathrm{i}}^2 < 0$ and carrier generation prevails.

2.3.2 Phonon Transition

Another process is the generation/recombination by phonon emission. This process is trap-assisted utilizing a lattice defect at the energy level E_{t} within the semiconductor band-gap. The

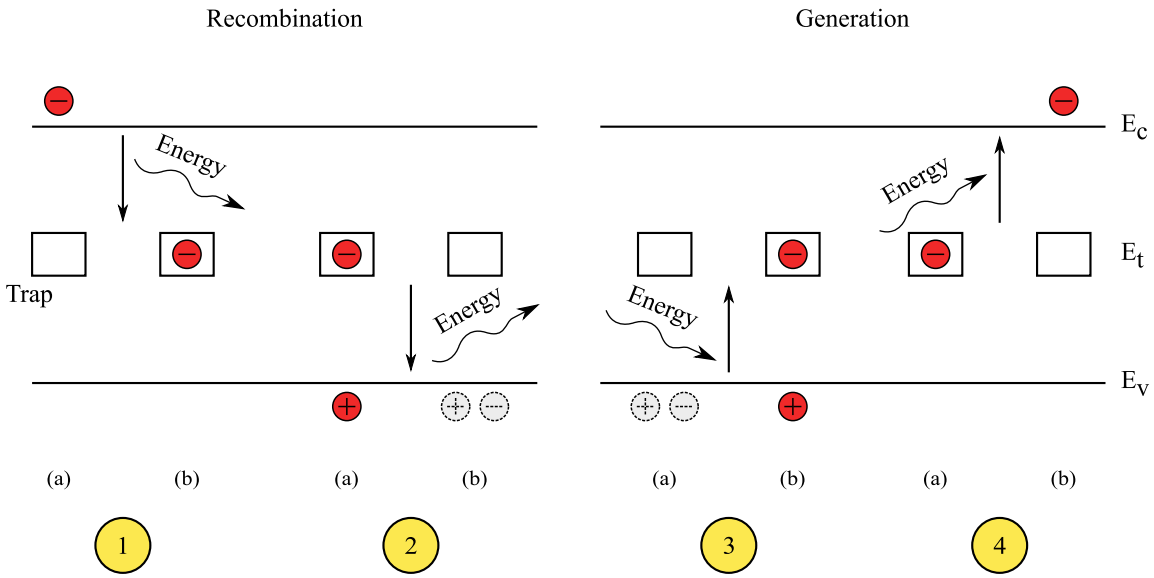


Figure 2.2: Four sub-processes in the Shockley-Read-Hall generation/recombination process. 1. electron capture, 2. hole capture, 3. hole emission, and 4. electron emission.

excess energy during recombination and the necessary energy for generation is transferred to and from the crystal lattice (phonon). A theory describing this effect has been established by Shockley, Read, and Hall [7, 8]. Therefore, the effect is throughout the literature referenced as Shockley-Read-Hall (SRH) generation/recombination. Four sub-processes are possible:

1. Electron capture. An electron from the conduction band is captured by an empty trap in the band-gap of the semiconductor. The excess energy of $E_c - E_t$ is transferred to the crystal lattice (phonon emission).

2. Hole capture. The trapped electron moves to the valence band and neutralizes a hole (the hole is captured by the occupied trap). A phonon with the energy $E_t - E_v$ is generated.

3. Hole emission. An electron from the valence band is trapped leaving a hole in the valence band (the hole is emitted from the empty trap to the valence band). The energy necessary for this process is $E_t - E_v$.

4. Electron emission. A trapped electron moves from the trap energy level to the conduction band. For this process additional energy of the magnitude $E_c - E_t$ has to be supplied.

These sub-processes are illustrated in Figure 2.2. (a) gives the initial electron/hole constellation, the arrows schematically depict the transition, and (b) gives the constellation after the sub-process.

Both, the recombination and the generation processes are two-step processes. The sequential occurrence of sub-processes 1 and 2 leads to recombination of an electron-hole pair. The excess energy of approximately the band-gap energy is transferred to the crystal lattice via lattice

vibrations, phonons. For the SRH generation of an electron-hole pair sub-processes 3 and 4 are responsible. Here external energy has to be supplied from the lattice.

To derive an expression for the total recombination rate R^{SRH}, rates for every sub-process are introduced. Here, acceptor like traps are assumed, which are neutral when empty and negatively charged when occupied by an electron. The derivation for donor traps, which are neutral when occupied by an electron and positively charged when empty, is similar and delivers the same result.

The electron capture rate v_{ec} is proportional to the electron concentration in the conduction band n, the concentration of empty traps N_{t}^{0}, and a proportionality constant k_{ec}. As the available electrons are spread in energy in the conduction band, we must consider the electron capture rate for different energies E. With the energy dependent distribution function for electrons $f_{\mathrm{e}}(E)$ and the density-of-states $g_{\mathrm{e}}(E)$ we get

$$\mathrm{d}v_{\mathrm{ec}} = k_{\mathrm{ec}}(E)\, N_{\mathrm{t}}^{0}\, f_{\mathrm{e}}(E)\, g_{\mathrm{e}}(E)\, \mathrm{d}E\,. \tag{2.50}$$

This is the differential electron capture rate at energy E. The total amount of conduction band electrons is

$$n = \int_{E_{\mathrm{c}}}^{\infty} f_{\mathrm{e}}(E)\, g_{\mathrm{e}}(E)\, \mathrm{d}E\,. \tag{2.51}$$

The hole capture rate v_{hc} is proportional to the hole concentration in the valence band p, the concentration of filled traps N_{t}^{-}, and a proportionality constant k_{hc}. Again, we consider the spread of the holes in energy,

$$\mathrm{d}v_{\mathrm{hc}} = k_{\mathrm{hc}}(E)\, N_{\mathrm{t}}^{-}\, f_{\mathrm{h}}(E)\, g_{\mathrm{h}}(E)\, \mathrm{d}E\,. \tag{2.52}$$

Here, $f_{\mathrm{h}}(E)$ is the distribution function for holes and $g_{\mathrm{h}}(E)$ the density-of-states. The total amount of holes in the valence band is

$$p = \int_{E_{\mathrm{v}}}^{\infty} f_{\mathrm{h}}(E)\, g_{\mathrm{h}}(E)\, \mathrm{d}E\,. \tag{2.53}$$

The hole emission rate v_{he} is proportional to the concentration of empty traps, the proportionality constant k_{he},

$$\mathrm{d}v_{\mathrm{he}} = k_{\mathrm{he}}(E)\, N_{\mathrm{t}}^{0}\, (1 - f_{\mathrm{h}}(E))\, g_{\mathrm{h}}(E)\, \mathrm{d}E\,. \tag{2.54}$$

And finally the electron emission rate v_{ee} is proportional to the concentration of filled traps and the proportionality constant k_{ee}

$$\mathrm{d}v_{\mathrm{ee}} = k_{\mathrm{ee}}(E)\, N_{\mathrm{t}}^{-}\, (1 - f_{\mathrm{e}}(E))\, g_{\mathrm{e}}(E)\, \mathrm{d}E\,. \tag{2.55}$$

The total trap concentration N_{t} is

$$N_{\mathrm{t}} = N_{\mathrm{t}}^{0} + N_{\mathrm{t}}^{-}\,, \tag{2.56}$$

and the fraction of occupied traps f_t is given by

$$f_\mathrm{t} = \frac{N_\mathrm{t}^-}{N_\mathrm{t}}\,, \quad 1 - f_\mathrm{t} = \frac{N_\mathrm{t}^0}{N_\mathrm{t}}\,. \tag{2.57}$$

With these definitions the net recombination rate for electrons becomes

$$\mathrm{d}R_\mathrm{e}^\mathrm{SRH} = \mathrm{d}v_\mathrm{ec} - \mathrm{d}v_\mathrm{ee} = \left[k_\mathrm{ec}(E)\, N_\mathrm{t}^0\, f_\mathrm{e}(E) - k_\mathrm{ee}(E)\, N_\mathrm{t}^-\, (1 - f_\mathrm{e}(E))\right]\, g_\mathrm{e}(E)\, \mathrm{d}E\,. \tag{2.58}$$

In thermal equilibrium ($np = n_0 p_0 = n_\mathrm{i}^2$) the net generation equals zero, which means that the respective capture and emission rates for electrons and holes must be equal

$$v_\mathrm{ec} = v_\mathrm{ee}\,, \quad v_\mathrm{hc} = v_\mathrm{he}\,. \tag{2.59}$$

From (2.58) we obtain using (2.59)

$$\frac{k_\mathrm{ee}(E)}{k_\mathrm{ec}(E)} = \frac{1 - f_\mathrm{t}}{f_\mathrm{t}} \frac{f_\mathrm{e}(E)}{1 - f_\mathrm{e}(E)}\,. \tag{2.60}$$

In thermal equilibrium f_t is given by Fermi-Dirac statistics

$$f(E) = \frac{1}{1 + \exp\left(\dfrac{E - E_\mathrm{F}}{\mathrm{k_B} T_\mathrm{L}}\right)}\,, \tag{2.61}$$

with the property

$$\frac{f}{1 - f} = \exp\left(-\frac{E - E_\mathrm{F}}{\mathrm{k_B} T_\mathrm{L}}\right)\,. \tag{2.62}$$

The ratio (2.60) then calculates as

$$\frac{k_\mathrm{ee}(E)}{k_\mathrm{ec}(E)} = \exp\left(\frac{E_\mathrm{t} - E_\mathrm{F}}{\mathrm{k_B} T_\mathrm{L}}\right) \exp\left(-\frac{E - E_\mathrm{F}}{\mathrm{k_B} T_\mathrm{L}}\right) = \exp\left(\frac{E_\mathrm{t} - E}{\mathrm{k_B} T_\mathrm{L}}\right)\,. \tag{2.63}$$

Using (2.63) we can further develop (2.58)

$$\mathrm{d}R_\mathrm{e}^\mathrm{SRH} = \left[N_\mathrm{t}^0\, f_\mathrm{e}(E) - \frac{k_\mathrm{ee}(E)}{k_\mathrm{ec}(E)}\, N_\mathrm{t}^-\, (1 - f_\mathrm{e}(E))\right]\, k_\mathrm{ec}(E)\, g_\mathrm{e}(E)\, \mathrm{d}E \tag{2.64}$$

$$= \left[(1 - f_\mathrm{t})\, f_\mathrm{e}(E) - \frac{k_\mathrm{ee}(E)}{k_\mathrm{ec}(E)}\, f_\mathrm{t}\, (1 - f_\mathrm{e}(E))\right]\, N_\mathrm{t}\, k_\mathrm{ec}(E)\, g_\mathrm{e}(E)\, \mathrm{d}E \tag{2.65}$$

$$= \left[1 - \frac{k_\mathrm{ee}(E)}{k_\mathrm{ec}(E)} \frac{f_\mathrm{t}}{1 - f_\mathrm{t}} \frac{1 - f_\mathrm{e}(E)}{f_\mathrm{e}(E)}\right]\, (1 - f_\mathrm{t}) f_\mathrm{e}(E) k_\mathrm{ec}(E) g_\mathrm{e}(E) N_\mathrm{t} \mathrm{d}E \tag{2.66}$$

$$= \left[1 - \exp\left(\frac{E_\mathrm{t} - E}{\mathrm{k_B} T_\mathrm{L}}\right) \exp\left(-\frac{E_\mathrm{t} - E_\mathrm{Ft}}{\mathrm{k_B} T_\mathrm{L}}\right) \exp\left(\frac{E - E_\mathrm{F}}{\mathrm{k_B} T_\mathrm{L}}\right)\right] \tag{2.67}$$

$$(1 - f_\mathrm{t}) f_\mathrm{e}(E) k_\mathrm{ec}(E) g_\mathrm{e}(E) N_\mathrm{t} \mathrm{d}E \tag{2.68}$$

$$= \left[1 - \exp\left(\frac{E_\mathrm{Ft} - E_\mathrm{F}}{\mathrm{k_B} T_\mathrm{L}}\right)\right]\, (1 - f_\mathrm{t}) f_\mathrm{e}(E) k_\mathrm{ec}(E) g_\mathrm{e}(E) N_\mathrm{t} \mathrm{d}E\,, \tag{2.69}$$

with the trap's quasi Fermi energy E_{Ft}.

Integrating over all possible electron energies gives the total electron recombination rate

$$R_{\mathrm{e}}^{\mathrm{SRH}} = \left[1 - \exp\left(\frac{E_{\mathrm{Ft}} - E_{\mathrm{F}}}{\mathrm{k_B}T_{\mathrm{L}}}\right)\right] (1 - f_{\mathrm{t}}) N_{\mathrm{t}} \int_{E_{\mathrm{c}}}^{\infty} f_{\mathrm{e}}(E) k_{\mathrm{ec}}(E) g_{\mathrm{e}}(E) \mathrm{d}E \,. \tag{2.70}$$

Typically a capture cross section $\sigma_{\mathrm{e}}(E)$ is introduced to rewrite k_{ec} as

$$k_{\mathrm{ec}}(E) = \sigma_{\mathrm{e}}(E)\, v_{\mathrm{th}}^{\mathrm{e}} \,, \tag{2.71}$$

with the thermal velocity for electrons

$$v_{\mathrm{th}}^{\mathrm{e}} = \sqrt{\frac{3\mathrm{k_B}T_{\mathrm{L}}}{m}} \,, \tag{2.72}$$

resulting in

$$R_{\mathrm{e}}^{\mathrm{SRH}} = \left[1 - \exp\left(\frac{E_{\mathrm{Ft}} - E_{\mathrm{F}}}{\mathrm{k_B}T_{\mathrm{L}}}\right)\right] (1 - f_{\mathrm{t}}) N_{\mathrm{t}} v_{\mathrm{th}}^{\mathrm{e}} \int_{E_{\mathrm{c}}}^{\infty} f_{\mathrm{e}}(E) \sigma_{\mathrm{e}}(E) g_{\mathrm{e}}(E) \mathrm{d}E \,. \tag{2.73}$$

For non-degenerate semiconductors near equilibrium a Maxwell-Boltzmann distribution can be assumed

$$f(E) = \exp\left(-\frac{E - E_{\mathrm{F}}}{\mathrm{k_B}T_{\mathrm{L}}}\right) \,, \tag{2.74}$$

and one obtains for the integral in (2.73)

$$n \langle\!\langle \sigma_{\mathrm{e}} \rangle\!\rangle = \int_{E_{\mathrm{c}}}^{\infty} \exp\left(-\frac{E - E_{\mathrm{F}}}{\mathrm{k_B}T_{\mathrm{L}}}\right) \sigma_{\mathrm{e}}(E) g_{\mathrm{e}}(E) \mathrm{d}E \,, \tag{2.75}$$

with the properties

$$n = N_{\mathrm{c}} \exp\left(-\frac{E_{\mathrm{c}} - E_{\mathrm{F}}}{\mathrm{k_B}T_{\mathrm{L}}}\right) \qquad \text{and} \quad K_{\mathrm{ec}} = N_{\mathrm{t}} \langle\!\langle k_{\mathrm{ec}} \rangle\!\rangle = N_{\mathrm{t}}\, v_{\mathrm{th}}^{\mathrm{e}} \langle\!\langle \sigma_{\mathrm{e}} \rangle\!\rangle \,, \tag{2.76}$$

where N_{c} is the effective density-of-states for electrons we have

$$R_{\mathrm{e}}^{\mathrm{SRH}} = \left[n - N_{\mathrm{c}} \exp\left(-\frac{E_{\mathrm{c}} - E_{\mathrm{F}}}{\mathrm{k_B}T_{\mathrm{L}}}\right) \exp\left(\frac{E_{\mathrm{Ft}} - E_{\mathrm{F}}}{\mathrm{k_B}T_{\mathrm{L}}}\right)\right] (1 - f_{\mathrm{t}}) K_{\mathrm{ec}} \,, \tag{2.77}$$

and

$$R_{\mathrm{e}}^{\mathrm{SRH}} = \left[n(1 - f_{\mathrm{t}}) - N_{\mathrm{c}} \exp\left(-\frac{E_{\mathrm{c}} - E_{\mathrm{t}}}{\mathrm{k_B}T_{\mathrm{L}}}\right) f_{\mathrm{t}}\right] K_{\mathrm{ec}} \,. \tag{2.78}$$

Introducing the auxiliary quantity

$$n_1 = N_{\mathrm{c}} \exp\left(-\frac{E_{\mathrm{c}} - E_{\mathrm{t}}}{\mathrm{k_B}T_{\mathrm{L}}}\right) \,, \tag{2.79}$$

we finally get for the electron recombination rate

$$R_{\mathrm{e}}^{\mathrm{SRH}} = (n(1-f_{\mathrm{t}}) - n_1 f_{\mathrm{t}})\, K_{\mathrm{ec}}\,. \tag{2.80}$$

Analogously the hole recombination rate can be obtained as

$$R_{\mathrm{h}}^{\mathrm{SRH}} = (p f_{\mathrm{t}} - p_1(1-f_{\mathrm{t}}))\, K_{\mathrm{hc}}\,, \tag{2.81}$$

by introducing

$$p_1 = N_{\mathrm{v}} \exp\left(\frac{E_{\mathrm{v}} - E_{\mathrm{t}}}{\mathrm{k_B} T_{\mathrm{L}}}\right)\,. \tag{2.82}$$

2.3.2.1 Dynamic Case

In transient simulations the capture and emission rates are not equal. Therefore, no further simplifications are possible and an additional equation has to be solved for each trap

$$\frac{\mathrm{d}n_{\mathrm{t}}}{\mathrm{d}t} = R_{\mathrm{e}}^{\mathrm{SRH}} - R_{\mathrm{h}}^{\mathrm{SRH}}\,. \tag{2.83}$$

This increases the computational effort significantly but is necessary for example to simulate the charge pumping effect (Section 4.1).

2.3.2.2 Stationary Case

In the stationary case electrons and holes always act in pairs thus the recombination rates for electrons and holes are equal,

$$R_n^{\mathrm{SRH}} = R_p^{\mathrm{SRH}} = R^{\mathrm{SRH}}\,. \tag{2.84}$$

Therefore, we can calculate f_{t} from (2.80) and (2.81) as

$$f_{\mathrm{t}} = \frac{k_{\mathrm{ec}} n + k_{\mathrm{hc}} p_1}{k_{\mathrm{ec}}(n+n_1) + k_{\mathrm{hc}}(p+p1)}\,. \tag{2.85}$$

Using this expression for the total recombination rate we get

$$\begin{aligned} R^{\mathrm{SRH}} &= R_n^{\mathrm{SRH}} = k_{\mathrm{ec}} N_{\mathrm{t}}\,(n(1-f_{\mathrm{t}}) - n_1 f_{\mathrm{t}}) && (2.86)\\ &= k_{\mathrm{ec}} N_{\mathrm{t}} \left(n(1 - \frac{k_{\mathrm{ec}} n + k_{\mathrm{hc}} p_1}{k_{\mathrm{ec}}(n+n_1) + k_{\mathrm{hc}}(p+p1)}) - n_1 \frac{k_{\mathrm{ec}} n + k_{\mathrm{hc}} p_1}{k_{\mathrm{ec}}(n+n_1) + k_{\mathrm{hc}}(p+p1)}\right) && (2.87)\\ &= k_{\mathrm{ec}} k_{\mathrm{hc}} N_{\mathrm{t}} \frac{np - n_1 p_1}{k_{\mathrm{ec}}(n+n_1) + k_{\mathrm{hc}}(p+p_1)}\,. && (2.88) \end{aligned}$$

It is very common to introduce carrier lifetimes for electrons and holes τ_n and τ_p

$$\tau_n = \frac{1}{k_{\mathrm{ec}} N_{\mathrm{t}}}, \quad \tau_p = \frac{1}{k_{\mathrm{hc}} N_{\mathrm{t}}}\,. \tag{2.89}$$

By using the capture cross sections for electrons and holes, σ_e and σ_h, and the thermal velocities v_th^e and v_th^h

$$\tau_n = \frac{1}{\sigma_\mathrm{e} v_\mathrm{th}^\mathrm{e} N_\mathrm{t}}, \quad \tau_p = \frac{1}{\sigma_\mathrm{h} v_\mathrm{th}^\mathrm{h} N_\mathrm{t}}, \tag{2.90}$$

we come to the final formulation of the Shockley-Read-Hall model for carrier generation/recombination

$$R^\mathrm{SRH} = \frac{np - n_\mathrm{i}^2}{\tau_p(n + n_1) + \tau_n(p + p_1)}. \tag{2.91}$$

Regarding the efficiency of trap centers it can be seen that the energy transfer necessary for generation/recombination is always approximately the band-gap energy, no matter where the trap energy level is. The reason is that carriers are transferred from one energy band-edge to the trap level and further to the other band-edge, giving in total the band-gap energy. But when the two sub processes capture and emission are considered, it can be seen that the further the trap energy is away from the mid-gap energy, the higher is the necessary energy for either capture or emission and the lower for the respective other process. The highest energy in this two-step process is always limiting the total generation/recombination. When the trap is located in the middle of the band-gap, the resulting energy barrier height is half the band-gap energy. As the trap is moved away from the mid-gap energy, the limiting energy barrier is increased and the probability of generation/recombination is reduced.

Impurities used for doping semiconductors are usually energetically situated very close to either the valence or the conduction band in order to be effective doping centers. They are therefore not very effective for carrier generation/recombination and are called “shallow” traps. “Deep” traps on the other hand are located close to the mid-gap which can be used to artificially increase the carrier generation or recombination.

2.3.2.3 Surface Generation/Recombination

For the investigation of NBTI the generation and recombination mechanisms at the silicon/dielectric interface are of major importance (Chapter 3). The Shockley-Read-Hall generation/recombination mechanism can also be applied to traps at the interface, which is for example obligatory for the simulation of the charge pumping effect (Section 4.1).

The derivation for recombination at surface traps is similar to the derivation for bulk traps. The major difference is the different unit for interface traps $[N_\mathrm{it}] = 1/\mathrm{cm}^2$ and the resulting unit for the surface recombination velocity $[R_\mathrm{it}^\mathrm{SRH}] = 1/\mathrm{cm}^2\mathrm{s}$.

2.3.2.4 Distributed Traps

As described in detail in Section 3.1.1, interface traps are not located on discrete energy levels but distributed in the band-gap instead. When accounting for the trap density-of-states $D_\mathrm{it}(E)$,

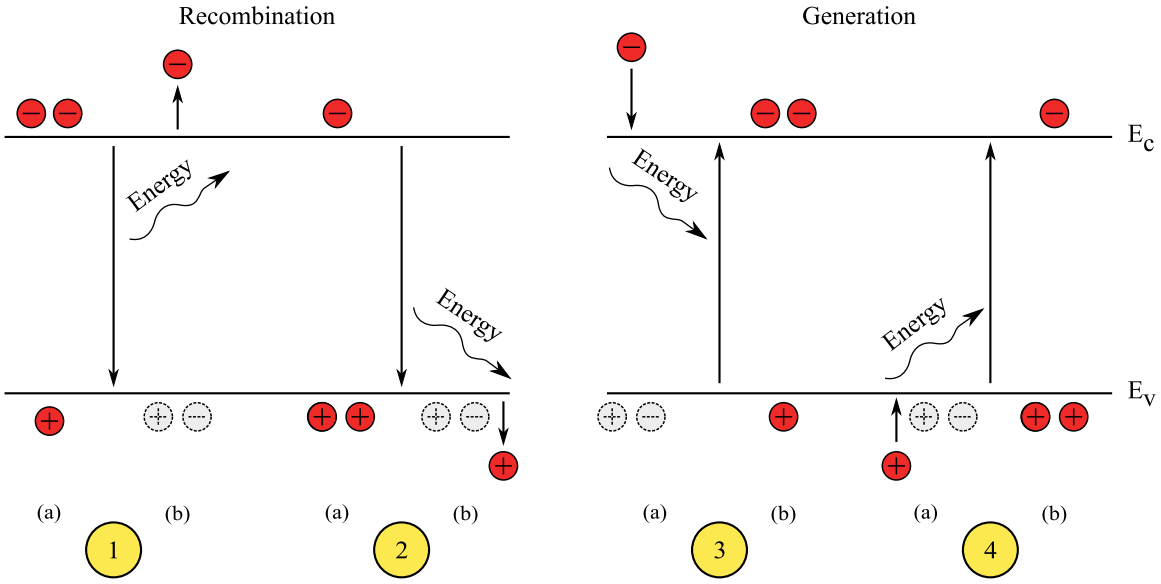


Figure 2.3: Four sub-processes in the Auger generation/recombination mechanism. 1. electron capture, 2. hole capture, 3. electron emission, and 4. hole emission.

we get for the interface trap concentration

$$N_{\mathrm{it}} = \int_{E_{\mathrm{v}}}^{E_{\mathrm{i}}} D_{\mathrm{it}}(E)\mathrm{d}E = \int_{E_{\mathrm{i}}}^{E_{\mathrm{c}}} D_{\mathrm{it}}(E)\mathrm{d}E\,. \tag{2.92}$$

The interface trap recombination rate is then obtained as

$$R_{\mathrm{it}}^{\mathrm{SRH}} = \int_{E_{\mathrm{v}}}^{E_{\mathrm{c}}} \frac{np - n_{\mathrm{i}}^2}{\tau_p(n + n_1) + \tau_n(p + p_1)} D_{\mathrm{it}}(E)\mathrm{d}E\,. \tag{2.93}$$

2.3.3 Auger Generation/Recombination

In the direct band-to-band Auger mechanism three particles are involved. During generation an electron hole pair is generated consuming the energy of a highly energetic particle. In the opposite process, when an electron hole pair recombines, the excess energy is transferred to a third particle. In detail the four possible processes are as follows:

1. Electron capture. An electron from the conduction band moves to the valence band neutralizing a hole in the valence band. The excess energy is transferred to an electron in the conduction band.

2. Hole capture. Again, an electron from the conduction band moves to the valence band and recombines with a valence hole. The excess energy is, in contrast to Process 1, transferred to another *hole* in the valence band.

3. Electron emission. A highly energetic electron from the conduction band transfers its energy to an electron in the valence band. The valence electron moves to the conduction band generating an electron hole pair.

4. Hole emission. A highly energetic hole from the valence band transfers its energy to an electron in the valence band which is then excited to the conduction band generating an electron hole pair.

These sub-processes are illustrated in Figure 2.3. (a) gives the initial electron/hole constellation, the arrows schematically depict the transition, and (b) gives the constellation after the sub-process.

As for the Shockley-Read-Hall effect a model can be derived by setting up rates for the four processes. For electron capture two electrons in the conduction band and one hole in the valence band are necessary. Using k_{ec} as the rate constant, the electron capture rate v_{ec} becomes

$$v_{\mathrm{ec}} = k_{\mathrm{ec}} n^2 p \,. \tag{2.94}$$

Analogical for hole capture where two holes and one electron are involved v_{hc} evaluates as

$$v_{\mathrm{hc}} = k_{\mathrm{hc}} n p^2 \,. \tag{2.95}$$

For electron and hole emission only one respective carrier is necessary

$$v_{\mathrm{ee}} = k_{\mathrm{ee}} n \,, \tag{2.96}$$
$$v_{\mathrm{he}} = k_{\mathrm{he}} p \,. \tag{2.97}$$

In thermal equilibrium the respective capture and emission rates are in equilibrium, and therefore

$$v_{\mathrm{ec},0} = v_{\mathrm{ee},0} \,, \quad k_{\mathrm{ec}} n_{\mathrm{i}}^2 = k_{\mathrm{ee}} \,, \tag{2.98}$$
$$v_{\mathrm{hc},0} = v_{\mathrm{he},0} \,, \quad k_{\mathrm{hc}} n_{\mathrm{i}}^2 = k_{\mathrm{he}} \,. \tag{2.99}$$

This leads us to the final model for the Auger recombination rate R^{AUG}

$$R^{\mathrm{AUG}} = v_{\mathrm{ec}} + v_{\mathrm{hc}} - v_{\mathrm{ee}} - v_{\mathrm{he}} = (k_{\mathrm{ec}} n + k_{\mathrm{hc}} p)(np - n_{\mathrm{i}}^2) \,. \tag{2.100}$$

Although the Auger mechanism is microscopically exactly the same as the mechanism during impact ionization described in the next section, the energy source is completely different. Whereas impact ionization relies on high current density, only a very large carrier density is of importance for Auger generation/recombination, as can be seen in the final formulation of (2.100).

2.3.4 Impact Ionization

Impact ionization is a pure generation process. Microscopically it is exactly the same mechanism as the generation part of the Auger process: a highly energetic carrier moves to the conduction or valence band, depending on the carrier type, and the excess energy is used to excite an electron

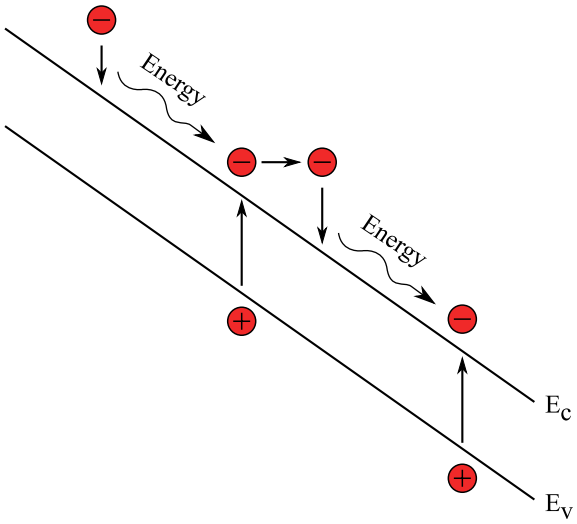


Figure 2.4: Impact ionization and avalanche multiplication. An energetic electron donates its energy to the generation of an electron hole pair. The newly generated electron can, due to the high electric field, obtain high energy and generate further carriers, leading to avalanche multiplication.

from the valence band to the conduction band generating another electron hole pair. The major difference is the cause of the effect. While it is purely the carrier concentration in the Auger mechanism, for impact ionization it is the current density.

Two partial processes can be distinguished:

1. Electron emission. A highly energetic electron from the conduction band transfers its energy to an electron in the valence band. The valence electron moves to the conduction band generating an electron hole pair.

2. Hole emission. A highly energetic hole from the valence band transfers its energy to an electron in the valence band which is then excited to the conduction band generating an electron hole pair.

Figure 2.4 depicts the effect of impact ionization and avalanche multiplication. The leftmost, highly energetic, electron excites a new electron/hole pair which gains energy and generates further carriers. The result is an avalanche multiplication of carrier generation.

As already mentioned, the generation rates are modeled proportional to the current densities $\vec{J}_n$ and $\vec{J}_p$ and can be written as:

$$v_\mathrm{e} = \alpha_\mathrm{e} \frac{|\vec{J}_n|}{\mathrm{q}} \,, \tag{2.101}$$

$$v_\mathrm{h} = \alpha_\mathrm{h} \frac{|\vec{J}_p|}{\mathrm{q}} \,, \tag{2.102}$$

with the ionization rates for electrons and holes, α_e and α_h. These rates are typically described with an exponential dependence upon the electric field component in the direction of the current flow E. With the critical electric fields for electrons and holes, $E_\mathrm{e}^\mathrm{crit}$ and $E_\mathrm{h}^\mathrm{crit}$, and the ionization rates at infinite field, α_e^∞ and α_h^∞, the ionization rates evaluate as

$$\alpha_\mathrm{e} = \alpha_\mathrm{e}^\infty \exp\left(-\left(\frac{E_\mathrm{e}^\mathrm{crit}}{E}\right)^{\beta_\mathrm{e}}\right), \tag{2.103}$$

$$\alpha_\mathrm{h} = \alpha_\mathrm{h}^\infty \exp\left(-\left(\frac{E_\mathrm{h}^\mathrm{crit}}{E}\right)^{\beta_\mathrm{h}}\right). \tag{2.104}$$

Here, β_e and β_h are additional model parameters, which are in the range of 1–2. The total impact ionization rate is now found as

$$R^\mathrm{II} = -v_\mathrm{e} - v_\mathrm{h} = -\alpha_\mathrm{e}\frac{|\vec{J}_n|}{\mathrm{q}} - \alpha_\mathrm{h}\frac{|\vec{J}_p|}{\mathrm{q}}. \tag{2.105}$$

The impact ionization rate does not actually depend on the local electric field but on the carrier temperature and, thus, on the high-energy tail of the distribution function. Therefore, the model is not very accurate, especially in small devices. Carriers need to travel a certain distance in the high electric field in order to gain energy. For the exact modeling semiconductor device equations of higher order are necessary.

2.4 Quantum Mechanical Effects

The classical semiconductor device equations from Section 2.1 imply that the mobile carriers, electrons and holes, behave like classical particles in the semiconductor. For large device dimensions this assumption gives very good results, but for small device geometries quantum mechanical effects like the quantum mechanical tunneling, described in Section 5.3, and the quantum mechanical confinement gain importance. The latter effect leads to a reduction of allowed states for electrons and holes near a Si/SiO_2 interface. In classical device simulations using the drift-diffusion approximation the peak of the electron concentration in the channel of a turned on n-channel MOSFET is calculated to be directly at the Si/SiO_2 interface. This is not correct as the number of allowed states is drastically reduced close to the interface and therefore the peak of the carrier concentration lies several angstroms away from the interface [9].

2.4.1 Quantum Confinement

For the modeling of NBTI the carrier concentration close to the Si/SiO_2 interface plays an important role (Section 6.4.4). The use of quantum confinement models reduces this carrier concentration and might have significant influence on the NBTI model used.

In classical device simulators quantum confinement is often accounted for by using additional quantum correction models. These models locally change the carrier density-of-states [10, 11] or they modify the conduction band edge close to the interface [12].

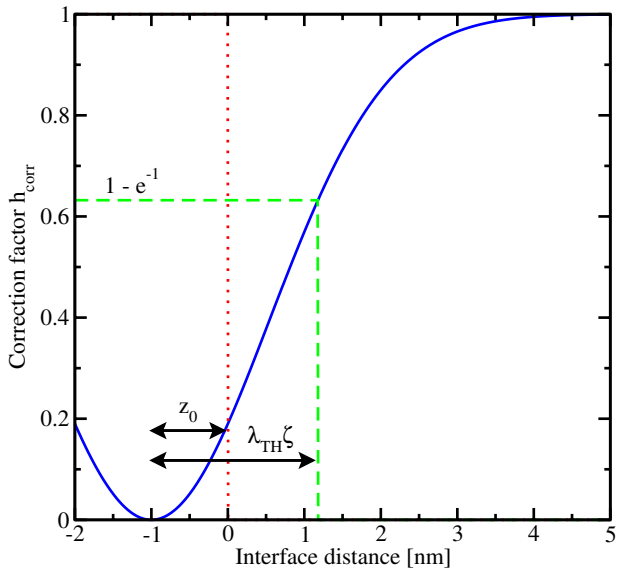


Figure 2.5: Plot of the DOS correction factor h_{corr} for $z_0 = 1\,\mathrm{nm}$ and $\zeta = 1$ at $T = 300\,\mathrm{K}$.

2.4.1.1 Density of States Correction

In classical device simulation the density-of-states (DOS) in homogeneous materials is modeled as a constant value throughout the device. In order to describe the quantum mechanical confinement a distance-dependent reduction of the DOS at the Si/SiO_2 interface has been proposed in [10, 11]

$$h_{\mathrm{corr}} = 1 - \exp\left(-\frac{(z+z_0)^2}{\zeta^2 \lambda_{\mathrm{TH}}^2}\right) , \tag{2.106}$$

where z is the distance to the Si/SiO_2 interface and z_0 shifts the whole function relative to the interface. ζ is a newly introduced parameter which enables the variation of λ_{TH} for calibration purposes. The symbol λ_{TH} denotes the thermal wavelength which is given by

$$\lambda_{\mathrm{TH}} = \frac{\hbar}{\sqrt{2m^* \mathrm{k_B} T}} , \tag{2.107}$$

where $\hbar$ is the reduced Planck constant, m^* is the effective carrier mass, $\mathrm{k_B}$ the Boltzmann constant, and T the temperature. The resulting DOS, N_{c}, is then calculated from the classical DOS $N_{\mathrm{c},0}$, which is normally modeled as a constant throughout the semiconductor, with the correction factor h_{corr} as

$$N_{\mathrm{c}} = N_{\mathrm{c},0} h_{\mathrm{corr}} . \tag{2.108}$$

The interplay of the different parameters and the distance to the Si/SiO_2 interface is schematically depicted in Figure 2.5. The parameter z_0 is important, because with $z_0 = 0$ the DOS at the interface becomes zero. This would cause numerical problems and reduce the convergence of the numerical solver. A positive number of z_0 shifts the correction function towards the dielectric,

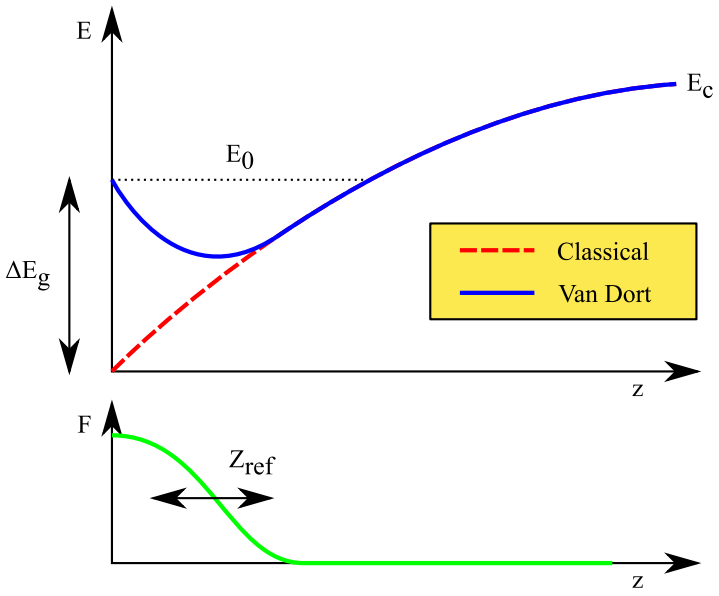


Figure 2.6: Band edge bending at the Si/SiO_2 interface. The classical band edge is corrected by the factor $\Delta E_g F(z)$.

approximately considering wave function penetration. The value of $\zeta\lambda_{TH}$ defines the effective depth of the correction. A high value, which can be achieved with $\zeta > 1$, leads to a reduction of the DOS even deep in the substrate.

Note that the correction factor does not depend on the bias and the band edge energies are not influenced. Hence, the model can be evaluated in a preprocessing step and does not impose any additional computational burden during iteration steps.

2.4.1.2 Conduction Band Edge Correction

An alternative approach is based on E_0, the first eigenvalue of the triangular energy well, as seen in Figure 2.6. This model was proposed by van Dort [12]

$$\Delta E_g = E_0 - E_c(0) = \frac{13}{9}\beta\left(\frac{\varepsilon_{si}}{4qk_BT}\right)^{1/3}|E_n|^{2/3}, \tag{2.109}$$

where the proportionality factor $\beta = 4.1 \times 10^{-8}$ eVcm is is found from the observed threshold voltage shift at high doping levels [13], ε_{si} is the permittivity of silicon, and E_n is the electric field at the Si/SiO_2 interface perpendicular to the interface.

The value of ΔE_g is multiplied with a distance-dependent weight function which has been introduced by Selberherr [14] for the modeling of surface roughness scattering in MOSFETs. The function is of the following form

$$F(z) = \frac{2\exp(-(z/z_{ref})^2)}{1+\exp(-2(z/z_{ref})^2)}, \tag{2.110}$$

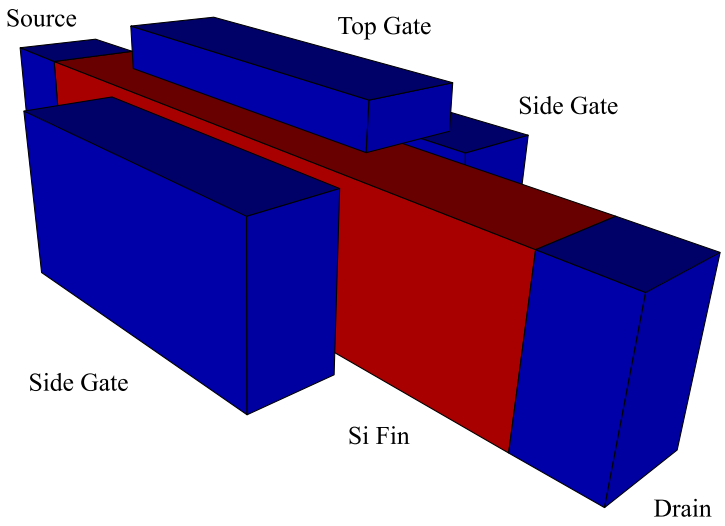


Figure 2.7: Device geometry of a triple-gate FinFET structure. Quantum confinement plays an important role in this device because of its small size and the formation of three, instead of one, channels.

where z_{ref} is the scaling factor for the interface distance. Thus, the resulting band edge energy with van Dort's quantum correction of the classical band edge energy E_{class} reads as follows

$$E_{\text{c}} = E_{\text{class}} + F(z)\Delta E_{\text{g}} \,. \tag{2.111}$$

Figure 2.6 depicts the distance dependent weight function F and the band edge energy for both, the classical approach and after quantum correction with van Dort's method.

2.4.1.3 Evaluation of Quantum Correction Models

For the evaluation of the quantum correction models a state-of-the-art three-dimensional n-channel FinFET device structure was chosen. The device geometry can be seen in Figure 2.7. The silicon fin has a cross section area of $6\times10\,\text{nm}^2$. The gate length is 20 nm with a gate oxide thickness of 1.5 nm. The source and drain regions are heavily n-type doped whereas the channel itself remains undoped.

Figure 2.8 depicts the electron concentration in a two-dimensional cut through the silicon fin in the middle of the channels. The gates are biased at 0.9 V with the source and drain contacts grounded. The classical simulation using the drift-diffusion approximation gives the highest magnitude of the electron concentration at the Si/SiO_2 interfaces below the gate contacts. It can be seen that the peak electron concentration is found in the top corners, as two respective gates couple to the channel, each of them attracting carriers. With the quantum confinement correction models the maximum carrier concentration is moved to the inside of the fin by a distance depending on the chosen model and its calibration parameters.

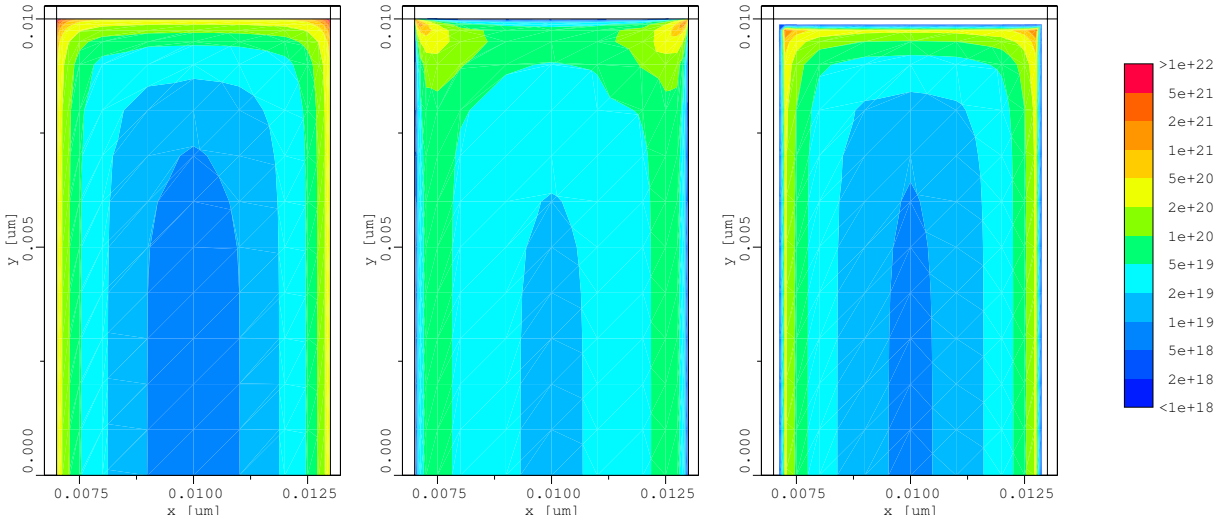


Figure 2.8: Electron concentration in a triple-gate FinFET for classical simulation (left), with the DOS correction model (middle), and the band edge energy correction model by Van Dort. The correction models force the peak of the carrier concentration away from the Si/SiO_2 interface into the substrate.

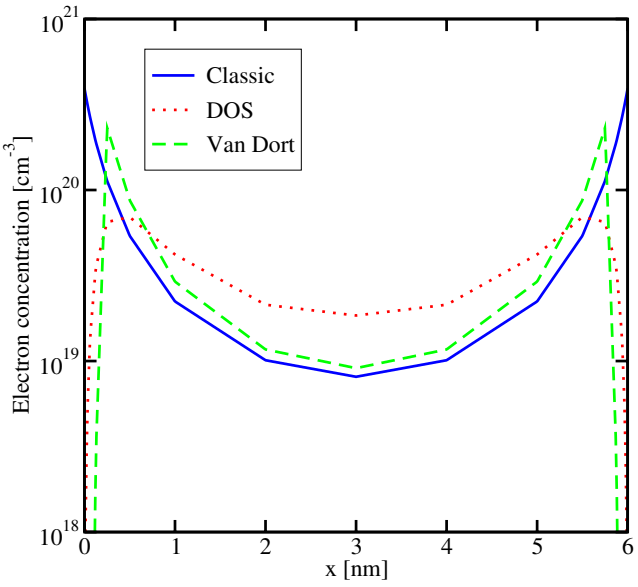


Figure 2.9: Electron concentration across the fin using classical device simulation and the confinement correction models.

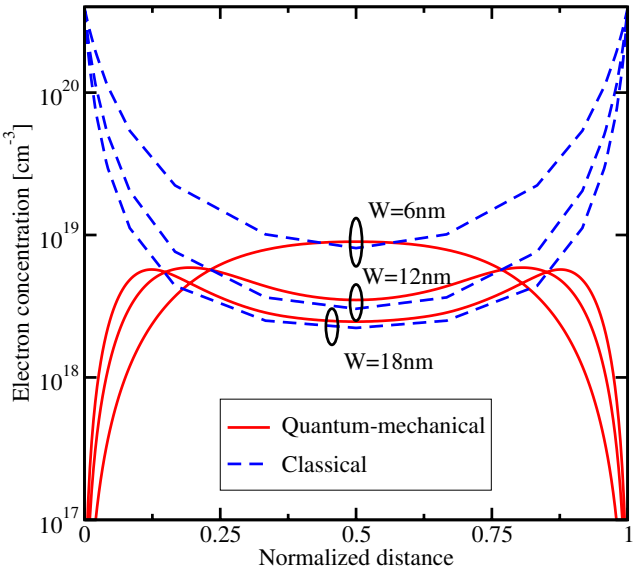


Figure 2.10: Comparison of classical and quantum mechanical carrier concentrations for different fin widths. The quantum mechanical calculations have been performed using a Schrödinger solver.

Figure 2.9 depicts a one-dimensional cut through the fin displaying the carrier concentrations for the different models for the same bias conditions. Comparing with Figure 2.10 reveals that qualitatively the DOS correction model delivers better results.

A comparison of FinFETs with different fin widths can be seen in Figure 2.10. It shows the electron concentration across the fin simulated with both, the classical drift-diffusion model and a Schrödinger solver, respectively. The fin widths are 6, 12, and 18 nm. At a fin width of 6 nm the electron concentration has its maximum in the center of the fin. This shape of the carrier concentration cannot be reproduced by the quantum correction models. With larger widths the maximum moves to the interfaces enabling a better fit of the correction models.

The channels in the silicon fin are displaced from the surface to the inside of the silicon and thus the drive current is reduced. Figure 2.11 depicts the drain current for a gate voltage of 0.9 V and different quantum correction mechanisms. Additionally to the triple-gate device the simulation has been performed with a double-gate structure, where the top gate from Figure 2.7 has been replaced with SiO_2. Simulation of the double-gate structure shows a reduced output current by a factor of approximately 20% due to the formation of only two channels.

Quantum correction leads to a considerable reduction of the saturation current. The DOS correction model yields reasonable results, but since it does not account for the band bending it must be calibrated for each bias point. Van Dort's model completely fails to reproduce the carrier concentration in the channel which may be due to the assumption of a triangular energy well. This assumption might be a too crude estimation for extremely thin channels. Therefore, these models can be very well used to describe the current reduction in very thin channel devices, but

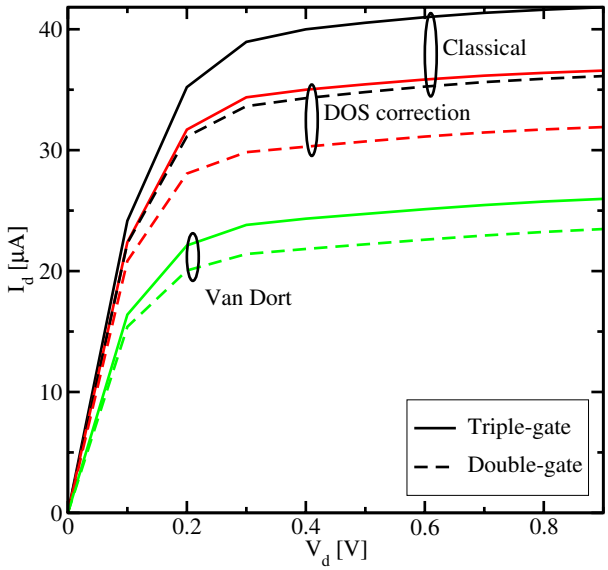


Figure 2.11: Comparison of the output characteristics of double- and triple-gate FinFETs at a gate voltage of 0.9 V. Quantum confinement correction reduces the drain current as no carriers are allowed at the Si/SiO_2 interface.

not when the shape of the carrier concentration is important. Here, the solution of Schrödinger's equation is necessary for accurate simulation of the carrier concentration.

Chapter 3

The Silicon/Silicon-Dioxide Interface

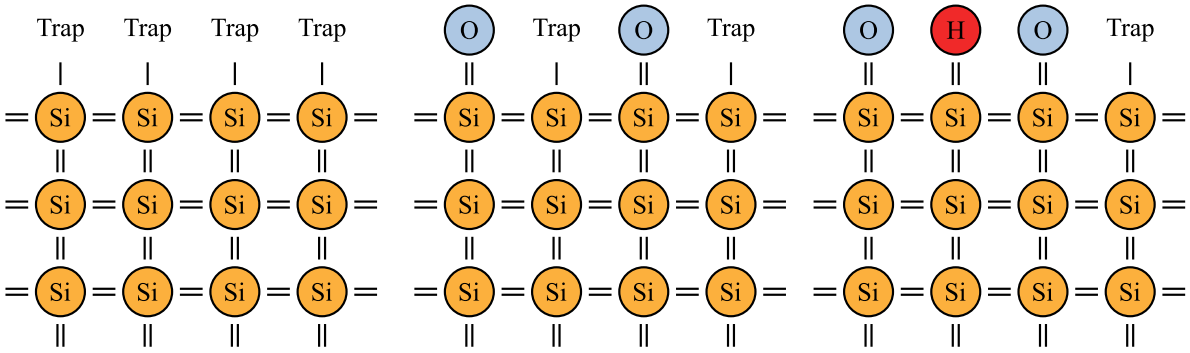
ONE of the most important things that led to the enormous success and continuous improvement of the CMOS technology are the excellent properties of the thermally grown Si/SiO_2 interface. The quality and abruptness of the interface is of high quality and it has a very low defect density [15].

The amorphous SiO_2 is grown on a pure silicon crystal. The lattice mismatch between those two materials is rather large, but the highly flexible angle of the Si-O-Si bond can compensate for a major part of this mismatch. The remaining, unbound silicon atoms at the interface form the prevalent imperfection of the interface, the silicon dangling bonds.

3.1 Silicon Dangling Bonds

The silicon atom possesses four valence electrons and therefore requires four bonds to fully saturate the valence shell. In the crystalline structure each silicon atom establishes bonds to its four neighboring atoms, leaving no unsaturated bond behind. At the surface of the silicon crystal atoms are missing and traps are formed as shown in Figure 3.1(a). The density of these interface states, D_{it}, in this regime is approximately $D_{\mathrm{it}} \approx 10^{14}\,\mathrm{cm}^{-2}\,\mathrm{eV}^{-1}$. After oxidation most interface states are saturated with oxygen atoms (Figure 3.1(b)). The density is then approximately $D_{\mathrm{it}} \approx 10^{12}\,\mathrm{cm}^{-2}\,\mathrm{eV}^{-1}$ [16].

This number is already a major improvement of the interface quality. But in an MOS transistor with a gate length of 100 nm and a gate width of 1 μm this density still translates to 1000 dangling bonds. With such a high number of interface defects a transistor would still not operate properly. Therefore, it is mandatory to increase the quality of the Si/SiO_2 interface in MOS device technology as much as possible. Each electrically active interface state leads to degradation of important transistor parameters such as the threshold voltage, the on-current, or the surface carrier mobility. To further improve the interface, the number of dangling valence bonds is further reduced by annealing the interface in forming gas with hydrogen atoms, as shown in Figure 3.1(c). The dangling silicon bonds are passivated by forming Si-H bonds. With this treatment the amount of electrically active interface states can be reduced to around $D_{\mathrm{it}} \approx 10^{10}\,\mathrm{cm}^{-2}\,\mathrm{eV}^{-1}$.



(a) Silicon surface (b) Unpassivated Si/SiO_2 interface (c) After hydrogen passivation

Figure 3.1: (a) At the silicon surface silicon atoms are missing and unpaired valence electrons exist forming electrically active interface traps. (b) After oxidation most interface states are saturated with oxygen bonds. (c) After annealing the surface with a hydrogen related species the amount of interface defects is further decreased.

This is an acceptable number and a first-class Si/SiO_2 interface can be formed. Ironically, exactly these Si-H bonds are the cause for NBTI, as extensively described in Chapter 6. The bonds can break at elevated temperatures and high electric fields due to their lower binding energy and re-activate the interface states.

The exact properties of the interface defects, which are trivalent silicon atoms with one unpaired valence electron

$$Si_3{\equiv}Si^{\bullet}\,, \tag{3.1}$$

or

$$Si_2O{\equiv}Si^{\bullet}\,, \tag{3.2}$$

depends on the exact atomic configuration and on the orientation of the substrate. While P_b centers (3.1) are formed on (111) oriented substrates, P_{b0} (3.1) and P_{b1} (3.2) centers can only exist on (100) orientations. Figures 3.2 and 3.3 depict the atomic configurations of all three trap types.

Recent works show [18–20] that all three types of P_b centers give rise to trap levels in the silicon band-gap. The charge state of the traps therefore depends on the Fermi-level.

3.1.1 Amphoteric Nature of Dangling Bonds

All three types of silicon dangling bonds investigated up to now (P_b, P_{b0}, and P_{b1}) are reported to be of amphoteric nature. Their energy distribution comprises of two distinct peaks in the silicon band-gap, as seen in Figure 3.6.

The two peaks have different properties regarding their possible charge states and their energetic positions depend on the type of the trap center. For P_b and P_{b0} these are:

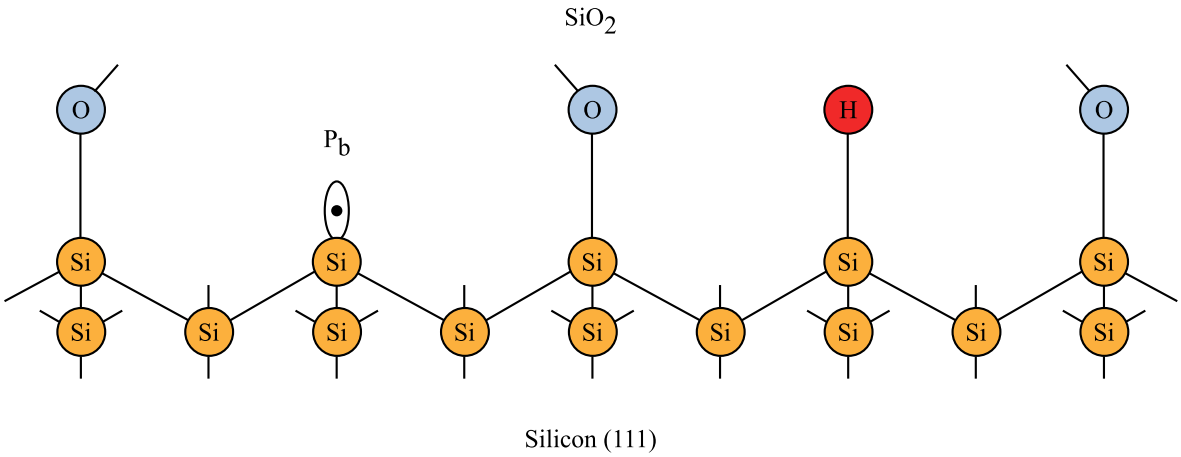


Figure 3.2: P_b defect located at a Si/SiO_2 interface with (111) orientation. The defect is formed by an unpaired valence electron of a silicon atom back-bonded to three other silicon atoms. The defect's trap energy lies in the silicon band-gap. Thus, the charge state of the trap depends on the Fermi-level and it is electrically active [17].

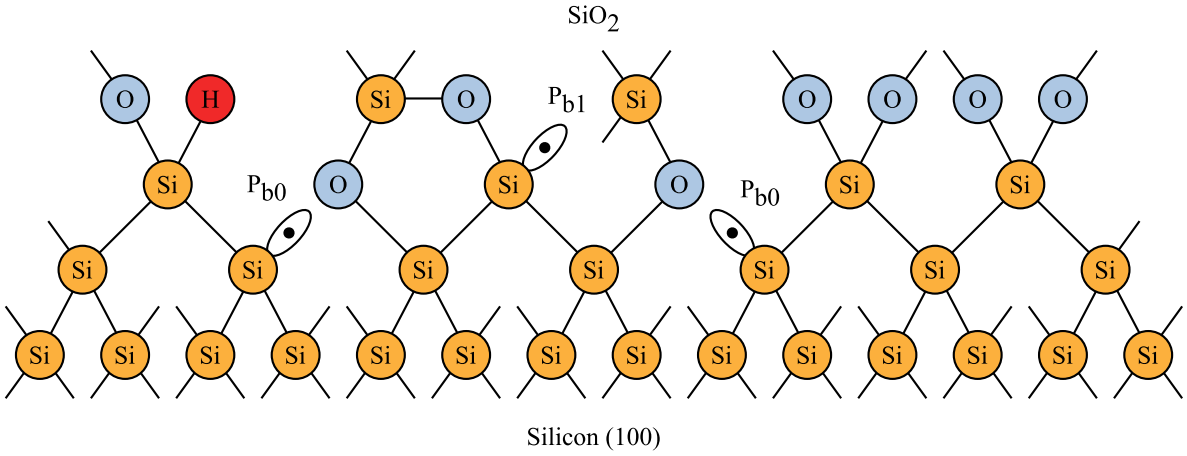


Figure 3.3: P_{b0} and P_{b1} defects located at a Si/SiO_2 interface with (100) orientation. The P_{b0} center is back-bonded to three silicon atoms and electrically very similar to the P_b center. The silicon atom of the P_{b1} center is back-bonded to two other silicon atoms and an oxygen atom. Both traps are electrically active as their energy lies in the silicon band-gap making their charge state Fermi-level dependent [17].

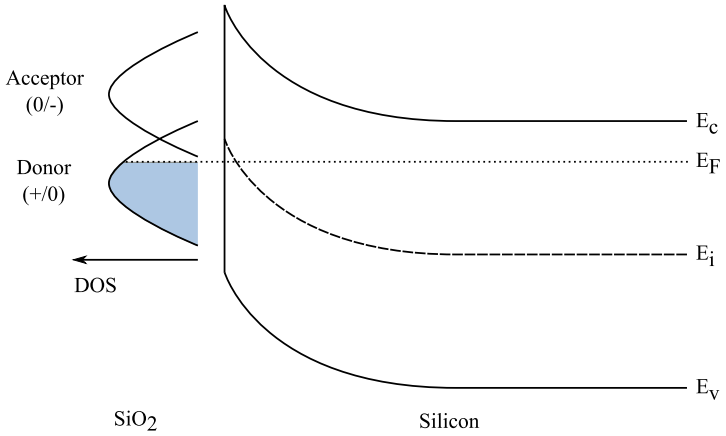


Figure 3.4: Energy diagram of a P_b center at the Si/SiO_2 interface in weak inversion. The trap consists of donor like states in the lower and acceptor like states in the upper half of the band-gap due to its amphoteric nature. The Fermi-level determines the filling and therefore the charge state. In this configuration the trap is slightly positively charged.

1. Donor-like energy levels: They are located in the lower half of the band-gap around 0.25 eV [21] above the valence band edge. The trap levels are positively charged when empty and electrically neutral when occupied by an electron. In the empty state they are diamagnetic, a very weak form of magnetism triggered by an external magnetic field changing the orbital motion of the atoms core electrons. With one unpaired electron the trap levels are paramagnetic, a stronger form of magnetism initiated by the presence of a magnetic field and unpaired electrons. The possible charge states of the P_b centers can be written as $P_b + h^+ = P_b^+$ and $P_b^+ + e^- = P_b$.

2. Acceptor-like energy levels: They are located in the upper half of the band-gap, around 0.85 eV [21] above the valence band. The trap levels are electrically neutral when empty and negatively charged when occupied by an electron. They are therefore paramagnetic when empty (the same state as the filled donor like level, with one electron in the P_b center), and diamagnetic when occupied, as the total amount of electrons in the P_b center is then two. The charge states are $P_b^- + h^+ = P_b$ and $P_b + e^- = P_b^-$.

P_{b1} centers have the same, amphoteric nature but the energy levels of the peaks are different, as shown in Figure 3.6(b).

Figure 3.4 illustrates the determination of the charge state of an amphoteric interface trap in weak inversion. At these conditions the upper peak is totally empty and therefore electrically neutral. The lower peak is filled to approximately two thirds and positively charged.

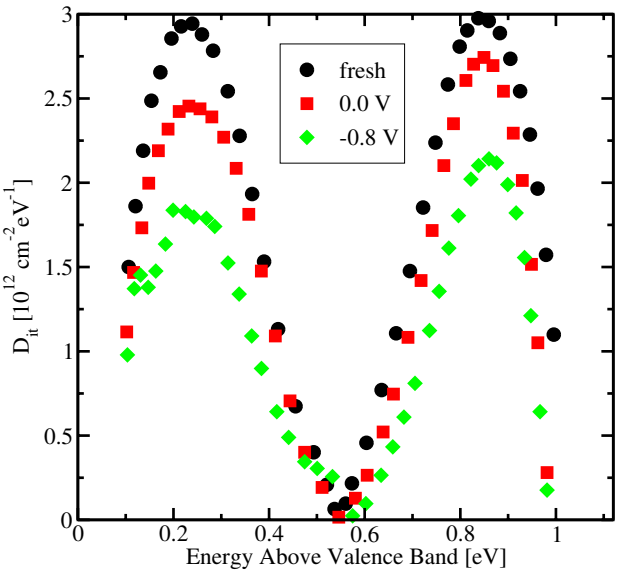
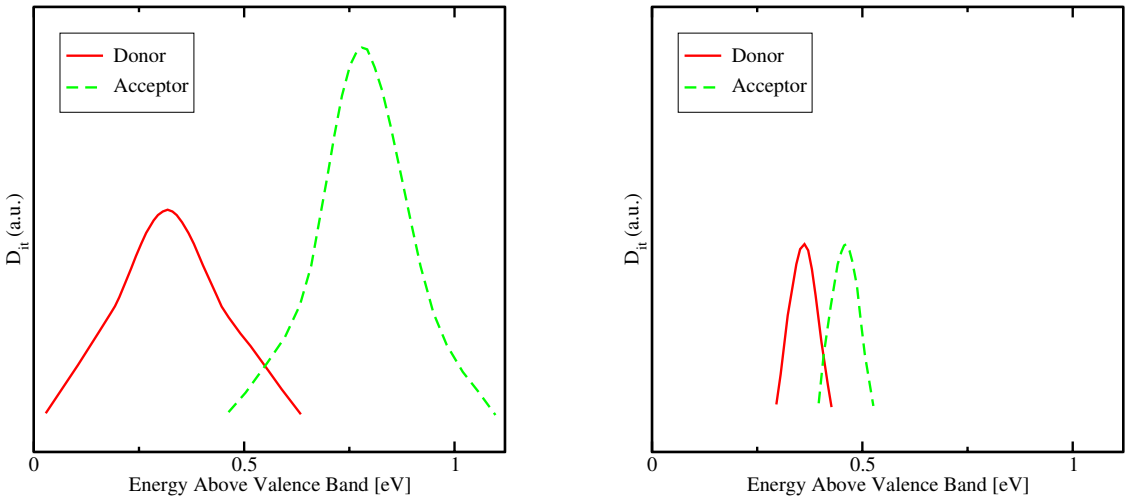


Figure 3.5: Energy distribution, D_{it}, of traps at the (111) Si/SiO_2 interface after Ragnarsson *et al.* [21]. The two peaks of Gauss'ian form are observed in the upper and lower half of the band-gap. After annealing at zero and -0.8 V bias the trap density is reduced.

3.1.2 Characterization of Trap Centers

A tool with the analytical power and sensitivity to identify the atomic-scale structure of the different traps is the electron spin resonance (ESR) measurement [22, 23]. The ESR measurement is suited to investigate the NBTI induced interface state generation [23–25]. The draw-back is that after stressing the device at the poly-gate with high voltages, the gate has to be etched off before ESR measurements can be performed. For the investigation of fully processed devices the spin-dependent recombination (SDR) technique can be used [26, 27].

The SDR technique can be explained by using the Shockley-Read-Hall (SRH) model (Section 2.3.2) for recombination and the Pauli exclusion principle as follows [28]. The device under test (DUT) is operated as a gate controlled diode with the source/drain to substrate slightly forward biased. With this forward bias the current is dominated by recombination at the trap centers of the Si/SiO_2 interface. At a certain gate voltage the recombination current has its maximum, as determined by the DCIV method described in Section 4.2. The DUT is exposed to a large DC magnetic field which is slowly varied to partially align the spins of the paramagnetic charge carriers and the paramagnetic trapping centers. Because of the Pauli exclusion principle it is not possible that a carrier is trapped by a trap having the same spin orientation. When the ESR condition is satisfied, the electron spins are flipped. This increases the probability of different spin orientations for carriers and traps and therefore increases the measured recombination current. This spin dependent increase in recombination current is evaluated in SDR.



(a) P_{b0} defect

(b) P_{b1} defect

Figure 3.6: Estimated density-of-states (DOS) distributions of (a) P_{b0} and (b) P_{b1} centers at the (100) Si/SiO_2 interface. The data are from Lenahan *et al.* [20] and give only a "crude, semi-quantitative estimation" as precise measurements are not yet available. Important details are the broader distribution of the P_{b0} centers and the distribution of most P_{b1} levels within the P_{b0} levels.

Early works [16, 29, 30] concentrated on the investigation of (111) oriented substrates, mainly due to the higher defect density compared to other wafer orientations and therefore easier experimental evaluation. Figure 3.5 gives the energy distribution of the interface trap density D_{it} for P_b centers at (111) oriented interfaces after Ragnarsson [21]. It can be seen that the interface trap energies are spread according to a Gauss'ian peak in the upper and another peak in the lower half of the band-gap. During passivation with hydrogen at elevated temperatures of 170 °C the electrically active dangling bonds are saturated with hydrogen atoms. The peaks of the maximum trap concentration stay approximately at the same energy levels whereas the concentration decreases continuously.

Recent works [18–21, 31] concentrate on the technologically much more important (100) substrate orientation. The standard silicon CMOS processes of most applications use this crystal orientation. The P_{b0} defects found at interfaces of this wafer orientation are very similar to the P_b centers found at (111) interfaces. The P_{b1} centers, in contrast, are found to comprise of completely different levels in energy. Lenahan *et al.* [20] gives an estimation of the energy distribution of P_{b0} and P_{b1} centers, as seen in Figure 3.6.

The number of P_{b1} centers is assumed to be lower than that of P_{b0} centers. Still, around the peak levels of the P_{b1} centers a small change in the Fermi-level can have a significant impact on the charge state because of these P_{b1} centers and their narrow distribution.

(a) E'_s (b) E'_{74}

Figure 3.7: Possible configurations of an E' oxide trapping center as found in amorphous SiO_2. It comprises of an unpaired electron localized on a silicon atom which is back-bonded to either three oxygen atoms (a) or two oxygen and one hydrogen atoms (b).

3.1.3 Effects of Trap Centers

Although the density of silicon dangling bonds is rather small, they still lead to worse device characteristics. Therefore, it is of major interest to reduce the amount of electrically active interface traps as much as possible.

It has been shown that by subjecting the MOS device to a post metallization anneal or to a forming gas anneal, the density of electrically active traps can be reduced dramatically. As a model for the passivation process the reactions

$$P_b + H_2 \rightarrow P_bH + H\,, \tag{3.3}$$
$$P_b + H \rightarrow P_bH\,, \tag{3.4}$$

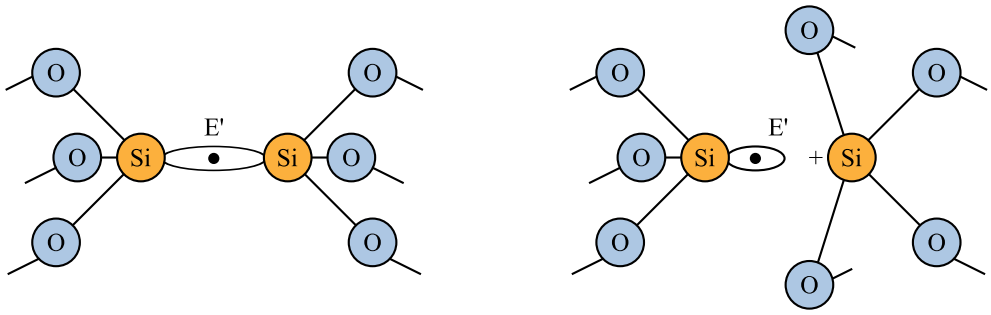
have been suggested in the literature [32–35].

The electrical activity of the P_b center is eliminated by a hydrogen atom as it shifts the energy levels of the trap out of the silicon band-gap [16], and thus passivates the defect.

During the negative bias temperature (NBT) stress exactly these hydrogen passivated silicon bonds can be broken. Broken bonds have, again, electrically active states in the band-gap which lead to additional charge on the interface and thus to degradation of important transistor parameters, as described in detail in Chapter 6.

3.2 E' Centers

Another defect, found in SiO_2 close to the Si/SiO_2 interface, is the E' center oxide defect. The exact atomic configuration is still debated [17, 27, 36–38] but possible configurations of E' centers involve an unpaired electron localized on a silicon atom back-bonded to three oxygen atoms or two oxygen atoms and one hydrogen atom, as shown in Figure 3.7. Variants proposed include a second silicon atom impacting the defect as depicted in Figure 3.8.



(a) E'_δ (b) E'_γ

Figure 3.8: Other possible configurations of E' centers. They comprise of a second silicon atom impacting the defect.

After very harsh negative bias temperature stressing conditions, those defects can be detected using resonance measurements [18, 22]. SDR measurements are only capable of sensing near interface oxide defects, as only those can interact with the interface traps and alter the recombination current. Therefore it is not clear whether more E' defects are located deeper in the SiO_2.

E' centers have levels near the middle of the SiO_2 band-gap where they allow hole capture from the silicon inversion layer. It was shown experimentally [36] that after hole capture an E' center can crack molecular hydrogen. This process might be important for the dissociation of silicon-hydrogen bonds at the Si/SiO_2 interface and therefore for the effect of NBTI.

Chapter 4

Characterization of Interfaces

THE excellent quality of the Si/SiO_2 interface is one of the main reasons for the success of the modern VLSI MOS technology. Therefore, it is very important to characterize the quality of this interface as precisely as possible.

4.1 Charge Pumping Method

The charge pumping method has shown to be a very reliable and also precise method allowing the in-depth analysis of the interface, directly in the MOSFET device. Additionally it only requires basic equipment and is relatively easy to set up.

The effect has been first reported by Brugler and Jespers in 1969 [39]. They reported a net DC substrate current when applying periodic pulses to the gate of a MOS transistor, while keeping source and drain grounded. The current was found to be proportional to the gate area and the frequency of the applied gate pulses. It was flowing in the opposite direction of the leakage current of the source and drain to substrate diodes. They showed that the current originates from recombination of minority and majority carriers at traps at the Si/SiO_2 interface. Therefore, the method can be used for measuring the interface trap density in MOSFETs for the evaluation of MOSFET degradation. The major breakthrough for the charge pumping method was the thorough investigation and correct explanation of the method, applied directly to MOSFET structures by Groeseneken *et al.* in 1984 [40].

4.1.1 Experimental Setup

The basic experimental setup for the charge pumping method can be seen in Figure 4.1 for an n-channel MOSFET. The source and drain to substrate diodes are reverse biased. The gate is pulsed between accumulation and inversion conditions while the charge pumping current is measured at the substrate. This current flows in the opposite direction of the source and drain to substrate diode leakage currents.

In the accumulation phase majority carriers, holes in case of an n-channel MOSFET, flood

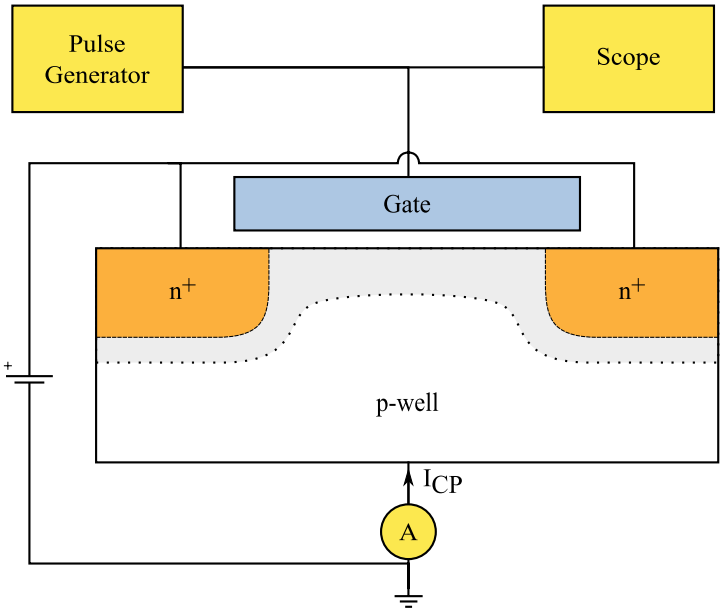


Figure 4.1: Basic experimental setup for the charge pumping measurement. The source to substrate and drain to substrate diodes are typically slightly reverse biased while the gate is pulsed between inversion and accumulation conditions. The substrate current is measured as the charge pumping current I_{CP}.

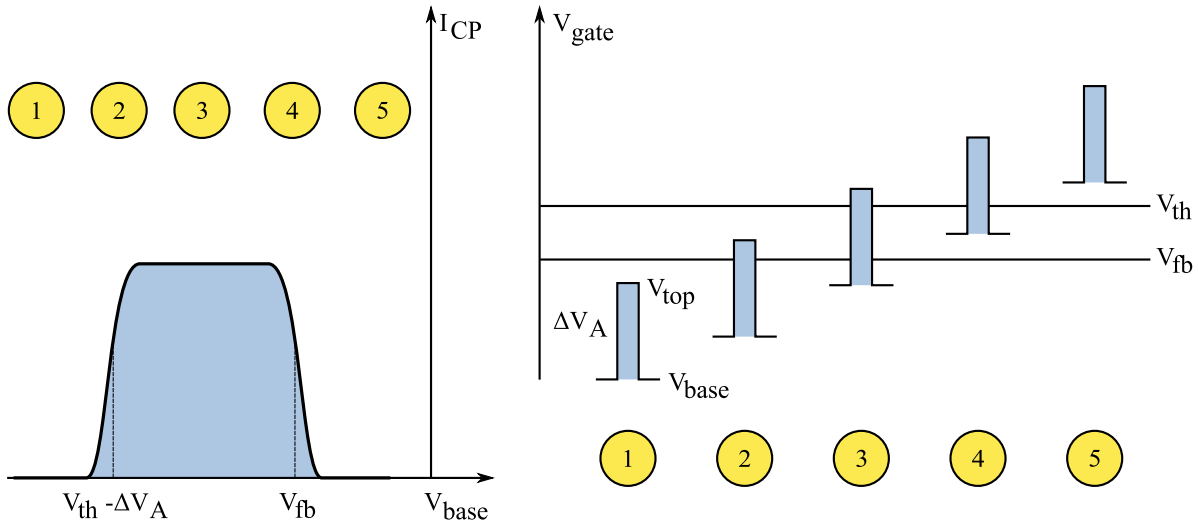


Figure 4.2: Base level sweep during a charge pumping measurement. As the base and also the top level of the gate pulse pass the flat-band and threshold voltage levels of the transistor, five different regimes can be distinguished.

the channel area and some of them become trapped in interface traps. When the gate pulse drives the transistor into inversion, the majority carriers leave the channel and move back to the substrate. Some trapped carriers with energies close to the valence band can be de-trapped through thermal emission before the channel becomes flooded by electrons and also move back to the substrate. The rest of the trapped holes recombines with channel electrons and leads to a net current. The same process occurs when the transistor is driven from inversion back to accumulation, with opposite carrier types.

The base level of the gate voltage pulse is swept to drive the MOSFET from accumulation to inversion. When the amplitude of the pulse is larger than the difference of threshold voltage and flat-band voltage, $\Delta V_{\mathrm{A}} > V_{\mathrm{th}} - V_{\mathrm{fb}}$, then five different regimes are observed as sketched in Figure 4.2. Regime 3, where the largest amount of traps in the band-gap is scanned, is the most important one. It can be described by a current model (Section 4.1.2) to calculate the interface trap density. The base level sweep charge pumping method was first proposed by Elliot [41] and the different regimes are governed by the following mechanisms:

Regime 1: The whole pulse is below the flat-band voltage and the substrate is in permanent accumulation. The interface traps are permanently filled with holes and therefore no recombination current is measured.

Regime 2: The top of the pulse reaches the region between the flat-band and the threshold voltages. In this phase the interface is moved from accumulation into strong depletion up to weak inversion. Here, the charge pumping current increases and the base voltage is around threshold voltage minus the pulse height. It could be assumed that the shape of the rising I_{CP} in this regime is determined by the recombination process in weak inversion. It has been shown, though, that other mechanisms may have an important influence. These can be surface potential fluctuations because of spatially non-uniformly distributed oxide charges [42, 43], acceptor and donor traps [43], or variations in the proximity of the source and drain regions. Also the modulation of the effective gate area by the gate voltage might influence the rising charge pumping current.

Regime 3: The base level voltage is below the flat-band voltage, and the top level of the pulse is above the threshold voltage, $V_{\mathrm{base}} < V_{\mathrm{fb}} < V_{\mathrm{th}} < V_{\mathrm{top}}$. In this regime the charge pumping pulse sweeps the substrate in the channel area from accumulation to complete inversion. At each time the transistor is pulsed from accumulation to inversion or back. The fast interface traps are filled with holes, or electrons, respectively, which then recombine with the opposite carrier type leading to a net current measurable as I_{CP}. In this regime the current has the highest magnitude.

Regime 4: The base level is between the flat-band and threshold voltages. The transistor only reaches weak accumulation, the interface traps are mainly negatively charged and are no longer flooded with holes, thus recombination is reduced and the charge pumping current goes

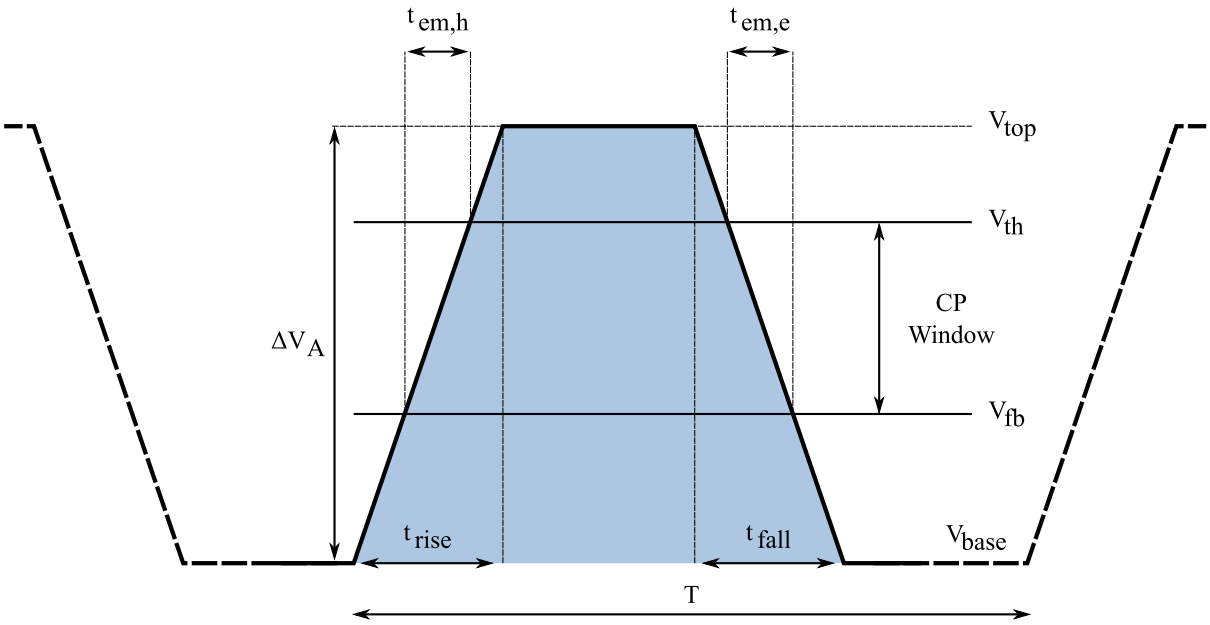


Figure 4.3: Charge pumping signal applied to the gate contact. The signal is characterized by the rise and fall times, t_{rise} and t_{fall}, and the amplitude ΔV_{A}. For the emission of electrons and holes from the traps only the window between the flat-band voltage V_{fb} and the threshold voltage V_{th} is significant.

down. The same surface potential fluctuation and gate area modulation effects as in Regime 2 can influence the characteristic of I_{CP}.

Regime 5: The transistor is completely in inversion during the whole pulse. The traps are filled with electrons and no holes reach the channel at any time. The measured substrate current only consists of the source and drain leakage currents.

4.1.2 Charge Pumping Current Model

The first comprehensive model for the charge pumping current was proposed by Groeseneken *et al.* in 1984 [40]. This model has been developed to capture the maximum charge pumping current which is obtained in Regime 3. Here the gate voltage pulse sweeps from below the flat-band voltage to above the threshold voltage. Therefore the substrate is driven from accumulation to inversion and back.

Only the fast interface traps situated between the two energy levels $E_{\text{em,e}}$ and $E_{\text{em,h}}$ in the

band-gap of the semiconductor can contribute to the charge pumping current,

$$E_{\mathrm{em,e}} = E_{\mathrm{i}} - \mathrm{k_B}T \ln\left(v_{\mathrm{th}}\sigma_{\mathrm{e}} n_{\mathrm{i}} t_{\mathrm{em,e}} + \exp\left(\frac{E_{\mathrm{i}} - E_{\mathrm{f,inv}}}{\mathrm{k_B}T}\right)\right), \tag{4.1}$$

$$E_{\mathrm{em,h}} = E_{\mathrm{i}} - \mathrm{k_B}T \ln\left(v_{\mathrm{th}}\sigma_{\mathrm{h}} n_{\mathrm{i}} t_{\mathrm{em,h}} + \exp\left(\frac{E_{\mathrm{i}} - E_{\mathrm{f,acc}}}{\mathrm{k_B}T}\right)\right), \tag{4.2}$$

where E_{i} is the intrinsic energy, $E_{\mathrm{f,inv}}$ and $E_{\mathrm{f,acc}}$ are the Fermi energies in inversion and accumulation, v_{th} is the thermal velocity, $\sigma_{\mathrm{e,h}}$ the capture cross sections of the traps, and n_{i} is the intrinsic carrier concentration. Traps outside this band cannot contribute to the current as the trapped charge becomes de-trapped instantly through thermal emission when the Fermi-level moves beyond the trap level, for trapped electrons, or above, for trapped holes.

The emission times $t_{\mathrm{em,e}}$ and $t_{\mathrm{em,h}}$ for electrons and holes can be calculated from the fall and rise times, t_{f} and t_{r}, as

$$t_{\mathrm{em,e}} = \frac{|V_{\mathrm{fb}} - V_{\mathrm{th}}|}{|\Delta V_{\mathrm{A}}|} t_{\mathrm{f}}\,, \tag{4.3}$$

$$t_{\mathrm{em,h}} = \frac{|V_{\mathrm{fb}} - V_{\mathrm{th}}|}{|\Delta V_{\mathrm{A}}|} t_{\mathrm{r}}\,, \tag{4.4}$$

and are illustrated in Figure 4.3. These are the times available for the emission of carriers from the fast traps.

The net CP current measured at the substrate can be obtained as

$$I_{\mathrm{CP}} = \overline{D_{\mathrm{it}}} 2qf A_{\mathrm{G}} \mathrm{k_B}T \ln\left(v_{\mathrm{th}} n_{\mathrm{i}} \sqrt{\sigma_{\mathrm{e}}\sigma_{\mathrm{h}}} \sqrt{t_{\mathrm{em,e}} t_{\mathrm{em,h}}}\right), \tag{4.5}$$

where f is the frequency and A_{G} the gate area.

The charge pumping current is directly related to the mean interface trap density $\overline{D_{\mathrm{it}}}$ in the channel, the size of the Si/SiO_2 interface channel area, the frequency f, and the pulse shape characterized by its rise and fall times. This makes the charge pumping method a perfect tool for the characterization of interface degradation.

4.1.3 Numerical Simulation

The device simulator Minimos-NT [44] is used for numerical analysis of the charge pumping effect. For each V_{base} of interest a transient simulation of the gate pulse is performed. The resulting currents can then be plotted versus the base voltage to obtain the typical charge pumping current I_{CP} versus V_{base} plot.

The device under test was a conventional n-channel MOSFET structure (Figure 4.4). The gate length, measured from source-substrate to substrata-drain junctions is 0.6 μm, the device width 100 μm, and the gate oxide thickness is 12 nm.

4.1.3.1 D_{it} Variation

The first simulation gives a comparison of the analytical current model (4.5) to numerical simulations using Minimos-NT. Here, the interface trap density N_{it} has been varied from $10^{10}\mathrm{cm}^{-2}$ up

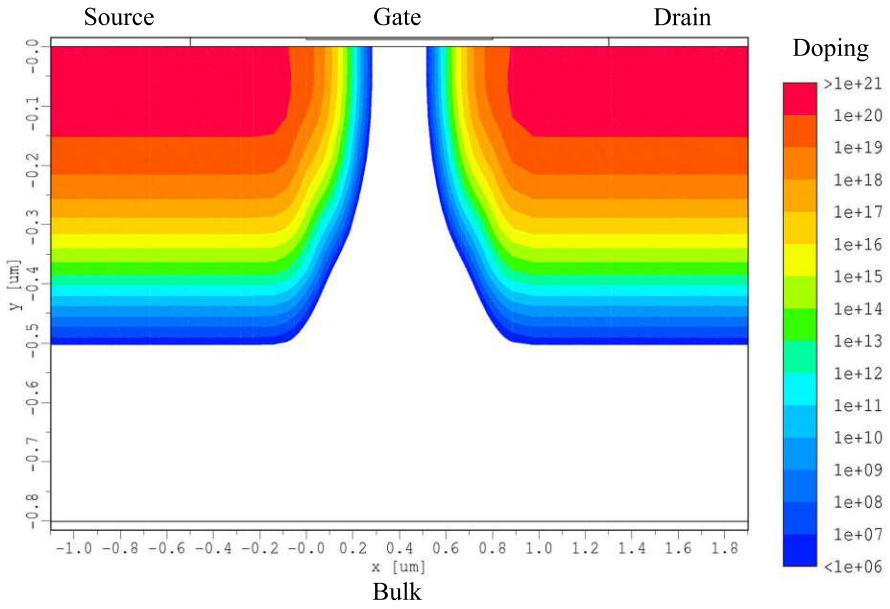


Figure 4.4: Device structure used for numerical charge pumping simulations. The n-channel MOSFET has a channel length of 0.6 μm (junction to junction) and the gate oxide thickness is 12 nm.

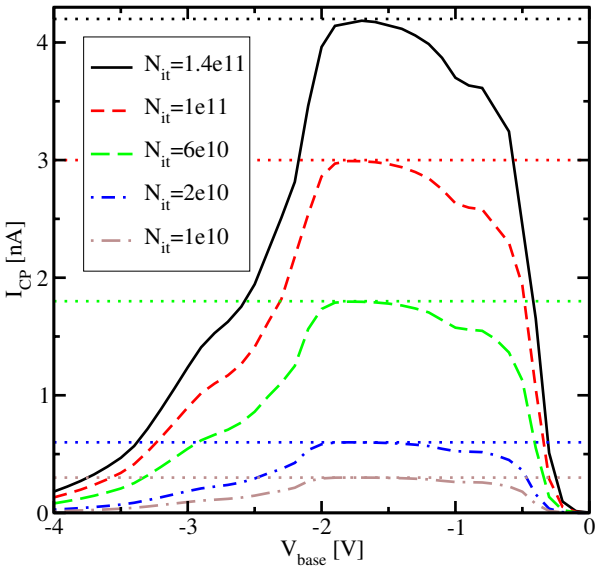


Figure 4.5: Charge pumping simulation results. The interface traps are of acceptor type, equally distributed in the band-gap, the concentration has been varied from $10^{10}\,\mathrm{cm}^{-2}$ up to $1.4 \times 10^{11}\,\mathrm{cm}^{-2}$, $\Delta V_\mathrm{A} = 3\,\mathrm{V}$, and there are no fixed interface charges. The dotted lines give the results from the current model obtaining excellent agreement with the numerical simulation results.

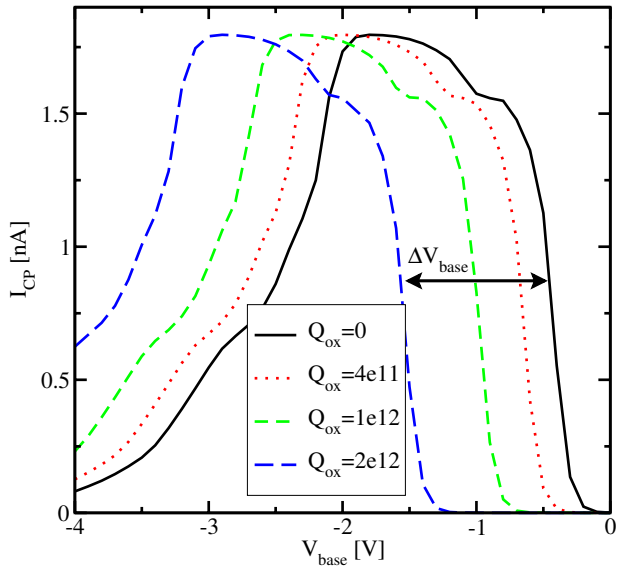


Figure 4.6: Charge pumping simulation to extract threshold voltage shifts. The interface trap density is constant at $6 \times 10^{10}\,\mathrm{cm}^{-2}$. Close to the Si/SiO_2 interface, fixed positive oxide charges Q_{ox} are generated. From the charge pumping signal and the resulting voltage shift ΔV_{base} the density of fixed interface charges can be calculated.

to $1.4 \times 10^{11}\mathrm{cm}^{-2}$, the pulse height is $\Delta V_A = 3\,\mathrm{V}$, and the rise and fall times are $t_r = t_f = 0.1\mu\mathrm{s}$. Figure 4.5 gives the resulting CP curves and also the analytic approximation using (4.5). The agreement is excellent with the peak values of Regime 3.

4.1.3.2 Q_{ox} Variation

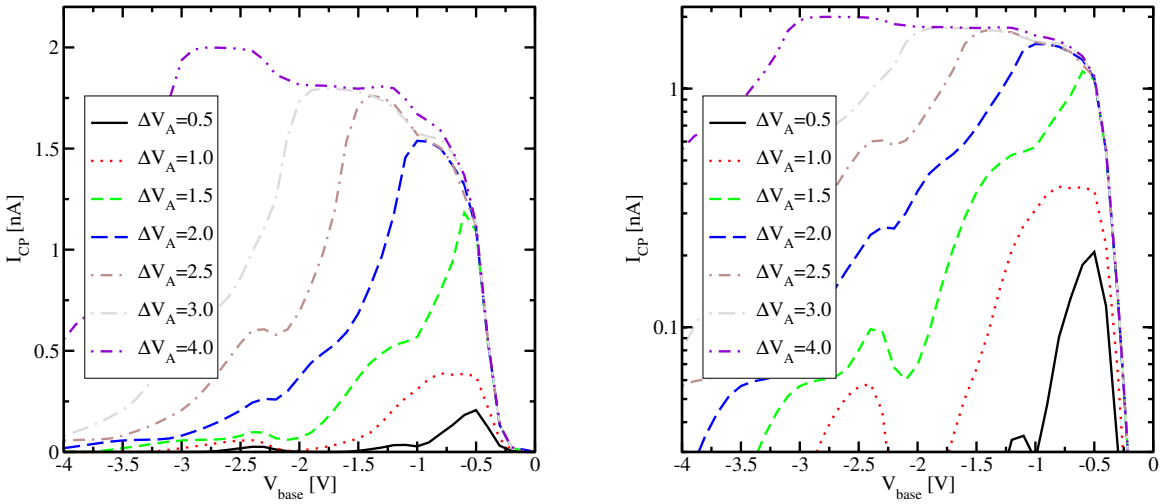
The charge pumping method is very well suited for the quantitative investigation of fixed oxide charges. Figure 4.6 shows the charge pumping currents of a MOSFET device with constant density of interface traps but varying amount of fixed charges at the Si/SiO_2 interface. The oxide charges Q_{ox} are assumed to be located directly at the interface. The resulting ΔV_{base} in the numerical simulations for an interface charge density $Q_{ox} = q_0 \cdot 2 \times 10^{12}\,\mathrm{cm}^{-2}$ is $\Delta V_{base} = -1.11\,\mathrm{V}$. This result is in excellent agreement with the analytical equation (2.31) presented in Section 2.2.1 and using the approximation for the plate capacitor

$$C_{ox} = \varepsilon \frac{A}{l}\,, \tag{4.6}$$

where A is the area and l is the dielectric thickness, predicting $\Delta V_{base} = -1.114\,\mathrm{V}$.

4.1.3.3 Pulse Amplitude Variation

Experimenting with the pulse height ΔV_A nicely illustrates the strongly increasing charge pumping current until ΔV_A surmounts $V_{th} - V_{fb}$ (Figure 4.7).



(a) Linear scale

(b) Logarithmic scale

Figure 4.7: Variation of the gate pulse height ΔV_A. The interface trap density is constant at $6 \times 10^{10}\,\mathrm{cm}^{-2}$. I_CP strongly increases until $\Delta V_\mathrm{A} > V_\mathrm{th} - V_\mathrm{fb}$. The marginal increase at higher ΔV_A is due to the decreasing emission times.

It can be seen that for increasing gate amplitudes the saturation current still slightly increases for $\Delta V_\mathrm{A} > V_\mathrm{th} - V_\mathrm{fb}$. This effect is because the transistor's depletion region is swept faster resulting in decreasing emission times for electrons and holes, (4.3) and (4.4). Therefore, the thermal emission is reduced and more traps can contribute to the charge pumping current.

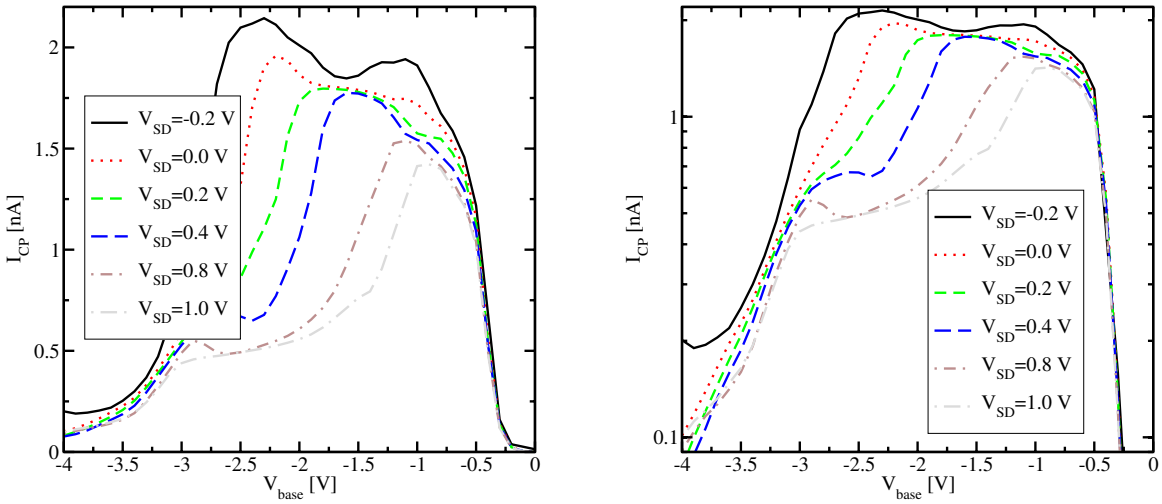
4.1.3.4 Reverse Bias

By increasing the reverse bias the charge pumping current is decreased, as shown in Figure 4.8. This reduction is due to two effects:

1. The body effect. It leads to an enlargement of the space charge region and therefore to an increase of the threshold voltage. This increased V_th in turn increases the emission times for electrons and holes, (4.3) and (4.4), and therefore to a reduction of I_CP as found from (4.5).

2. Due to the increase of the space charge regions around source and drain during accumulation the effective channel gate area is reduced. Therefore, less interface traps can contribute to the charge pumping current.

For $V_\mathrm{sd} < -0.2\,\mathrm{V}$ the charge pumping current is dominated by the source/drain to substrate diode current and cannot be used for the evaluation of interface traps.



(a) Linear scale

(b) Logarithmic scale

Figure 4.8: Variation of the source/drain to substrate bias. The interface trap density is kept constant at $6 \times 10^{10}\,\mathrm{cm}^{-2}$. At high bias the source/drain to substrate diodes are turned off. At lower source and drain voltages the source/drain to substrate diodes are turned on and the diode currents dominates the charge pumping current (below $V_{\mathrm{sd}} = -0.2\,\mathrm{V}$).

4.1.3.5 Temperature Dependence

As the thermal emission process is, as its name already suggests, strongly temperature dependent, so is the charge pumping current. At higher temperatures more carriers can be de-trapped before recombining with the opposite carrier type and I_{CP} is reduced (Figure 4.9).

4.2 DCIV Method

Another experimental method for the investigation of the interface quality is the direct-current current-voltage (DCIV) method. It can be used to determine both the interface trap density N_{it} and the amount of fixed oxide charges Q_{ox}.

The experimental set-up is illustrated in Figure 4.10. The source and drain to substrate diodes are slightly forward biased, typically with around $V_{\mathrm{e}} = V_{\mathrm{s}} = V_{\mathrm{d}} \approx 0.33\,\mathrm{V}$. The gate bias is swept from inversion to slight accumulation. During this sweep the substrate current I_{b} is measured. The substrate current originates from recombination of carriers at the Si/SiO_2 interface.

For each degradation level there is a clear peak of I_{DCIV} at a certain $V_{\mathrm{g}} = V_{\mathrm{peak}}$, as shown in Figure 4.11. The peak height above the base line, ΔI_{DCIV} is approximately proportional to the

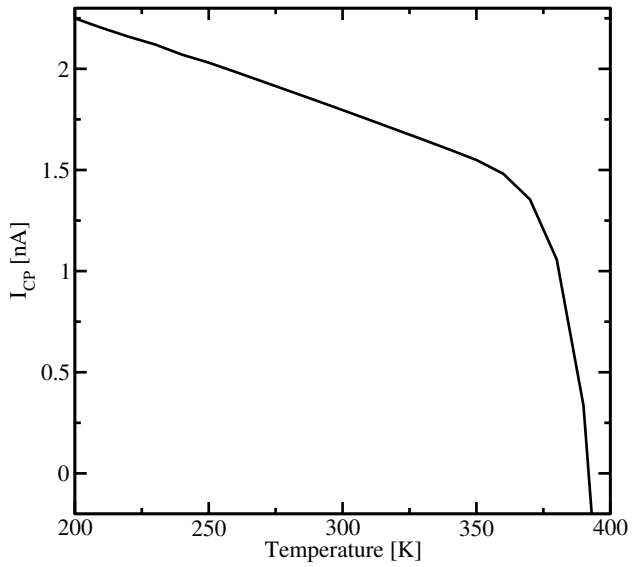


Figure 4.9: Simulated temperature dependence of the charge pumping current. Base voltage and interface trap density are kept constant ($V_{\text{base}} = -1.8\,\text{V}$ and $N_{\text{it}} = 6 \times 10^{10}\,\text{cm}^{-2}$), only the temperature is steadily increased. Higher temperatures support the thermal emission of trapped carriers and therefore reduce the measured charge pumping current.

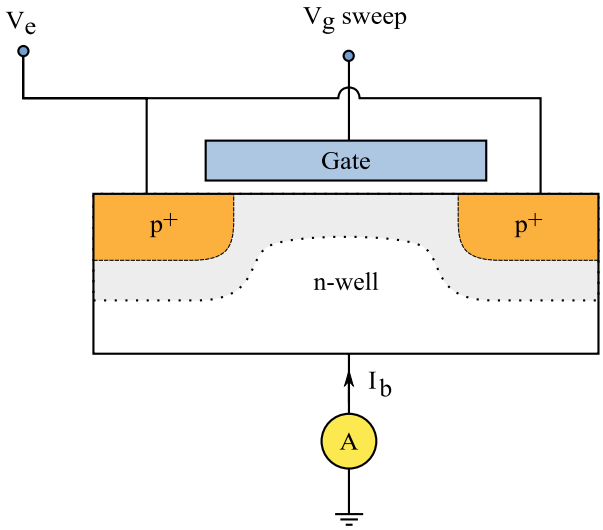


Figure 4.10: Configuration of the DCIV measurement. The source and drain to substrate diodes are slightly forward biased. As the gate to substrate voltage V_g is swept, a current at the substrate contact can be measured.

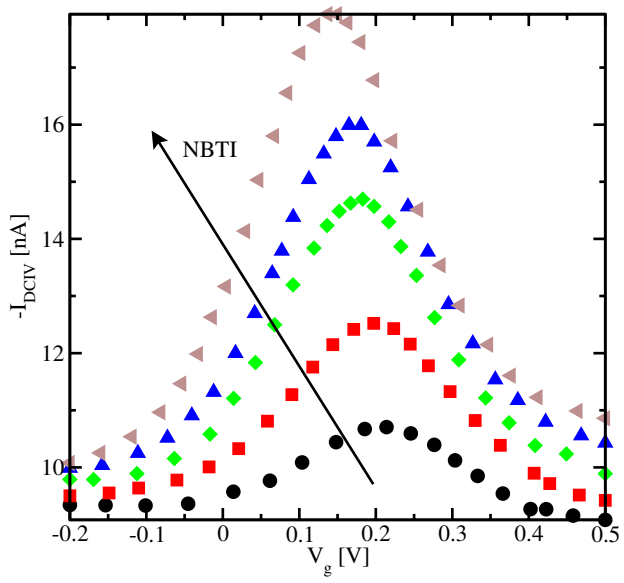


Figure 4.11: DCIV currents after different stress times. With increasing interface degradation the maximum I_{DCIV} increases. The shift of the peak position to lower voltage indicates fixed oxide or interface charges. Data are from Zhu *et al.* [45].

effective interface trap density N_{it} as [46, 47]

$$\Delta I_{\mathrm{DCIV}} = \frac{1}{2}\mathrm{q}_0 n_{\mathrm{i}} \sigma v_{\mathrm{th}} D_{\mathrm{it}} A \mathrm{q}_0 |V_{\mathrm{e}}| \exp\left(\frac{\mathrm{q}_0 |V_{\mathrm{e}}|}{2\mathrm{k_B} T}\right) . \tag{4.7}$$

Here, q_0 is the elementary charge, n_{i} the intrinsic carrier concentration, σ the geometric mean of the capture cross sections for electrons and holes, equal to $\sqrt{\sigma_{\mathrm{e}}\sigma_{\mathrm{h}}}$, v_{th} the thermal velocity, A the effective gate area, V_{e} the forward bias applied to the source-substrate and drain-substrate junctions, $\mathrm{k_B}$ Boltzmann's constant, and T the temperature.

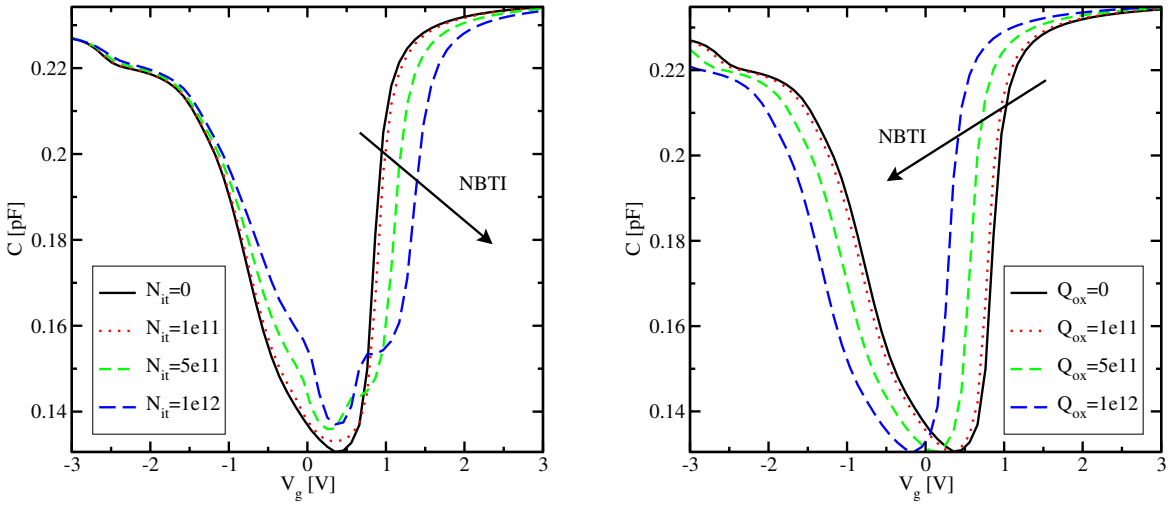
To determine Q_{ox}, the oxide charge density, the peak position of the DCIV curve can be used. As described in Section 2.2.1, the peak position is directly proportional to the amount of oxide charges and can be calculated using Equation 2.31.

4.3 Capacitance-Voltage Characteristics

The capacitance-voltage (CV) measurement can be used to determine the interface quality [48]. During the CV measurement the device is swept from accumulation to inversion while constantly measuring the capacitance as

$$C = \frac{\Delta Q}{\Delta V} . \tag{4.8}$$

Figure 4.12(a) depicts the simulated CV characteristics of a p-type MOSFET at different degradation levels. The density-of-states is assumed constant. It is a low frequency simulation



(a) Interface traps

(b) Oxide charges

Figure 4.12: Capacitance-voltage simulation demonstrating the influence of (a) interface traps N_{it} and (b) fixed oxide charges Q_{ox}. While interface traps lead to deformation of the CV curve, the introduction of oxide charges shifts the flat-band voltage and therefore the CV characteristics.

($f = 1\,\mathrm{kHz}$) where the substrate carriers are in thermal equilibrium and follow the gate voltage change. An increase in the interface trap density N_{it} is followed by deformation of the characteristics as the traps can dynamically be charged and discharged. An introduction of fixed oxide charges shifts the flat-band voltage and therefore the whole CV characteristics (Figure 4.12(b)) as predicted by Equation 2.31.

Increasing the frequency changes the CV results [5]. As the interface traps are not fast enough and can therefore not follow the gate voltage variation their impact on the CV characteristics is reduced as shown in Figure 4.13. For very high frequencies the inversion layer cannot build up fast enough and the capacitance is drastically reduced.

Of major importance in the CV results is the trap density-of-states. Figure 4.14 gives the simulation results for different trap energy distributions. The exponential distribution has most traps close to the band edge. These traps can be filled and emptied even close to flat-band voltage conditions more easily and lead to an increase in the capacitance.

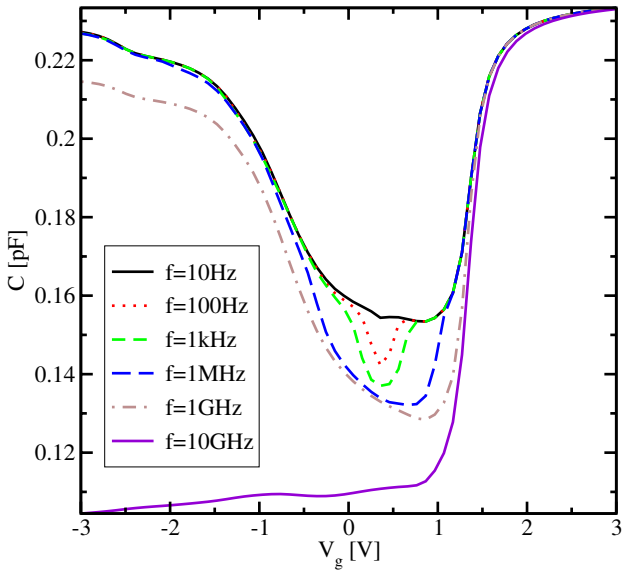


Figure 4.13: Frequency variation. With increasing frequency the traps cannot follow the signal anymore and are not visible in th CV characteristics. For the very high frequency $f = 10\,\mathrm{GHz}$ also the inversion layer charges cannot follow the signal and drastically reduce the capacitance in the inversion regime.

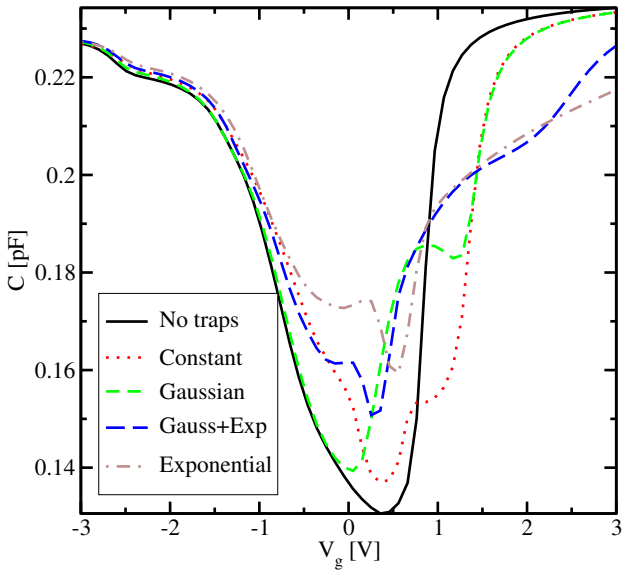


Figure 4.14: Defect density-of-states variation. The interface trap density and the frequency are constant at $N_{\mathrm{it}} = 10^{12}\,\mathrm{cm}^{-2}$ and $f = 1\,\mathrm{kHz}$. The variation of the DOS changes the capacitance as the charge trapping behavior is different.

Chapter 5

Dielectric Degradation and Failure Mechanisms

ONE elementary aspect of semiconductor devices is their reliable performance for a determined period of time. There are many extrinsic and intrinsic degradation mechanisms which can lead to device and circuit failure. Due to aggressive scaling of device geometries, increasing electric fields across the device, and the utilization of new materials many degradation mechanisms which were formerly of minor importance can now lead to severe reliability concerns.

In this chapter a range of degradation mechanisms affecting the gate dielectric is described and their implication on device failure is discussed. Special focus is put upon mechanisms concerning the silicon/silicon-dioxide interface as its properties play a very important role for negative bias temperature instability. The effect of trap-assisted tunneling and its implication on the threshold voltage due to trapped carriers, similar to charge trapping in NBTI (Section 6.6), are examined in detail.

5.1 Hot Carrier Degradation

In general "hot carriers" are particles that attain a very high kinetic energy from being accelerated by a high electric field. These energetic carriers can be injected into normally forbidden regions of the device, as the gate dielectric, where they can get trapped or cause interface states to be generated. These defects then lead to threshold voltage shifts and transconductance degradation of MOS devices. To avoid, or at least minimize hot carrier degradation, several device design modification can be made. These are for example a larger channel length, double diffusion of source and drain, and graded drain junctions to name a few.

For the injection of hot carriers into the dielectric there are four distinguished injection mechanisms [49]: channel hot-electron (CHE) injection, drain avalanche hot-carrier (DAHC) injection, secondary generated hot-electron (SGHE) injection, and substrate hot-electron (SHE) injection.

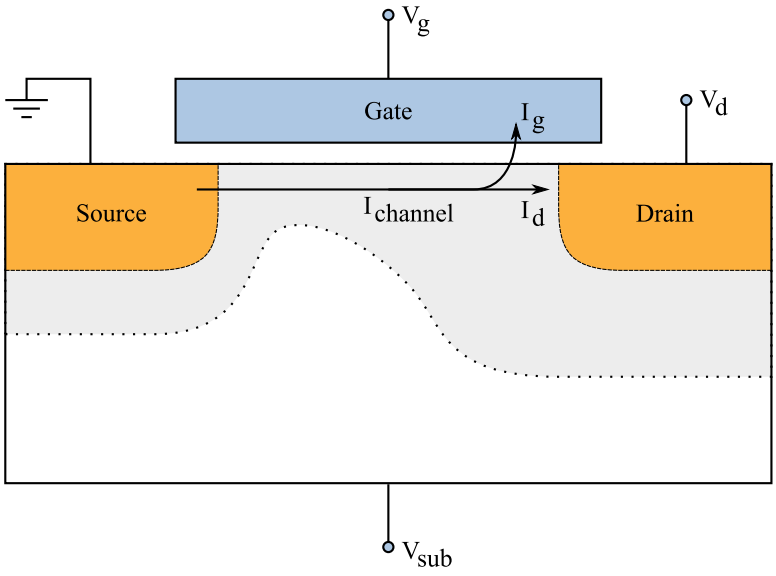


Figure 5.1: At CHE stress conditions, $V_{gs} \approx V_{ds}$, "lucky electrons" gain enough energy while drifting across the channel and are injected into the dielectric causing a gate current, interface and oxide degradation. The oxide electric field is the same as at the drain pulsed voltage measurement technique (Section 6.3.3).

5.1.1 Channel Hot-Electron Injection

Where the gate voltage is approximately equal to the drain voltage the channel hot-electron (CHE) injection is at a maximum. Figure 5.1 depicts this conditions where "lucky electrons" [50, 51] which are attracted by the high gate voltage gain enough energy from the electric field across the channel to surmount the Si/SiO_2 barrier at the drain end of the channel. The CHE injection can be measured as a gate current which has a maximum around $V_{gs} \approx V_{ds}$. For lower gate voltages the field does not attract electrons to the gate electrode anymore. For higher drain voltages the electric field at the drain leads to avalanche multiplication due to impact ionization (Section 2.3.4) and hot electrons *and* hot holes are injected reducing the measured gate current.

5.1.2 Drain Avalanche Hot-Carrier Injection

At stress conditions with high V_{ds} and lower V_{gs} the drain avalanche hot-carrier (DAHC) injection is significant [49] (Figure 5.2). It is caused by the injection of holes and electrons generated by avalanche multiplication as described in Section 2.3.4. The carriers gain their energy due to a high electric field in the drain region. Measurement of DAHC is difficult as both carrier types are injected simultaneously. Additionally some of the generated carriers lead to a bulk current.

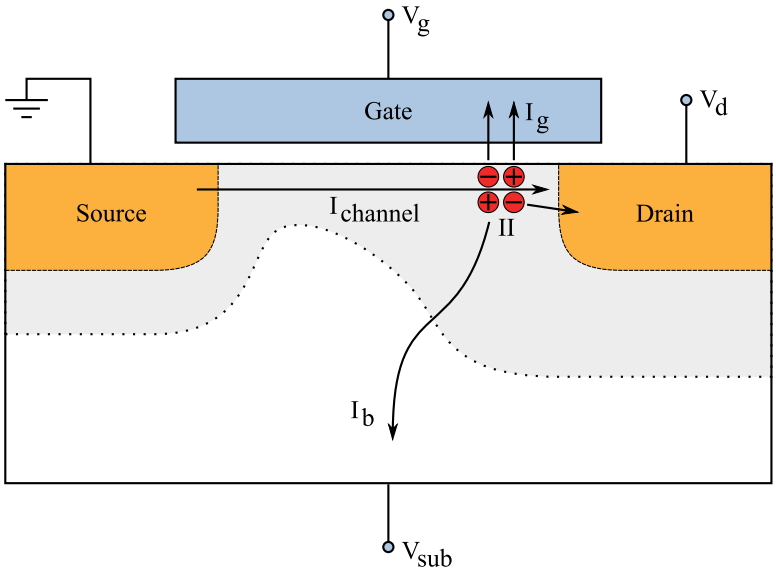


Figure 5.2: Hot carriers lead to impact ionization generating electron-hole pairs. In the drain avalanche hot-carrier injection regime hot electrons and hot holes are injected into the dielectric. Additionally some of the carriers form a bulk current.

5.1.3 Secondarily Generated Hot-Electron Injection

The origin of secondarily generated hot-electron (SGHE) injection is firstly a photo induced generation process (Section 2.3.1) [52, 53]. Photons are generated in the high field region near the drain and induce a generation process for electron-hole pairs. The second effect is the avalanche multiplication (Section 2.3.4) near the drain region leading to the injection of both, electrons and holes into the dielectric. The injection process is supported by the substrate bias which is additionally driving carriers to the interface.

5.1.4 Substrate Hot-Electron/Hole Injection

The substrate hot-electron or hot-hole (SHE/SHH) injection is a result of a high positive or negative bias at the substrate backside V_{sub} [49]. This leads to carriers in the substrate being driven to the Si/SiO_2 interface which gain further kinetic energy in the surface depletion region. The substrate carriers are either generated by optical generation or by electrical injection from a buried p-n junction. These carriers can eventually overcome the energy barrier at the interface and are injected into the oxide.

SHE/SHH is often used to investigate the insulator qualities and for reliability tests [54, 55]. The advantage is that the energetic carriers at the interface are uniformly distributed along the channel in contrast to the injection mechanisms described above, where the maximum of the injection is near the drain end of the channel. Therefore the stress conditions at the interface are well defined. As the surface potential ϕ_{surf} is pinned, the oxide field is solely controlled by

the gate voltage. The potential drop in the substrate is determined by the substrate voltage V_{sub}. Therefore, several important conditions like the oxide field, the carrier energy, and the current intensity can be adjusted independently.

5.2 Dielectric Wearout and Breakdown

The proper operation of MOS transistors relies on the insulating properties of the dielectric layer. Each dielectric material has a maximum electric field it can intrinsically sustain (dielectric strength). Applying a higher field leads to breakdown which destroys the insulating properties and allows current to flow. At lower electric fields the insulator can wear-out after some time and finally break down completely. This time-dependent dielectric breakdown (TDDB) is a very important reliability aspect for MOS structures [56].

For dielectric breakdown two scenarios are distinguished, extrinsic and intrinsic breakdown. Extrinsic breakdown is due to defects in the dielectric which can be introduced during different processing steps while intrinsic breakdown is because of the nature of the dielectric itself. It occurs at a certain electric field, defining the dielectric strength. As the insulating layers are getting thinner the probability of an external defect and therefore the probability of an extrinsic failure is decreasing. Hence intrinsic failure is the most likely problem for today's dielectrics.

5.2.1 Measurement of Breakdown

For thicker oxides, above 6 nm, the time to breakdown can be defined as the moment when the oxide layer abruptly loses its insulating properties. This is called a hard breakdown and can be detected as a large jump in the current-time or voltage-time curve, depending whether constant voltage or constant current stress has been applied [57].

For very thin dielectrics and lower stress voltages, especially in large area capacitors, a soft breakdown is possible, which is very difficult to detect in the output curve as it changes only slightly. A significant change can be found in the gate current noise after soft breakdown. Therefore a breakdown detection method based on measurement of the increasing noise has been proposed [58].

5.2.2 Models

Several models have been proposed to characterize dielectric wearout and breakdown. Their target is to explain the mechanisms involved in dielectric degradation.

5.2.2.1 Anode Hole Injection Model

The anode hole injection model [59, 60] suggests that electrons that tunnel through the dielectric can, when reaching the anode, transfer their excess energy to an electron deep in the valence band of the anode. This electron is promoted to the lower edge of the conduction band leaving a hot hole behind. These hot holes can then tunnel back into the dielectric generating oxide

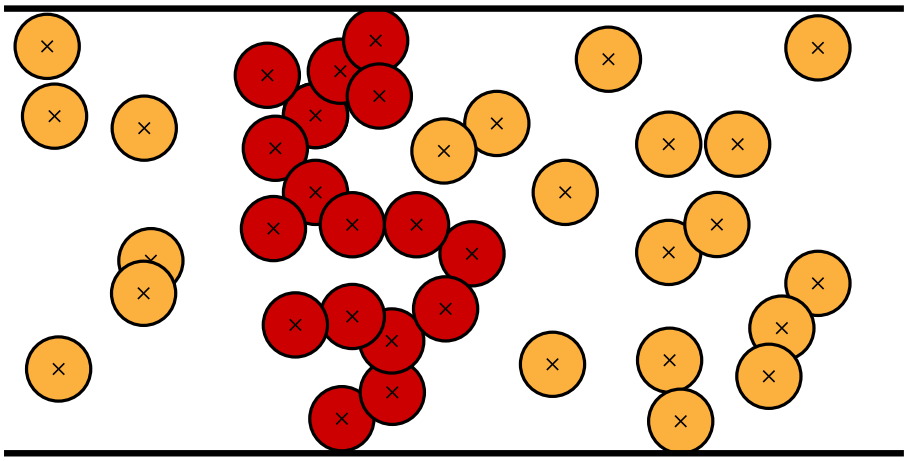


Figure 5.3: Schematic illustration of the percolation model for intrinsic oxide breakdown simulations. It is based on trap generation and conduction via these traps. A breakdown path is formed by spheres connecting anode and cathode.

traps. The model suggests that these traps increase the current density through the dielectric due to trap-assisted tunneling (Section 5.3.2) leading to a runaway process followed by dielectric breakdown.

According to this model breakdown occurs when a critical hole fluence $Q_{\mathrm{h,crit}}$ is reached. It has been shown that for a 11 nm oxide $Q_{\mathrm{h,crit}}$ is about 0.1 C/cm^2 [59]. However, it has been observed that this number decreases for thinner oxides [60, 61] which cannot be explained by the model. There is also no physical explanation for the link between the breakdown event and the critical hole fluence.

5.2.2.2 Electron Trap Generation Model

The electron trap generation model claims a critical electron trap density (CETD) generated during stress to be necessary to trigger oxide breakdown. It presents the breakdown event as the formation of a conducting path of traps connecting the cathode to the anode [62–65].

Experiments show, though, that at thinner dielectrics the CETD for breakdown is decreased [60, 61, 66]. This is in contradiction with the electron trap generation model which predicts a critical trap density independent of the oxide thickness.

5.2.2.3 Percolation Model

The breakdown model proposed by Degraeve *et al.* [66] is based on the percolation concept. It links both the anode hole injection model and the electron trap generation model. Degraeve *et al.* showed that the electron trap generation is correlated with the generated hole fluence and not with the oxide field and thickness. This implies that either the correlation is casual, i.e. holes are involved in the generation of traps, or that holes and traps are both related to

a common parameter, which, for example, can be the energy of the injected electrons at the anode.

Figure 5.3 illustrates the percolation concept. During stress electron traps are generated in the dielectric at random positions. Around these traps, spheres with a constant radius r are defined. These spheres give the conductive area around the traps. The radius is the only free parameter of the model. As soon as cathode and anode are connected by overlapping spheres, a conducting path is formed and the breakdown condition is reached. The critical electron trap density (CETD) can now be calculated as the number of traps divided by the volume of the simulated dielectric.

The percolation model predicts that for decreasing oxide thickness the CETD decreases, and due to the direct correlation also $Q_{\mathrm{h,crit}}$. This is in perfect agreement with experimental findings [60, 61, 66]. So thinner oxides have a weaker "hole fluence immunity". This is an intrinsic property of the breakdown mechanism and is due to the lower number of traps necessary to form a conducting path through the dielectric. The second finding from the percolation model is the enhanced statistical spread of oxide breakdown for thinner oxides. As for thinner oxides only a few traps are necessary to form a conducting path, the statistical spread on the trap density is larger.

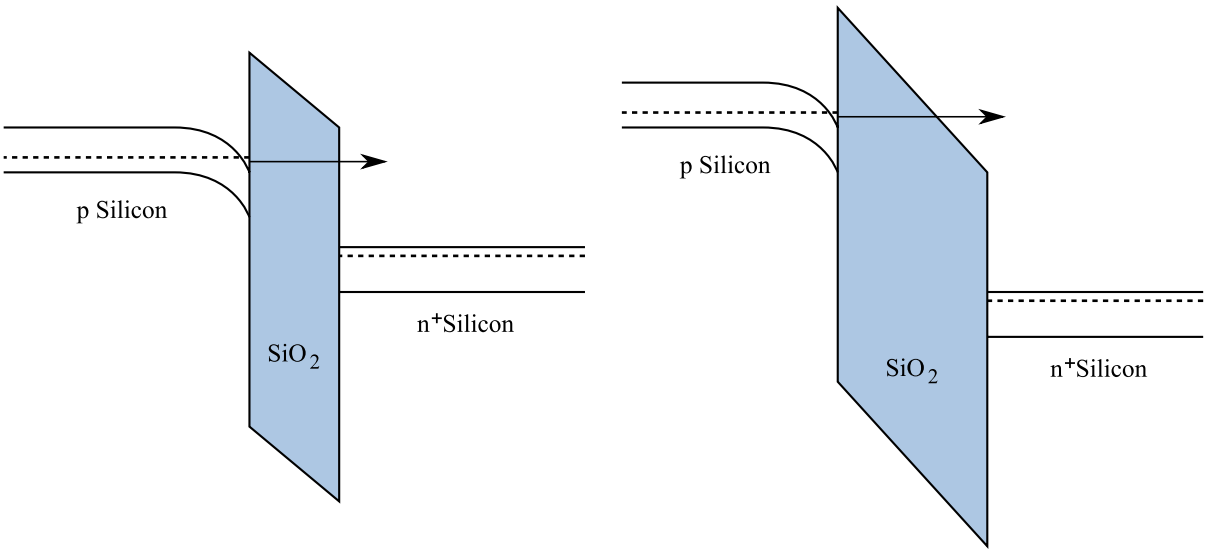
5.3 Quantum Mechanical Tunneling

Due to constant downscaling of gate-dielectric thicknesses in modern MOS devices the effect of tunneling has drastically gained relevance. Quantum mechanical tunneling describes the transition of carriers through a classically forbidden energy state. This can be an electron tunneling from the semiconductor through a dielectric, which represents an energy barrier, to the gate contact of an MOS structure. Even if the energy barrier is higher than the electron energy, there is quantum mechanically a finite probability of this transition. The reason lies in the wavelike behavior of particles on the quantum scale where the wave function describes the probability of finding an electron at a certain position in space. As the wave function penetrates the barrier and can even extend to the other side, quantum mechanics predict a non-zero probability for an electron to be on the other side.

Figure 5.4 depicts the energy band diagrams for two tunneling mechanisms in an MOS structure consisting of a p-type bulk silicon, SiO_2 dielectric, and a n^+ polycrystalline silicon gate.

5.3.1 Direct Tunneling

In Figure 5.4(a) the energy band conditions for the direct tunneling regime are shown. Here, the electrons from the inverted silicon surface can tunnel directly through the forbidden energy barrier formed by the dielectric layer to the poly-gate. Direct tunneling is strongly gaining significance when the dielectric layer gets thinner.



(a) Direct tunneling (b) Fowler-Nordheim tunneling

Figure 5.4: Energy band diagrams showing tunneling mechanisms in an MOS structure. (a) At direct tunneling conditions the electron tunnels through the whole SiO_2 energy barrier. (b) At Fowler-Nordheim conditions the electron tunnels through a part of the barrier to the conduction band of the insulator. From there it can flow to the anode.

A common approach to model the tunneling current is the Tsu-Esaki formula [67]

$$J = \frac{4\pi q_0 m}{h^3} \int_{E_{\min,0}}^{E_{\max}} TC(E_x, m_{\text{diel}}) N(E_x)\, \mathrm{d}E_x \ , \tag{5.1}$$

where m is the electron mass in silicon and m_{diel} the electron mass in the dielectric [68]. $TC(E_x, m_{\text{diel}})$ is the transmission coefficient and $N(E_x)$ the supply function which is given as

$$N(E_x) = \int_0^{\infty} (f_1(E) - f_2(E))\, \mathrm{d}E_x \, . \tag{5.2}$$

Here, f_1 and f_2 denote the energy distribution functions near the interfaces [69].

5.3.1.1 Fowler-Nordheim Tunneling

The energy band conditions for Fowler-Nordheim tunneling, which is a special case of direct tunneling, are depicted in Figure 5.4(b). The electrons do not tunnel directly to the other side of the barrier. Instead they tunnel from the silicon inversion layer to the conduction band of the

SiO_2 layer from where they are transported to the gate contact. The Fowler-Nordheim regime is significant for thicker dielectrics and sufficiently high electric fields.

A thorough investigation of modeling and simulation of tunneling mechanisms can be found in the thesis of Gehring [69].

5.3.2 Trap-Assisted Tunneling

As the reduction of the applied voltages does not keep up with the miniaturization of actual devices the electric fields across dielectric layers are constantly increasing. Especially for non-volatile memory cells high electric fields are necessary in order to achieve quick write and erase cycles. Due to the repeated high-field stress, defects can arise in the dielectric leading to tunneling currents, even at low fields. This stress-induced leakage current (SILC) [70, 71] plays a major role in the determination of the retention times of non-volatile memory cells.

There are many approaches to model trap-assisted tunneling (TAT). One is to model the defect assisted tunneling process including a single trap. For each trap position tunneling from the cathode to the trap and further to the anode is considered. From this the trap occupancy function can be calculated which is used to compute the tunneling current [72]. This single-TAT approach works very well for slightly degraded devices or devices with thin gate dielectrics. For thicker dielectrics with a high defect density it is reasonable to assume that also the interaction of two or more traps in the tunneling process takes place [73].

For the modeling of multi-trap assisted tunneling (multi-TAT) approaches like a two-trap process [74, 75] or a multi-trap process considering hopping of carriers between distinct defects [76] have been presented. Recently, anomalous charge loss in floating-gate memory cells has been reported [73], where a two-trap model was used to reproduce the measured data.

For correct modeling of such highly degraded devices a new approach is proposed as part of this work. It rigorously computes TAT current assisted by multiple traps [77, 78]. In this model hopping processes between all oxide defects are taken into account. In addition the filling of oxide traps with carriers, leading to space charge in the oxide and therefore to a shift of the threshold voltage, is accounted for. This shift can lead to circuit failure as the timing parameters of the device are degraded.

The model for the simulation of SILC in highly degraded devices is based on inelastic, phonon-assisted tunneling [79] with the tunneling-rate proposed by Herrmann and Schenk [72]. These approaches are extended for modeling the interaction of multiple traps in the tunneling process.

5.3.2.1 Inelastic Phonon-Assisted Tunneling

The defect-assisted tunneling process of an electron from the cathode to the anode via a trap is considered as a two-step process. Electrons are captured from the cathode, relax to the energy level of the trap by emitting one or more phonons with the energy $\hbar\omega$, and are then emitted to the anode. This process is inelastic as the electron energy is not conserved during the tunneling process. Figure 5.5 depicts this process including the phonon emission.

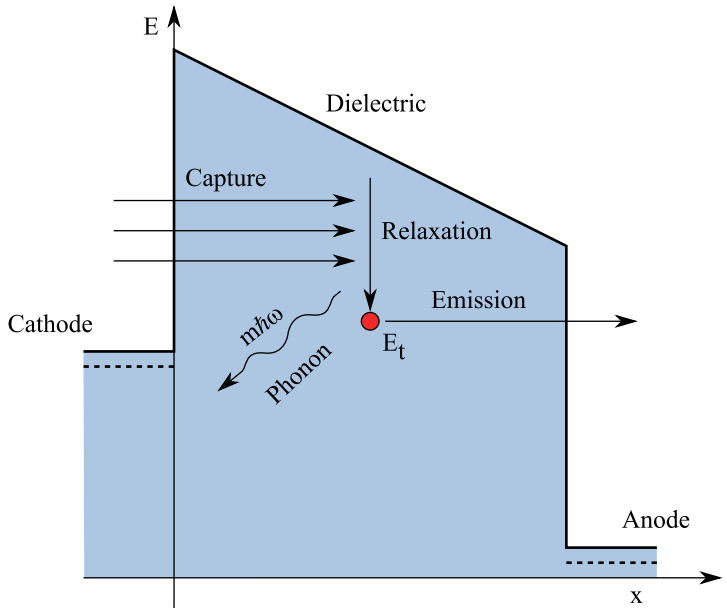


Figure 5.5: Inelastic tunneling process including a sole trap. The excess energy of the tunneling electron is released by means of phonon emission.

5.3.2.2 Single-Trap Assisted Tunneling

In the single-trap assisted tunneling approach the tunneling current for each trap is calculated separately. The total tunneling current is afterwards superimposed from the individual contributions. Although the interaction between neighboring traps is of importance in highly degraded dielectrics (Section 5.3.2.3), for less degraded devices this approach can be used [72, 80].

The tunneling current density is modeled as the sum of capture or emission rates of each trap, R_{c_i} or R_{e_i}, which are equal in the stationary case, $R_{c_i} = R_{e_i} = R_i$, multiplied by the trap cross section Δx_i,

$$J = q \sum_i R_i \Delta x_i \,. \tag{5.3}$$

The energetic position of the trap, E_T, with respect to the conduction band edge determines the trap cross section [81]

$$\Delta x_i = \frac{\hbar}{\sqrt{2 m_{diel} E_T}} \left(\frac{4\pi}{3}\right)^{1/3} , \tag{5.4}$$

where m_{diel} denotes the electron mass in the dielectric, which is used as a model calibration parameter.

The single-TAT and the multi-TAT models differ in the way the capture and emission rates are calculated. When only single-trap processes are considered (see Figure 5.5) the rates are

determined by [72]

$$R_{c_i} = \tau_{c_i}^{-1} N_{t_i} (1 - f_{t_i}) \,, \tag{5.5}$$

$$R_{e_i} = \tau_{e_i}^{-1} N_{t_i} f_{t_i} \,. \tag{5.6}$$

Here, τ_{c_i} and τ_{e_i} are the capture and emission times and N_{t_i} is the trap concentration. As the capture and emission rates are equal in the stationary case, the trap occupancy f_{t_i} can be directly calculated as

$$f_{t_i} = \frac{\tau_{c_i}^{-1}}{\tau_{c_i}^{-1} + \tau_{e_i}^{-1}} \,, \tag{5.7}$$

where the inverse capture and emission times are defined as as [72, 82]

$$\tau_{c_i}^{-1} = \int_{E_0}^{\infty} g_C(E) c_n(E) T_C(E) f_C(E) \, dE \,, \tag{5.8}$$

$$\tau_{e_i}^{-1} = \int_{E_0}^{\infty} g_A(E) e_n(E) T_A(E) (1 - f_A(E)) \, dE \,. \tag{5.9}$$

In these expressions, $g_C(E)$ and $g_A(E)$ denote the density of states in the cathode and anode, respectively, T_C and T_A the transmission coefficients from the cathode and the anode, and the symbols c_n and e_n are computed as

$$c_n(E) = c_0 \sum_m L_m \delta(E - E_m) \,, \tag{5.10}$$

$$e_n(E) = c_0 \exp\left(-\frac{E - E_t}{k_B T_L}\right) \sum_m L_m \delta(E - E_m) \,, \tag{5.11}$$

with

$$c_0 = \frac{(4\pi)^2 \Delta x_i^3 (\hbar \Theta_0)^3}{\hbar E_{g,SiO2}} \,, \tag{5.12}$$

$$\hbar \Theta_0 = \left(\frac{q_0^2 \hbar^2 F^2}{2 \, m_{diel}}\right)^{1/3} \,. \tag{5.13}$$

The summation index m gives the number of discrete phonon emissions, E_m is the phonon energy, and L_m is the multiphonon transition probability [72]. The symbols f_C and f_A represent the Fermi-distributions, F the electric field in the dielectric, and $E_{g,SiO2}$ the band gap of SiO_2. The transmission coefficients were evaluated by a numerical WKB method which yields reasonable accuracy for single-layer dielectrics. This model has been used in a more or less similar form by various authors [79, 80, 82].

5.3.2.3 Multi-Trap Assisted Tunneling

For highly degraded devices the isolated calculation for each trap is not sufficient anymore [74]. Anomalous charge loss in memory cells has been observed and was explained by conduction

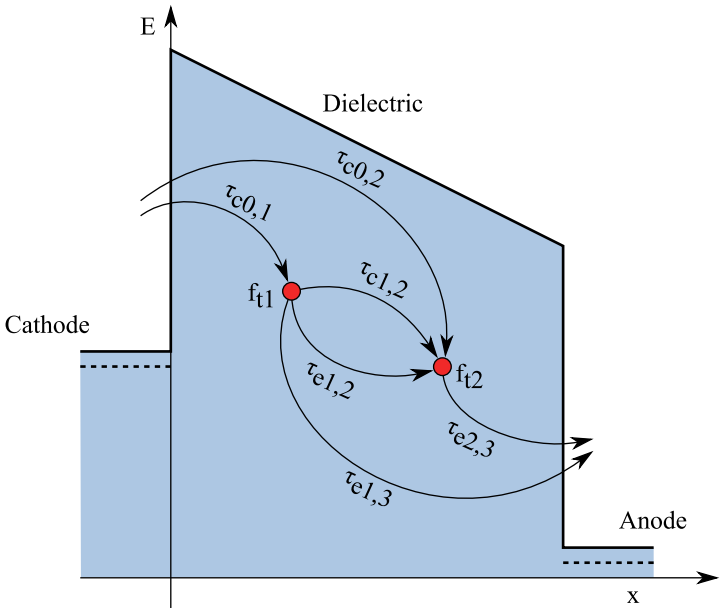


Figure 5.6: Multi-trap assisted tunneling process. The tunneling rate R_i of a specific trap is determined by all capture and emission times to and from the trap.

through a second trap [74]. The single-trap model can be extended for this case, and the rate equations become (see Figure 5.6)

$$\overbrace{\tau_{\mathrm{c}_{0,1}}^{-1} N_{\mathrm{t}_1}(1-f_{\mathrm{t}_1})}^{R_{\mathrm{c}_1}} - \overbrace{(\tau_{\mathrm{e}_{1,2}}^{-1} N_{\mathrm{t}_1} f_{\mathrm{t}_1}(1-f_{\mathrm{t}_2}) + \tau_{\mathrm{e}_{1,3}}^{-1} N_{\mathrm{t}_1} f_{\mathrm{t}_1})}^{R_{\mathrm{e}_1}} = 0\,, \tag{5.14}$$

$$\underbrace{\tau_{\mathrm{c}_{0,2}}^{-1} N_{\mathrm{t}_2}(1-f_{\mathrm{t}_2}) + \tau_{\mathrm{c}_{1,2}}^{-1} N_{\mathrm{t}_2} f_{\mathrm{t}_1}(1-f_{\mathrm{t}_2})}_{R_{\mathrm{c}_2}} - \underbrace{\tau_{\mathrm{e}_{2,3}}^{-1} N_{\mathrm{t}_2} f_{\mathrm{t}_2}}_{R_{\mathrm{e}_2}} = 0\,, \tag{5.15}$$

where instantaneous transitions between occupied and free traps are assumed.

For thicker dielectrics it is quite reasonable to assume that an arbitrary number of traps assist in the conduction process. We therefore extend the model to n traps where the capture and emission rates are evaluated as

$$R_{\mathrm{c}_k} = \sum_{i=0}^{k-1} \tau_{\mathrm{c}_{i,k}}^{-1} N_{\mathrm{t}_k} f_{\mathrm{t}_i}(1-f_{\mathrm{t}_k})\,, \tag{5.16}$$

$$R_{\mathrm{e}_k} = \sum_{i=k+1}^{n+1} \tau_{\mathrm{e}_{k,i}}^{-1} N_{\mathrm{t}_k} f_{\mathrm{t}_k}(1-f_{\mathrm{t}_i})\,. \tag{5.17}$$

The values for f_{t_0} and $f_{\mathrm{t}_{n+1}}$, which are the trap occupation probabilities at the cathode and the anode, are set to 1 and 0, respectively. This way the cathode acts as a perfect electron source and the anode as an electron sink.

From the capture and emission rates the following equation system can be set up

$$\sum_{i=0}^{k-1} \tau_{\mathrm{c}_{i,k}}^{-1} N_{\mathrm{t}_k} f_{\mathrm{t}_i}(1-f_{\mathrm{t}_k}) = \sum_{i=k+1}^{n+1} \tau_{\mathrm{e}_{k,i}}^{-1} N_{\mathrm{t}_k} f_{\mathrm{t}_k}(1-f_{\mathrm{t}_i})\,, \quad k=1,2,\ldots n\,, \tag{5.18}$$

from which the values of all trap occupation probabilities have to be calculated. This is performed using a Newton method with the cost function F_k for a trap at position k

$$F_k(\vec{f}) = \sum_{i=0}^{k-1} \tau_{\mathrm{c}_{i,k}}^{-1} f_{\mathrm{t}_i}(1-f_{\mathrm{t}_k}) - \sum_{j=k+1}^{n} \tau_{\mathrm{e}_{k,j}}^{-1} f_{\mathrm{t}_k}(1-f_{\mathrm{t}_j}) = 0\,, \quad 1 \le k \le n\,, \tag{5.19}$$

and the values of the derivatives in the Jacobian matrix

$$\frac{\partial F_i}{\partial f_{\mathrm{t}_j}} = \begin{cases} \tau_{\mathrm{e}_{i,j}}^{-1} f_{\mathrm{t}_i} & \text{for } i<j\,, \\ -\sum_{k=0}^{i-1} \tau_{\mathrm{c}_{k,i}}^{-1} f_{\mathrm{t}_k} - \sum_{k=i+1}^{n} \tau_{\mathrm{e}_{i,k}}^{-1}(1-f_{\mathrm{t}_k}) & \text{for } i=j\,, \\ \tau_{\mathrm{c}_{j,i}}^{-1}(1-f_{\mathrm{t}_i}) & \text{for } i>j\,. \end{cases} \tag{5.20}$$

A typical number of unknowns of the equation system is 15. Depending on the dielectric thickness and trap energy the result does not improve when further increasing this number. The computational effort remains negligible compared to the total device simulation time. The multi-trap assisted tunneling current density can then be obtained from the capture or emission rates

$$J = q \sum_i \left[\sum_{j=0}^{i-1} \tau_{\mathrm{c}_{j,i}}^{-1} N_{\mathrm{t}_i} f_{\mathrm{t}_j}(1-f_{\mathrm{t}_i}) \right] \Delta x_i\,. \tag{5.21}$$

5.3.2.4 Influence on the Threshold Voltage

By coupling this model to the semiconductor device equations (Section 2.1) the simulation of the effect of charged defects on the threshold voltage of memory devices is possible. The space charge density in the dielectric is calculated from the trap occupation function as

$$\rho_{\mathrm{TAT}}(x) = \mathrm{q} f_{\mathrm{t}}(x) N_{\mathrm{t}}(x)\,, \tag{5.22}$$

and added to Poisson's equation as described in Section 2.1.2 with

$$\rho_{\mathrm{TAT}} = \rho_n\,. \tag{5.23}$$

The trapped electrons shift the threshold voltage to more positive voltages and can lead to circuit failure.

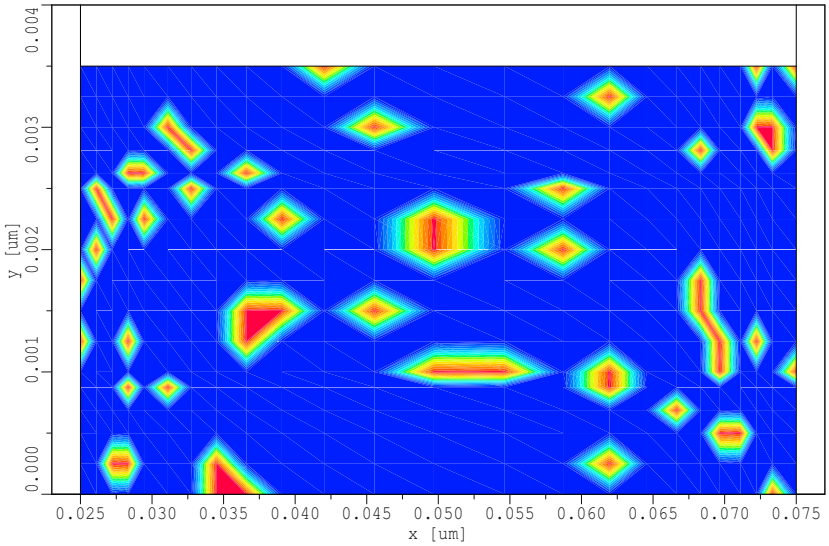


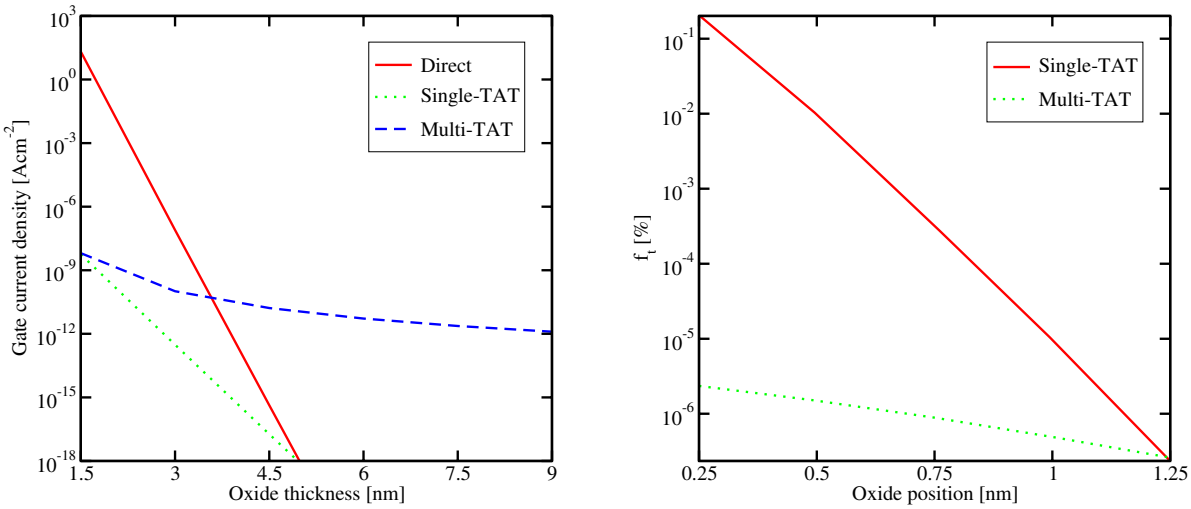
Figure 5.7: The formation of traps randomly distributed across the oxide of an MOS structure as simulated using Minimos-NT. Each trap contributes to additional tunneling current.

5.3.3 Evaluation of the New Model

The implementation of these models into the device- and circuit-simulator Minimos-NT [44] allows the two- and three-dimensional study of single- and multi-trap assisted tunneling as well as direct tunneling mechanisms. Figure 5.7 gives an example of a highly degraded dielectric with traps randomly distributed across the oxide. The dielectric can now be used for the simulation of multi-trap assisted tunneling.

Figure 5.8(a) shows a comparison of SILC simulation results assuming three different tunneling mechanisms, namely direct tunneling, single-TAT, and multi-TAT. The models have been applied to a set of MOS transistors with gate dielectric thicknesses ranging from 1.5 nm up to 9 nm. The gate is biased at 1 V, source and drain are kept at 0 V. For both, the single-TAT and the multi-TAT simulations, the trap energy is set to 2.8 eV below the dielectric conduction band with a constant trap density of $9 \times 10^{17}\,\mathrm{cm}^{-3}$ across the oxide.

In the multi-trap simulation the tunneling current is several orders of magnitude higher than in the single-trap simulation. This is due to the fact that the multi-TAT current includes the single-TAT component as a limiting case. The multi-TAT model considers the capture and emission processes from the cathode and to the anode, respectively, and also the capture and emission processes involving all other trap centers. This leads to the comparably high multi-TAT component in devices with thicker oxides. It has to be considered, though, that this high current is mainly due to the assumption of uniformly distributed trap concentrations across the oxide,



(a) Oxide thickness variation

(b) Trap occupancy along the channel

Figure 5.8: (a) SILC simulations for a set of MOS devices at 1 V gate bias. The oxide was assumed to have a constant trap concentration. The multi-TAT current is always larger than the single-TAT current. (b) The trap occupancy f_t in the oxide of a 1.5 nm MOS transistor at 1 V gate bias.

which implies that the total amount of traps is much higher in thick dielectrics. The direct tunneling component loses importance for thicker dielectrics but dominates for thin dielectrics as found in logic CMOS devices. For devices with thicker oxides and higher trap densities multi-TAT processes become increasingly important.

Figure 5.8(b) depicts the resulting trap occupancy within the oxide. An MOS transistor with 1 V gate bias was simulated. It can be seen that the trap occupancy f_t is remarkably lower in the multi-TAT case. The reason is the higher probability for electrons to tunnel to one of the neighbor traps compared to tunneling to the anode as it is the only possibility in the single-TAT model.

Figure 5.9 outlines the threshold voltage V_{th} for different oxide thicknesses. The direct tunneling model, applying the commonly used Tsu-Esaki approach, does not account for the filling of traps in the oxide. Therefore the threshold voltage is not shifted compared to the simulation without a tunneling model. The new multi-TAT model predicts an increase in V_{th}. This higher threshold voltage is due to the filled and therefore negatively charged traps.

For thicker dielectrics with high defect density the inclusion of multiple traps is crucial for the simulation of both, quantum mechanical gate currents and the shift of threshold voltages. Thin dielectrics, on the other hand side, may be as well treated using the single-TAT model.

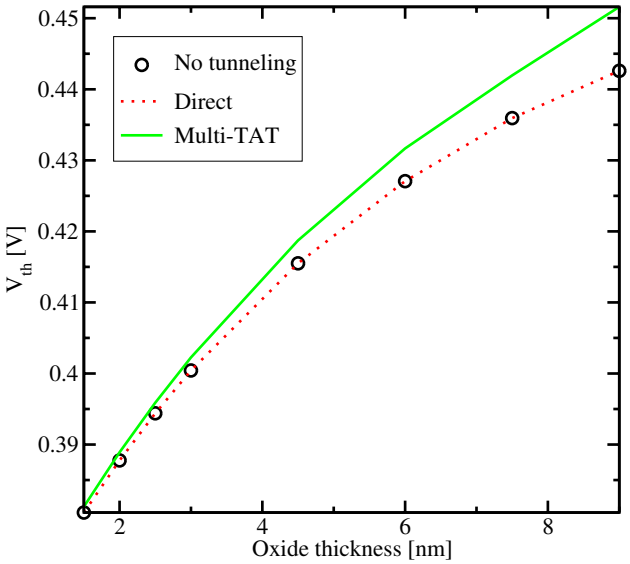


Figure 5.9: Comparison of the threshold voltage V_{th} of MOSFET structures with different oxide thicknesses. Thicker devices suffer from a higher shift due to the availability of more trapped charges.

Chapter 6

Negative Bias Temperature Instability

BIAS temperature instability (BTI) is a degradation phenomenon affecting mainly MOS field effect transistors. The highest impact is observed in p-channel MOSFETs which are stressed with negative gate voltages at elevated temperatures. The stress conditions for this negative bias temperature instability (NBTI) typically lie below 6 MV/cm for the gate oxide electric field and temperatures ranging between 100–300 °C. Higher electric fields can cause additional degradation due to hot carriers (Section 5.1) and should be avoided for the evaluation of NBTI. A very interesting aspect of device degradation caused by NBTI is its capability to anneal to a certain extend when the stress conditions are diminished.

The most important transistor parameters which degrade because of NBTI are:

- decreasing transconductance g_{m},
- decreasing linear drain current $I_{\mathrm{d,lin}}$ and saturation current $I_{\mathrm{d,sat}}$,
- decreasing channel mobility μ_{eff},
- decreasing subthreshold slope S,
- increasing off current I_{off}, and
- increasing absolute value of the threshold voltage V_{th}.

The consequence can be a reduced circuit switching speed as charging times for interconnect or load capacitances are increased or even circuit failures.

The effect of NBTI has already been reported 40 years ago [83], but gained much attention in recent years [84–87] due to modern semiconductor technologies. The following aspects have been found to lead to increasing susceptibility to NBTI:

- higher oxide electric fields due to oxide scaling,

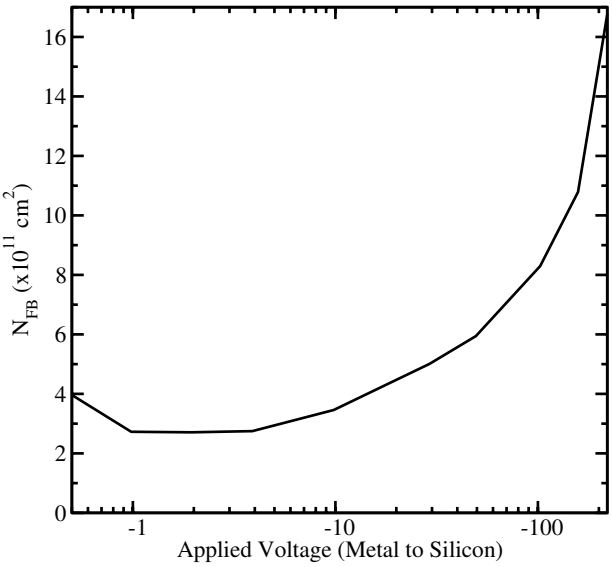


Figure 6.1: First report of the NBTI effect by Miura and Matukura in 1966 [83]. The graph shows the saturated value of the accumulated electron density at flat-band conditions N_{FB} for different stress voltages. The stress temperature was 300 °C.

- higher temperatures due to higher power dissipation,
- replacement of buried p-channel MOSFETs with surface devices,
- the introduction of CMOS elevating the importance of p-channel MOSFETS, and
- nitrided oxides with a higher permittivity.

In this chapter the observed characteristics of NBTI are outlined and the recent understanding of the involved physical degradation processes described. An important topic discussed here is the measurement and characterization of the level of degradation followed by state-of-the-art modeling approaches presented by different groups as found in literature. A new model is presented which was implemented in the numerical device simulator Minimos-NT [44] and gives excellent agreement with measurement data as shown in Chapter 7.

6.1 First Report of NBTI

The effect of negative bias temperature instability has first been published by Miura and Matukura in 1966 [83]. The authors investigated a 300 nm SiO_2, thermally grown in a dry oxygen atmosphere, on an n-type silicon substrate. The metal contact was formed of aluminum. This MOS structure was stressed at a temperature of 300 °C at different gate voltages. The stress at each voltage was retained until the degradation saturated, which was then measured.

Figure 6.1 shows the measured figure of merit, N_{FB}. It represents the accumulated electron density in the silicon bulk close to the Si/SiO_2 interface at flat-band conditions. N_{FB} is extracted from the flat-band voltage which is obtained by capacitance-voltage measurements (Section 4.3).

At that time the effect was remarkable, as a major concern was the migration of mobile ions due to the electric field. For negative gate bias stress this would suggest positive ions from the oxide to be attracted by the negative gate electrode leading to a "more negative" dielectric and thus to a *decrease* of N_{FB} for p-channel MOSFETs. This effect can be seen in Figure 6.1 for gate voltages down to $-4\,\mathrm{V}$. At more negative stress voltages another degradation mechanism prevails, leading to positive charging of the dielectric and/or the Si/SiO_2 interface and therefore to an *increase* of N_{FB}.

Miura and Matukura proposed an electrochemical reaction under the influence of the strong electric field at the Si/SiO_2 interface. This reaction leads to positively charged oxygen vacancies in the SiO_2 film. As this mechanism proceeds at higher electric fields it dominates the ion migration process which saturates at large biases.

6.2 The NBTI Time Exponent

The degradation of transistors due to NBTI is often found to follow a power law in time over a wide range of decades as

$$\Delta V_{\mathrm{th}} = A \exp\left(\frac{E}{E_{\mathrm{ref}}}\right) \exp\left(-\frac{E_{\mathrm{a}}}{\mathrm{k_B}T}\right) t^n \,, \tag{6.1}$$

where the threshold voltage shift ΔV_{th} depends exponentially on the electric field stress E and on the time t with an exponent n. The symbol A is a pre-factor and E_{ref} the reference electric field. The temperature dependence is here modeled to follow Arrhenius' law, $\exp(-E_{\mathrm{a}}/(\mathrm{k_B}T))$, with the activation energy E_{a}.

The exact physical mechanism for the threshold voltage degradation is still not clear. Especially the introduction of the temperature dependence as Arrhenius' law is debated [88].

The most important value researchers try to gain from analytic formalisms as (6.1) is the capability to extrapolate the degradation for longer times and/or lower stress fields. Here, the evaluation of the parameters has to be as precise as possible. A small error in the time exponent n, for example, can lead to under- or overestimation of the product lifetime by several years.

6.3 Physical Mechanisms of NBTI

Although the effect of bias temperature instability has been reported more than 40 years ago by several groups [83, 89, 90] there is still much controversy about the physical mechanisms behind the degradation and the exact causes for BTI are not yet fully understood. However, broad agreement has been found that when MOSFETs are stressed with a constant gate voltage at an elevated temperature, positive charge builds up either at the Si/SiO_2 interface or in the gate oxide layer. This charge leads to degradation of the transistor parameters.

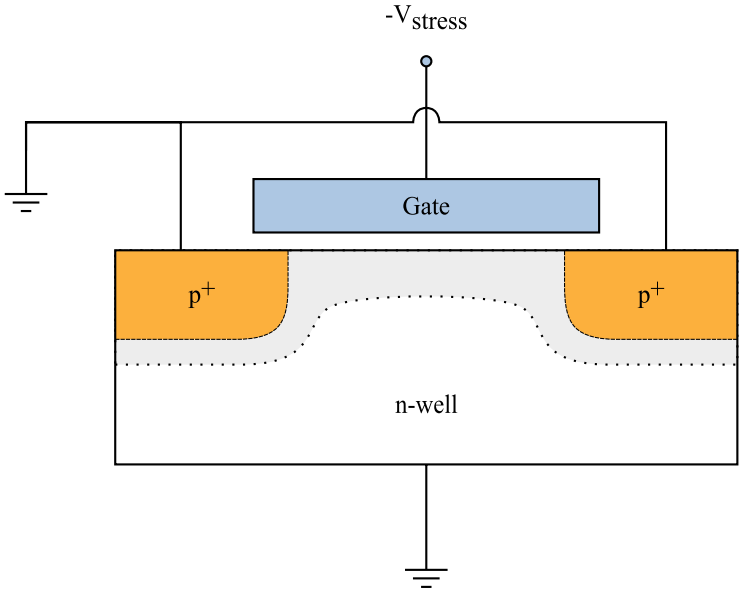


Figure 6.2: The typical set-up for NBTI investigation. The gate is negatively biased while the source, drain, and substrate contacts are grounded. These conditions are applied at elevated temperatures and for a certain period of time. Due to the symmetry of source and drain no channel hot carriers are generated.

One of the problems researchers face is that the exact transition from crystalline silicon to amorphous silicon-dioxide and the possible defects are not totally clear. Chapter 3 gives an overview of the resent research results concerning this interface and its defects.

In this section the proposed physical mechanisms that lead to BTI are presented and their validity discussed.

6.3.1 Basic Experimental Set-Up

The set-up for the observation of negative bias temperature (NBT) degradation of either a MOSFET or an MOS capacitor is schematically depicted in Figure 6.2. The substrate and in the case of the MOSFET also the source and drain contacts are grounded, while the gate is negatively biased. These bias conditions are applied at elevated temperatures, typically ranging between 100 and 200 °C, for a certain period of time.

To extract the level of degradation the stress is interrupted several times for short measurement cycles. During these measurement intervals typically the threshold voltage is measured performing a gate voltage sweep while measuring the current at the slightly forward biased drain contact. The interface trap density can be extracted using the charge pumping (Section 4.1), DCIV (Section 4.2), or CV (Section 4.3) methods.

This is the basic experimental set-up which was formerly used to obtain NBTI data found in literature. The problem with this method is the strong recovery of NBTI induced damage during

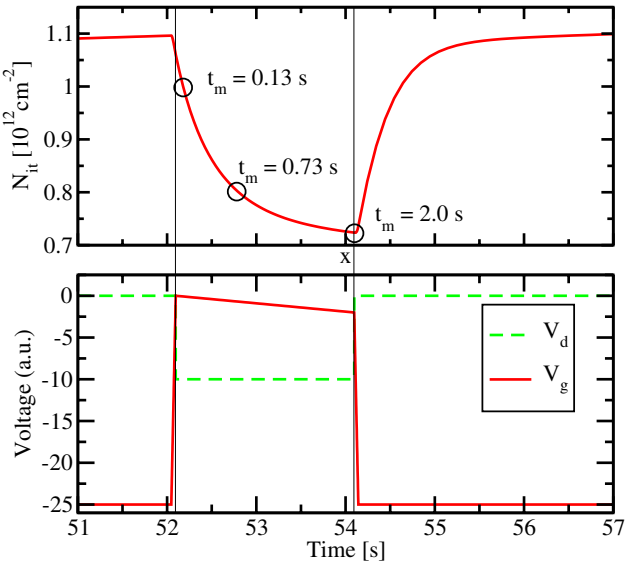


Figure 6.3: Measurement cycle during NBTI characterization. When the gate voltage V_g is switched from stress to measurement conditions (lower figure) a strong recovery effect can be seen (upper figure). For best accuracy the measurement delay has to be reduced as much as possible.

the first milliseconds or even microseconds after the stress is released. Figure 6.3 qualitatively illustrates the measurement error obtained due to the fast recovery of NBTI degradation. When the stress gate voltage is removed, a part of the degradation relaxes. As a consequence the measured degradation of figures of merit, such as the threshold voltage or the interface trap density, is strongly dependent on the delay between the end of the stress conditions and the time where the value of the degradation is measured. This fast transient effect has been shown by Ershov *et al.* in 2003 [91] and led to the development of advanced measurement methods as described below.

6.3.2 Faster Determination of the Threshold Voltage

To reduce the delay between stress and measurement, Kaczer *et al.* [92] proposed an improved measurement method to reduce the amount of relaxation. Instead of performing a full I_d vs V_g sweep, the gate voltage is reduced from stress condition to a constant value around the initial threshold voltage, $V_g \approx V_{th}$. At this bias condition only one value for the drain current, I_d, is measured. Using the unstressed I_d/V_g characteristic, an approximation for the new threshold voltage can be extracted by horizontally shifting the unstressed characteristic as outlined in Figure 6.4.

As only one drain current has to be obtained, instead of a range of drain currents for different gate biases, this approach is drastically faster and the error due to the fast recovery is reduced.

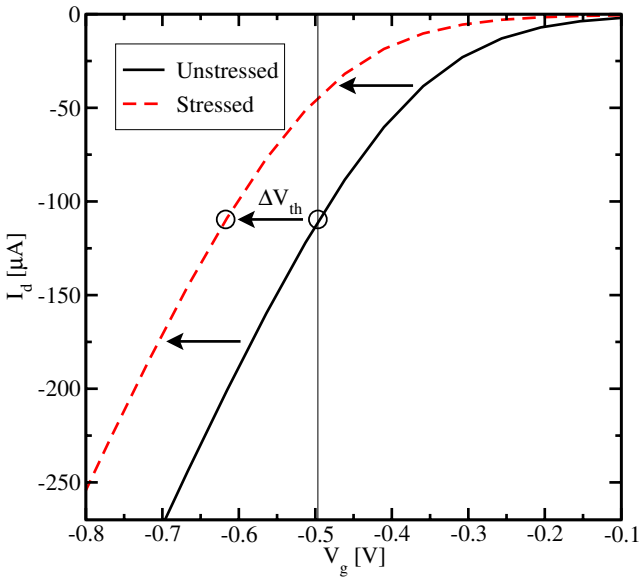


Figure 6.4: Fast method for determining the threshold voltage shift from a single point measurement. The drain current I_d is measured for $V_\mathrm{g} \approx V_\mathrm{th}$. The estimated value for ΔV_th is extracted from horizontally shifting the initial I_d-V_g curve.

This technique still suffers from a measurement delay of some tens or hundreds of milliseconds. In order to completely remove the, normally unintentional, recovery during the measurement conditions on-the-fly measurement methods with zero delay have been proposed [93, 94], which are discussed in the next sections.

6.3.3 Drain Pulsed Voltage

To completely remove the relaxation of the device degradation during measurement cycles, the gate voltage is constantly set to stress bias, $V_\mathrm{g} = V_\mathrm{stress}$. The drain voltage is normally kept at zero volt, but it is periodically pulsed with gate stress voltage, $V_\mathrm{d} = V_\mathrm{stress}$. During the pulses the drain current is measured [94].

One concern using this method might be the following: although the stress gate voltage is never removed during measurement, biasing the drain contact with the same voltage still leads to a drastic change of the electric field at the Si/SiO_2 interface, especially close to the drain contact. As the electric field has been found to be one of the major driving forces of NBTI degradation, this change of the field will most likely lead to altered degradation behavior.

Figure 6.5 depicts the electric field at the Si/SiO_2 interface of a MOSFET biased at symmetric stress conditions with $V_\mathrm{g} = V_\mathrm{stress}$ and $V_\mathrm{d} = 0\,\mathrm{V}$ and also with $V_\mathrm{d} = V_\mathrm{g} = V_\mathrm{stress}$. It can be clearly seen that the electric field component normal to the interface changes severely when applying a high drain voltage. This might invalidate the main argument for this measurement method, of not changing the degradation behavior during measurement.

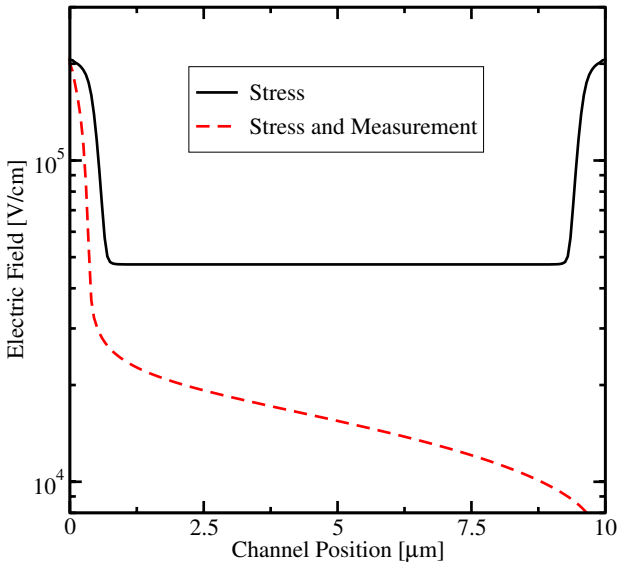


Figure 6.5: The electric field component normal to the Si/SiO_2 interface of an MOS transistor. The solid line is for symmetric bias conditions where $V_g = 1.2\,V$ and $V_s = V_d = V_{sub} = 0\,V$. The electric field at the interface changes drastically when the drain contact is biased with $V_d = V_g$ (dashed line).

The amount by which the electric field is altered during measurement depends strongly on the exact device geometry, the doping concentrations, and also the stress bias. Appropriate simulations have to be made to evaluate the impact on the electric field before using this measurement for device qualification.

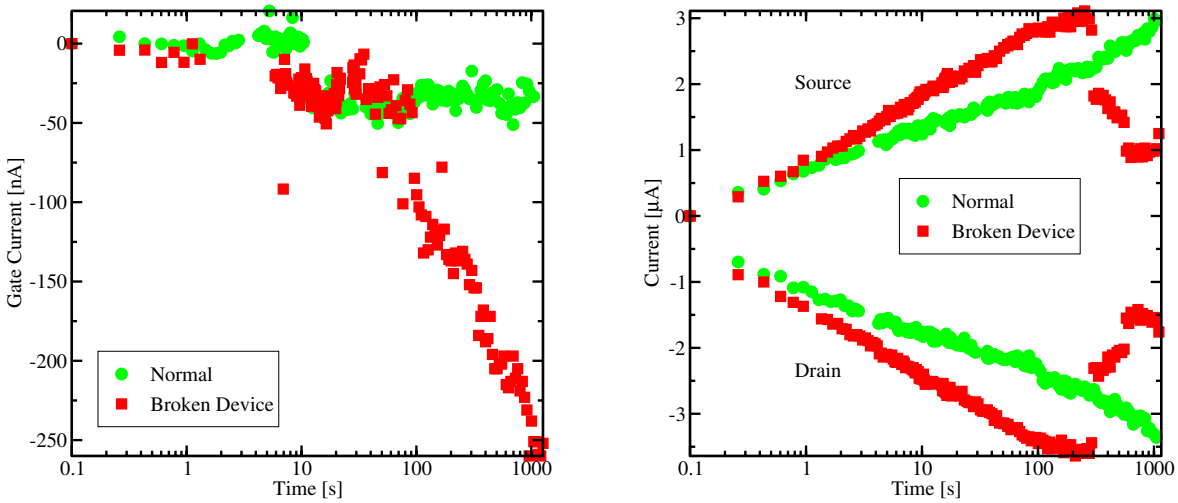
Another problem of this method is the high channel current during measurement. It can lead to additional degradation mechanisms due to hot carrier stress (Section 5.1), as hot carrier injection into the oxide.

For circuit design the threshold voltage is in most cases of more interest than the drain current in saturation. The next on-the-fly method extracts the threshold voltage while keeping the gate stress approximately constant.

6.3.4 Gate Pulsed Voltage

By using the gate pulsed voltage (GPV) technique, the transconductance g_m, the drain current in the linear regime $I_{d,lin}$, and the threshold voltage shift ΔV_{th} can be extracted while keeping the gate stress nearly constant [93, 94].

Using this method, the source and bulk contacts are grounded while the drain contact is slightly biased with typically $-25\,mV$. The stress at the gate contact is permanently applied and superimposed with pulses which are small compared to the stress voltage. With the small drain voltage, as compared to the drain pulsed voltage method, neither the problem of the severely changing interface electric field nor hot carrier degradation are imminent.



(a) Gate dielectric damage

(b) Device damage

Figure 6.6: Observed terminal voltages during stress. Abrupt changes in the terminal currents indicate an additional degradation mechanism such as dielectric breakdown (Section 5.2). Data are from IMEC.

The transconductance g_m of the transistor, which is defined as the change in drain current as a result of a change of gate voltage, can then be extracted as

$$g_m = \frac{\partial I_{d,lin}}{\partial V_g} = -\frac{\partial I_{d,lin}}{\partial V_{th}} . \tag{6.2}$$

From this relation also the threshold voltage can then be extracted.

6.3.5 Observation of Terminal Currents

During stress tests it is often useful to measure the source, drain, and gate currents, I_s, I_d, and I_g. Devices which suffer from oxide breakdown during stress can then be detected as I_g should stay very low and I_s should be approximately equal to I_d. Sudden changes in terminal voltages, as seen in Figure 6.6, are not typical for NBTI and suggest an additional degradation mechanism.

6.3.6 Importance of Initial Degradation

The effect of fast recovery of NBTI degradation has received much attention in recent years. Many groups proposed enhanced measurement schemes, as the on-the-fly measurements described above, to measure the true V_{th} shift free from any recovery effects [93–98]. Still, there is much controversy about different reported NBTI characteristics ranging from power-law exponents (Section 6.2) of around 0.16 [98] or much smaller [96] to logarithmic dependence [93].

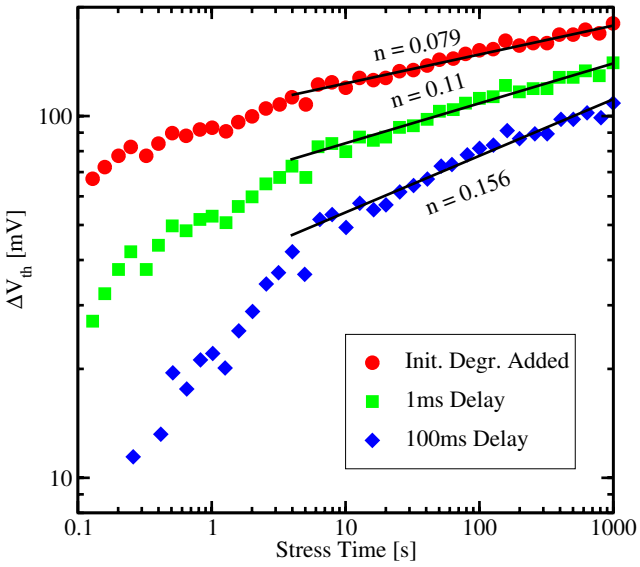


Figure 6.7: On-the-fly measurements with slow determination of the initial V_{th} after Shen *et al.* [99]. As the first measurement already suffers from notable degradation, all the following ΔV_{th} are underestimated. As this effect is pronounced for lower stress times, it can be observed as a kink around 8 seconds.

As the major problem leading to the spread of the literature data Shen *et al.* identified the slow measurement of the initial degradation before stressing the device [99]. The first measurement is compared to the subsequent measurements to extract the threshold voltage shift. As the first measurement already captures a certain amount of degradation this shift is underestimated. This underestimation is especially serious at short stress times and results in a kink as observed in a log-log plot and, probably more important, in completely different, apparent slopes in the power-law of the degradation.

Figure 6.7 depicts data from [99] showing the apparent threshold voltage shift over time in logarithmic scale. It can be seen that for long delay times of 100 ms at the initial measurement, the kink in the power law is very pronounced. But as the initial measurement time is drastically reduced to 1.04 ms, and therefore the accuracy of this measurement increased, the kink becomes less obvious. The authors showed that by assuming a shift of the threshold voltage of 40 mV even at the fast initial measurement, and adding it to the following data they can achieve a clear power-law with the time exponent of $n = 0.079$. This slope varies drastically when the amount of initial degradation is not added and might explain the wide spread of the data reported in literature.

To conclude, not only the speed of the degradation measurements during stress is important in order to spot the fast transients during relaxation, but also the accuracy of the initial degradation measurement is of highest importance and can be improved by very high speed measurement methods [100].

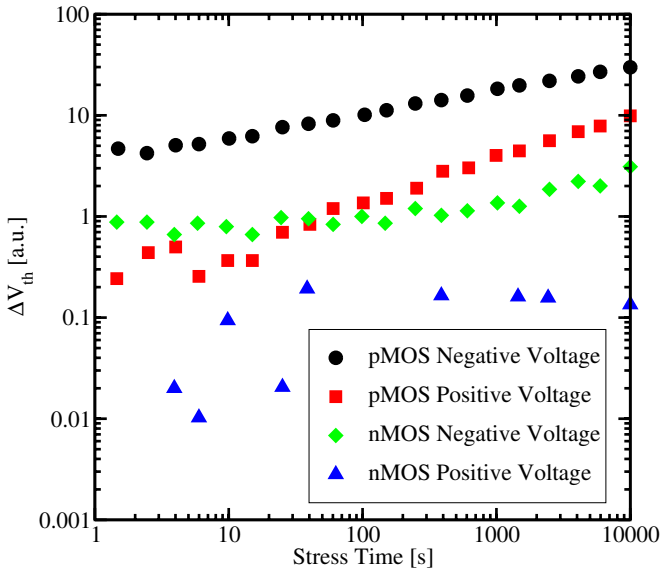


Figure 6.8: The effect of bias temperature instability for p- and n-channel MOS structures with positive and negative voltage stress. The highest degradation can be found at p-channel MOSFETs stressed with negative gate voltages, which is the typical NBTI stress. The measurement data are from Huard *et al.* [86].

6.3.7 pMOSFET vs. nMOSFET

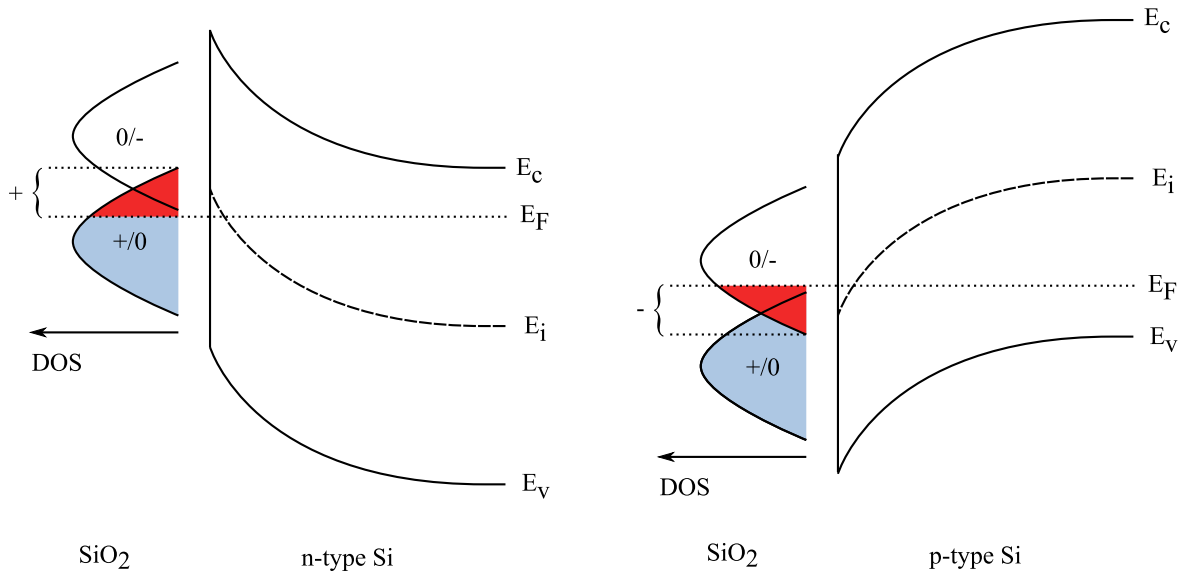
The effect of bias temperature instability can be observed in both, p-channel and n-channel MOSFETs. However, Figure 6.8, presenting data from Huard *et al.* [86], shows that p-channel MOSFETs with negative gate voltage stress are much more susceptible to this kind of degradation.

It has been reported that for NBTI degradation channel cold holes are important [86, 101]. As the n-channel MOSFET biased into accumulation also has holes at the surface of the substrate, the threshold voltage shift should be similar to p-channel MOSFETs. Therefore, the lack of holes can not be the cause for the different degradation behavior.

6.3.7.1 Charge States

One explanation for the different susceptibility of bias temperature stress (BTS) is based on the charge states of the interface traps, N_{it}, and the oxide traps, N_{ot}.

Oxide traps are positively charged by hole trapping in both, the n- and p-channel MOSFET. Therefore their contribution to a V_{th} shift should be comparable. The charge state of the interface traps, on the other hand side, depends on the Fermi-level as pointed out in Chapter 3. The threshold voltage of a MOSFET is reached when the surface of the substrate is driven into weak inversion. Therefore the charge state at the silicon-dielectric interface under inversion conditions is very important.



(a) p-channel MOSFET

(b) n-channel MOSFET

Figure 6.9: Energy diagram of a (a) p-channel MOSFET and an (b) n-channel MOSFET at the Si/SiO_2 interface in weak inversion. The Fermi-level of the left substrate is energetically located in the donor trap resulting in a positive net charge at the interface. The Fermi-level of the right substrate, in contrast, is energetically located in the acceptor trap leading to negative interface charge.

The Fermi-level at the interface of a p-channel MOSFET biased into inversion is below the mid-gap energy. As Figure 6.9(a) illustrates, in this regime the acceptor like trap levels are empty, and the donor like trap levels are partially filled, assuming the trap energy levels to be symmetrically aligned around mid-gap. The net charge resulting from this configuration is positive and leads to a total interface charge for p-channel MOSFETs of

$$Q_{\text{tot}} = Q_{\text{ot}} + Q_{\text{it}} \,. \tag{6.3}$$

For the n-channel MOSFET the situation is completely different. Figure 6.9(b) illustrates the energy diagram of an nMOSFET in inversion. Here, the Fermi-level at the interface is located above the mid-gap energy. The donor like trap levels in the lower half of the band gap are therefore completely filled and the acceptor like trap levels are partially filled. The result is a negative charge due to the interface traps which leads to a total interface charge for n-channel MOSFETs of

$$Q_{\text{tot}} = Q_{\text{ot}} - Q_{\text{it}} \,. \tag{6.4}$$

In p-channel MOSFETs, both, the interface and the oxide traps contribute to the threshold voltage shift. In n-channel devices the charge state of the interface traps reduces the shift introduced by the oxide traps and the effect of NBTI is therefore less pronounced in n-channel MOSFETs.

6.3.7.2 Availability of Hydrogen

Another possible explanation for the different degradation behavior can be found in the model of Tsetseris *et al.* [102]. As described in Section 6.5, the model proposes the importance of atomic hydrogen, H^+, for weakening and breaking of Si-H bonds at the interface. These protons originate from the substrate of the transistor and are initially bond to dopants, usually phosphorus in the case of a pMOS (P-H bonds).

The activation energy for dissociation of P-H bonds is usually very high. Only in the depletion region the bonds are weakened and can be cracked more easily. During NBT stress a p-channel MOSFET is driven into inversion and a depletion region is formed in the substrate. In this region the hydrogen atoms can be dissociated and cause degradation at the interface. For positive stress no depletion region is formed. Therefore no hydrogen can be provided by dissociating P-H bonds, and the level of degradation is reduced.

In the n-channel MOSFET the substrate is typically doped with boron and hydrogen can only be supplied from B-H bonds. As their binding energy is much higher, also in depletion regions, n-channel MOSFETs are much less susceptible to bias instability, no matter which polarity the stress voltage has.

6.3.7.3 Work Function Difference

Another reason for different degradation behavior of p- and n-channel MOSFETs might be found in the fact that for a given oxide electric field the gate voltages differs [84].

The Fermi-level of a p-type poly gate lies approximately at the valence band edge $E_F = E_v$ while the Fermi-level of an n-type poly gate is located at the conduction band edge energy $E_F = E_c$. For NBT stress the p-type poly gate of the p-channel device is depleted while the n-type poly gate of the n-channel device is driven into accumulation.

Oxide and interface charges (Q_{ox} and Q_{it}) generated by NBTI degradation further shift the relevant gate voltages as described in Section 2.2.1.

Due to this asymmetry for n- and p-type gate contacts, the oxide electric fields are not equal for a given gate voltage and at a given oxide electric field the gate voltages differ. This has to be considered when comparing the two types of transistors.

6.3.8 Influence of Channel Carrier Transport

During NBT stress conditions the voltages of substrate, source, and drain contacts are typically equal (Section 6.3.1) or close to equal (Section 6.3.4). Under these conditions, the charge transport in the channel is negligible and so is its impact during NBTI investigations. The only displacement of channel carriers, holes in case of NBTI on p-channel MOSFETs, is due to thermal activity.

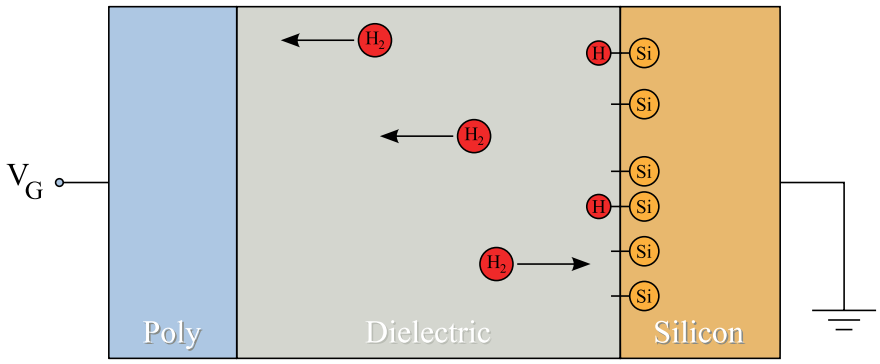


Figure 6.10: Schematic representation of the reaction-diffusion (R-D) model. Electrically inactive Si-H bonds at the Si/SiO_2 interface are broken and the hydrogen diffuses into the dielectric leaving behind an electrically active interface trap N_{it}. Here, H_2 diffusion is assumed.

6.4 Reaction-Diffusion Model

The reaction-diffusion (R-D) model is pioneering for the description of NBTI. It was first proposed by Jeppson and Svensson in 1977 [103] and is capable of reproducing the time evolution of device degradation due to negative bias temperature stress for a wide range of measurements.

The model describes the device degradation as a combination of two effects. In the first place, a field-dependent electrochemical reaction at the Si/SiO_2 interface. Electrically inactive, passivated silicon dangling bonds (Section 3.1), Si-H, are broken. Here an electrically active interface state N_{it} and a mobile, hydrogen related species X are formed,

$$\mathrm{Si-H} \rightleftharpoons \mathrm{Si^{+}} + \mathrm{X_{interface}}\,. \tag{6.5}$$

In the second place, the model describes the transport of the hydrogen species away from the interface into the dielectric,

$$\mathrm{X_{interface}} \rightleftharpoons \mathrm{X_{bulk}}\,. \tag{6.6}$$

Also the reverse process is possible: transport of a diffusing hydrogen species back to the interface and re-passivation of a $Si^{\bullet}$ dangling bond.

Figure 6.10 gives a schematic illustration of the model. In the figure hydrogen molecules, H_2, are assumed as the diffusing species. The process at the interface is modeled by a rate equation as

$$\frac{\partial N_{it}(t)}{\partial t} = \underbrace{k_f(N_0 - N_{it}(t))}_{\text{generation}} - \underbrace{k_r N_{it}(t) N_X(0,t)^{1/a}}_{\text{annealing}}, \tag{6.7}$$

where k_f is the interface-trap generation and k_r the annealing rate. The symbol N_0 denotes the initial number of electrically inactive Si-H bonds and $N_X(0,t)$ is the surface concentration of

the diffusing species. The value of a gives the order of the reaction. In the original publication, neutral hydrogen, H^0, was proposed [103] which is obtained with $a = 1$. For molecular hydrogen, H_2, $a = 2$. In this case the molecule is assumed to be formed in the vicinity of the interface

$$\mathrm{Si-H} + h \;\rightleftharpoons\; \mathrm{Si^+} + \tfrac{1}{2}\mathrm{H_2}\,. \tag{6.8}$$

The Si-H bond is broken and captures a hole, leading to a positively charged interface state and molecular hydrogen is formed.

The equilibrium of the forward and backward reaction is controlled by the hydrogen density at the interface $N_\mathrm{X}(0,t)$. Thus, the transport mechanism of the hydrogen species away from the interface characterizes the degradation mechanism, controlling the device parameter shift. The original reaction-diffusion model describes the transport as a purely diffusive mechanism which is described by the diffusion equation

$$\frac{\partial N_\mathrm{X}(x,t)}{\partial t} = D\,\vec{\nabla}^2 N_\mathrm{X}(x,t)\,. \tag{6.9}$$

Here, D is the diffusivity of the hydrogen species in the dielectric. The influx of the newly created species has to be considered as

$$a\frac{\partial N_\mathrm{it}}{\partial t}\,. \tag{6.10}$$

For each generated interface trap a hydrogen is released, thus

$$N_\mathrm{it} = \int_x N_\mathrm{X}\,\mathrm{d}x\,. \tag{6.11}$$

The R-D model assumes the interface states to be the only contribution to device parameter shift. But especially thick high-voltage oxides also show the generation of oxide charges due to hole trapping [86], leading to an additional parameter shift. Therefore it has been suggested [85] to separate the oxide charge contribution from the measurement results, before the R-D model can put into agreement to them. Suggested methods are the estimation of bulk-trap concentration for every stress voltage, temperature, and oxide thickness or trying to avoid the generation of bulk-traps by optimizing the stress conditions to that aspect.

6.4.1 Properties of the R-D Model

For the stress phase the solution of the R-D model can be split up into five different regimes. They are distinguished by different time exponents n (Section 6.2) for the degradation and are depicted in Figure 6.11.

Regime 1: In the very early stage of the stress phase the amount of free hydrogen, both, at the interface and in the dielectric N_X is very low. The amount of already broken Si-H bonds at the interface N_it is close to zero. Thus, Equation 6.7 is solely limited by the forward reaction rate k_f and transforms to

$$\frac{\partial N_\mathrm{it}(t)}{\partial t} \approx k_\mathrm{f} N_0\,, \tag{6.12}$$

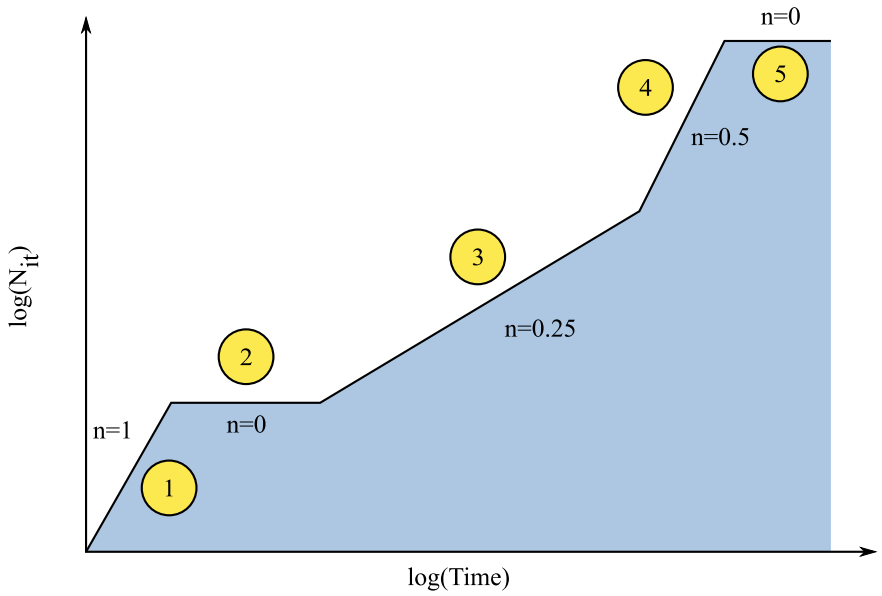


Figure 6.11: The five different regimes for the time exponent n, as obtained from the reaction-diffusion model.

with the solution for the interface trap generation

$$N_{\mathrm{it}}(t) \approx k_{\mathrm{f}} N_0 t \,, \tag{6.13}$$

when assuming

$$N_{\mathrm{it}}(0) = 0 \,, \tag{6.14}$$

having a time dependence of $n = 1$.

Regime 2: After some time, when the amount of hydrogen at the interface is considerable, the forward reaction reaches a quasi-equilibrium with the backward reaction

$$k_{\mathrm{f}}(N_0 - N_{\mathrm{it}}) \approx k_{\mathrm{r}} N_{\mathrm{it}} N_{\mathrm{X}} \,, \tag{6.15}$$

and, as N_0 is very large with respect to N_{it}

$$k_{\mathrm{f}} N_0 \approx k_{\mathrm{r}} N_{\mathrm{it}} N_{\mathrm{X}} \,. \tag{6.16}$$

The diffusion process has not removed a considerable amount of hydrogen from the interface yet, therefore the amount of interface traps equals the amount of hydrogen at the interface

$$N_{\mathrm{it}} = \int_x N_{\mathrm{X}} \,\mathrm{d}x \approx N_{\mathrm{X,interface}} \,. \tag{6.17}$$

We therefore get

$$N_{\mathrm{it}} \approx \sqrt{\frac{k_{\mathrm{f}} N_0}{k_{\mathrm{r}}}} \,. \tag{6.18}$$

As this equation is not time dependent we can write

$$N_{\mathrm{it}} \approx \sqrt{\frac{k_{\mathrm{f}} N_0}{k_{\mathrm{r}}}} t^0 \,. \tag{6.19}$$

with a resulting time dependence of $n = 0$.

As long as the diffusion of hydrogen away from the interface has not reached a considerable magnitude, there is no further degradation of the interface.

Regime 3: Regime three starts when the diffusion of hydrogen away from the interface sets in and acts as limiting factor for the degradation. In this phase the diffusion front has not reached the poly gate, therefore

$$t_{\mathrm{ox}} > \sqrt{4Dt} \,. \tag{6.20}$$

An analytical solution has been found to be [103]

$$N_{\mathrm{it}} \approx 1.16 \sqrt{\frac{k_{\mathrm{f}} N_0}{k_{\mathrm{r}}}} D^{1/4} t^{1/4} \,. \tag{6.21}$$

Numerically the solution is, depending on the group, somewhat larger [85] or smaller [104] then the analytical solution of $n = 0.25$ as it is based on some simplifications.

In the reaction-diffusion model, this is the dominating regime in the typical lifetime of a MOSFET. It sets in after some seconds stress and, depending on the exact conditions, is dominating for several orders of magnitude in time, lasting up to several years.

Regime 4: When the diffusion front reaches the poly gate contact the time exponent changes again. In the model it is assumed that the gate electrode acts as absorber for the diffusing species or, in other words, the diffusivity is significantly higher in the poly than in the dielectric. For this case a time exponent of $n = 0.5$ is derived. Although a slight increase of the time exponent can be found in some measurement data for thin dielectrics, it is way below 0.5.

Regime 5: When, theoretically, all interface bonds N_0 are broken and

$$N_{\mathrm{it}} \approx N_0 = \mathrm{const} \,, \tag{6.22}$$

no further degradation can occur in this model. Therefore the change in N_{it} is zero and so is the time exponent n. As this saturation condition would only occur ever at extremely long stress times, or, at very high stress conditions which would lead to other degradation mechanisms as TDDB (Section 5.2), it has not yet been observed experimentally.

6.4.2 R-D Model vs. Fast Recovery

It has been shown that the measured degradation and also the extracted time exponent strongly depend on the time delay during measurement. Using fast measurement methods to spot the

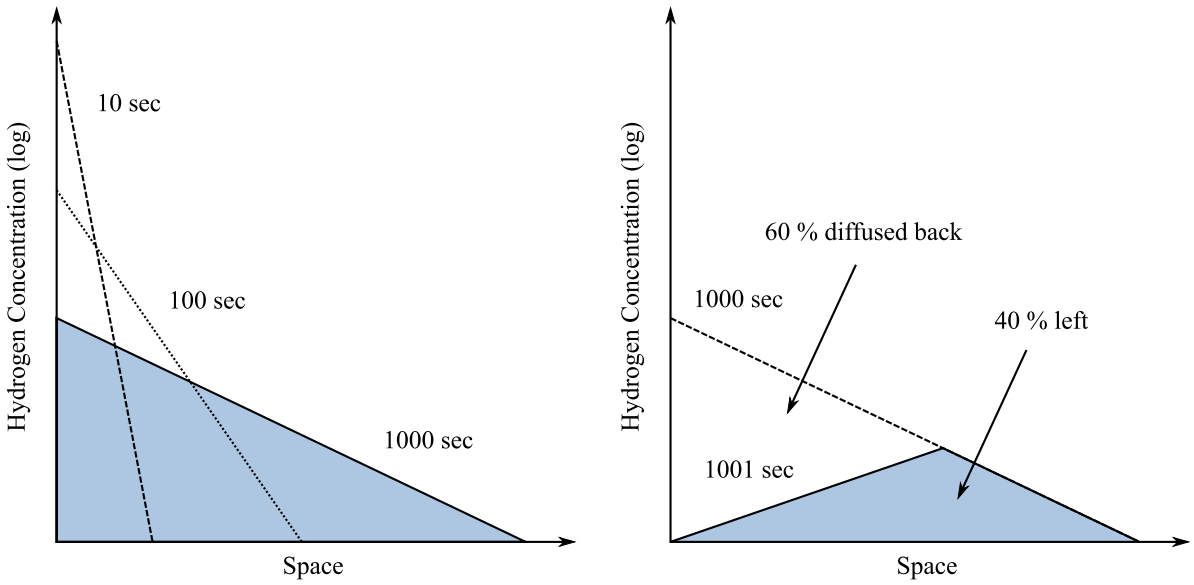


Figure 6.12: Hydrogen profiles in the dielectric for 1000 second stress (left) and after the recovery of 60 % of the degradation. It is difficult to argue that the diffusion of more than half of the hydrogen which took 1000 seconds to diffuse into the dielectric can diffuse back to the interface in one second [99].

full degradation during NBT stress reveals a very low time exponent around $n = 0.07 - 0.1$ (see Section 6.3.6). The reaction-diffusion model, in contrast, predicts a time exponent of $n = 0.25$ for neutral hydrogen diffusion, H^0. By assuming molecular hydrogen being the diffusing species a time exponent of $n = 0.16$ can be achieved. Here, an additional reaction step at the interface generates molecular hydrogen before diffusion, $2H^0 \rightarrow H_2$. Only by introducing higher order chemical reactions at the interface the time exponent could be further reduced to fit the model to state-of-the-art measurement results.

The fast recovery effect poses another inconsistency of the pure R-D model with measurement data [99]. Even for very long stress times of more than 1000 seconds the immediate recovery of V_{th} in the first second after stress is more than 60 % using the fast measurement set-up.

The R-D model assumes diffusion limited degradation and also diffusion limited relaxation. When considering a stress time of 1000 seconds a certain amount of hydrogen related species diffuses into the dielectric as shown in Figure 6.12. The total amount of hydrogen in the dielectric must equal the number of interface traps N_{it} (6.11). N_{it} is in turn directly proportional to the shift of the threshold voltage ΔV_{th}. To argue a 60 % recovery during the first second in the framework of the R-D model, 60 % of the hydrogen must diffuse back to the interface and anneal the dangling silicon bonds within this second. This would result in a hydrogen profile in the dielectric, as seen on the right hand side of Figure 6.12, implying that the backward diffusion must be orders of magnitude faster than the forward diffusion [99].

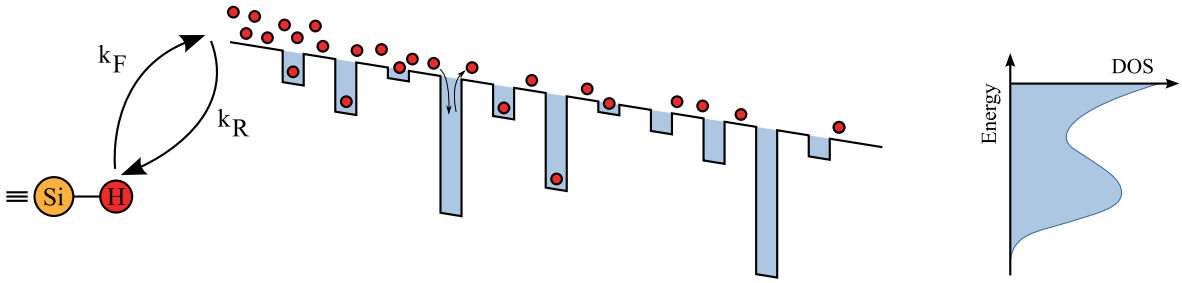


Figure 6.13: Schematic illustration of dispersive transport. Hydrogen is dissociated from the interface and transported into the SiO_2 via a diffusive mechanism. In the dielectric the mobile species sees traps of different energies leading to dispersive transport. The energy diagram on the right hand side shows a possible density-of-states (DOS) energy distribution of the traps.

6.4.3 Dispersive Transport

Instead of using the standard diffusion equation [103, 105] to describe hydrogen transport in the dielectric it has been shown [88, 106–112] that correct modeling of transport in an amorphous materials must consider its dispersive [113] nature.

Dispersion arises when the mobile species experiences different barrier heights at different positions in space. Figure 6.13 gives a schematic illustration where the dielectric contains hydrogen traps of different energy. The trapping probability at each trap is the same. But for de-trapping the deeper traps pose a higher barrier than the shallow ones. This implies that at the beginning of the trapping events the energy distribution of trapped carriers is proportional to the density-of-states (DOS). But after some time, as de-trapping preferably occurs for shallow traps, the peak of the energy distribution moves to deeper traps as illustrated in Figure 6.14. Thus, the equilibrium distribution is totally different from the DOS.

In the modeling approach, the species N_X is separated into two distinct contributions. The conducting, N_c, and trapped, N_t, particles. The trapped particles are distributed in energy where the density at a trap energy-level E_t is given as $\rho(\vec{x}, E_t, t)$. The trapped particles do not contribute to the transport. To introduce dispersive transport into the reaction-diffusion model, (6.9) transforms to

$$\frac{\partial N_c(\vec{x},t)}{\partial t} + \frac{\partial N_t(\vec{x},t)}{\partial t} = D\vec{\nabla}^2 N_c(\vec{x},t)\,. \tag{6.23}$$

At each trap energy-level a rate equation describes the dynamics between trapping and de-trapping as

$$\frac{\partial \rho(E_t)}{\partial t} = cN_c\Big(g(E_t) - \rho(E_t)\Big) - r(E_t)\rho(E_t)\,. \tag{6.24}$$

Here, c is the capture rate, $r(E_t)$ the energy-dependent release rate, and $g(E_t)$ is the trap DOS.

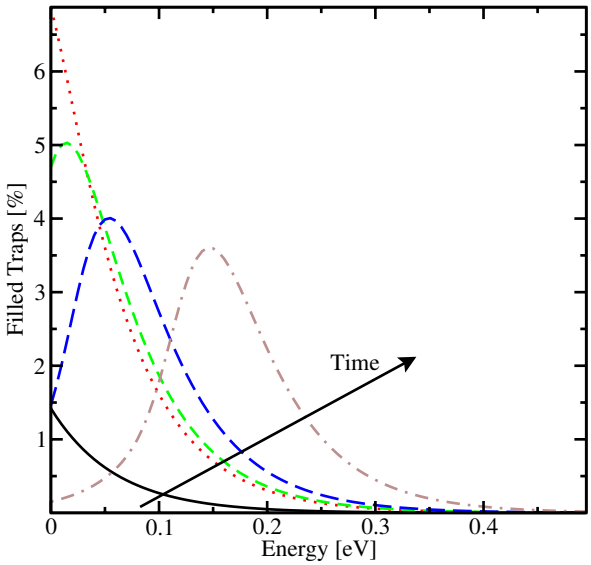


Figure 6.14: At small times the energy distribution of trapped carriers is similar to the density-of-states (DOS) as the trapping probability for each trap is the same. After time the peak of the energy distribution moves towards deeper states. The reason is the higher de-trapping probability for shallow traps.

When considering the trap distribution in exponential form [92, 113]

$$g(E_\mathrm{t}) = \frac{N_\mathrm{t}}{E_0} \exp\Big(-\frac{E_\mathrm{c} - E_\mathrm{t}}{E_0}\Big) \,, \tag{6.25}$$

and with introduction of the dispersion parameter

$$\alpha = \frac{\mathrm{k_B} T}{E_0} \,, \tag{6.26}$$

N_it can, for neutral species, be expressed with a power-law as [114]:

$$\Delta N_\mathrm{it}(t) = \Big(a\Big(\frac{k_\mathrm{f} N_0}{k_\mathrm{r}}\Big)^a\Big)^{1/(1+a)} \Big(\frac{D}{\nu_0}\frac{N_\mathrm{t}}{N_\mathrm{c}}\Big)^{1/(2+2a)} \Big(\frac{1+a}{1+a\alpha/2}\Big)^{1/(1+a)} (\nu_0 t)^{(1-\alpha/2)/(1+a)} \,, \tag{6.27}$$

with ν_0 being the release rate coefficient. Here, only hydrogen in the conductive state can contribute to the reverse rate.

When assuming atomic hydrogen ($a = 1$) the slope calculates from (6.27) as $n = 1/2 - \alpha/4$ while when assuming molecular hydrogen ($a = 2$) H_2 results in $n = 1/3 - \alpha/6$. In both cases the time exponent n is increasing for an increasing dispersion parameter α. So for an increasing number of deep traps the degradation increases [104, 111].

For complex trap distribution functions no analytic formulation can be found. An example is a combination of shallow band-tail traps (the exponential distribution) and deep Gaussian traps describing deep traps (right hand side of Figure 6.13) [111]. For such a DOS only the numerical solution is possible as obtained with the numerical device simulator Minimos-NT.

6.4.4 Coupling to the Semiconductor Device Equations

As NBTI is strongly dependent on the stress conditions, the available models include at least the dependence on the electric field (the bias) and the temperature. But also the hole concentration at the interface might be of interest [85].

In the one-dimensional NBTI models the values of these quantities are accounted for in the model parameters, as k_f and k_r in the reaction-diffusion model for example. For each device, stress voltage, and temperature these parameters are extracted (Section 7.1.2) and are assumed constant for these stress conditions. But as degradation is continuing during stress, the underlying parameters, especially the electric field across the interface, change. So the electric field at the Si/SiO_2 interface is decreasing when the NBTI induced degradation leads to positive charge build up at this interface. Therefore, it might not be appropriate to consider the forward reaction rate k_f as a constant.

Another point is the two-, or three-dimensional investigation of a device structure. Here the electric field and the hole concentration, and therefore also the magnitude of degradation, can change drastically across the channel of the device when the stress conditions are not uniform (Section 6.3.3). Thus, the model parameters are not uniformly distributed at the Si/SiO_2 interface and a one-dimensional consideration of the model might not be accurate enough [115].

To solve these issues it is important to include the NBTI models into a numerical device simulation. The distributed quantities, obtained from the semiconductor device equations (Section 2.1) as the electric field, hole and electron concentration can be directly used in the degradation model. The resulting charge densities from the NBTI model can be included in the device equation leading to a fully self-consistent coupling of those two sets of equations.

The reaction-diffusion model might be linked to the semiconductor device equations with the forward reaction rate as follows [85]

$$k_\mathrm{f} = k_{\mathrm{f},0} \frac{p_\mathrm{s}}{p_\mathrm{ref}} \exp\left(\frac{E_\mathrm{ox}}{E_\mathrm{ref}}\right) . \tag{6.28}$$

Where p_s and E_ox are the hole concentration and the electric field at the interface, p_ref and E_ref are reference values. By using solution variables of the semiconductor equations the NBTI model can be applied to arbitrary device geometries.

6.5 Tsetseris' Model

There is still no comprehensive atomic-scale scenario describing the effect of BTI including the exact mechanism of degradation, the role of different species, or the origin of low activation energies. By means of first-principles calculations it is possible to investigate the atomic configuration of the semiconductor's structure, including dopants, unwanted defects, and the binding and dissociation energies of atomic bonds.

Tsetseris and co-workers investigated the subject of BTI at atomic level using the ab-initio simulation code VASP [116]. They propose a proton based dissociation model [102] to describe

the BTI induced breaking of Si-H bonds at the silicon-dielectric interface, which is described in the following.

The breaking of passivated Si-H bonds at the Si/SiO_2 interface and the generation of dangling silicon bonds, $Si^{\bullet}$, has been suggested as direct mechanism in the reaction-diffusion model

$$\text{Si-H} \rightarrow \text{Si}^{\bullet} + \text{H}\,. \tag{6.29}$$

Ab-initio calculations have shown [102] that the removal of hydrogen from the Si-H bond and its transport to a remote Si-Si bond raises the energy of the system by 1.9 eV. Adding the migration barrier leads to a necessary dissociation activation energy of 2.4 eV which is in good agreement with the experimentally obtained value of 2.6 eV [117]. When holes are present in the channel the activation energy upon H removal decreases to 2.1 eV. These values show, however, that only hot carriers in the channel, which are not available in typical BTI stress conditions (Section 6.3.8), can obtain enough energy to break Si-H bonds . Rashkeev *et al.* [118] propose an alternative depassivation reaction which adds a proton,

$$\text{Si-H} + \text{H}^{+} \rightarrow \text{Si}^{\bullet} + \text{H}_2\,. \tag{6.30}$$

The proton weakens the Si-H bond and after dissociation creates molecular hydrogen, H_2, which diffuses into the dielectric.

Through ab-initio calculations [119] a dissociation energy barrier of 0.95 eV is obtained, when the Fermi-level is at the valence band edge, which is the situation of an n-type substrate in inversion. Due to the lower energy barrier, this process can be activated by BTI temperatures.

As source for the required protons the authors suggested hydrogen bound to substrate dopants. The activation energy to dissociate a P-H complex has calculated to be 1.3 eV, which is in good agreement with an experimental value of 1.18 eV [120]. These energy values are valid in n-type silicon at flat band conditions where hydrogen exists as H^{-}.

In the depletion region the stability of the P-H complex changes dramatically. In this region the preferred charge state of hydrogen changes from negative to neutral, H^0, resulting in a much lower dissociation activation energy for a P-H complex of only 0.3–0.4 eV. For a certain period of time the hydrogen atom located in the inversion layer stays neutral and then transfers into a positively charged proton, H^{+}, by picking up a hole.

At negative bias stress conditions the free H^{+} protons drift to the Si/SiO_2 interface. As the energy barrier to cross the interface is very high (1 eV) the protons migrate rapidly along the interface. As both, dangling bonds and the protons are positively charged it is very unlikely that a proton passivates an interface trap. They will preferably be located close to Si-H bonds which they can break by forming H_2 (6.30). Figure 6.15 gives a schematic overview of the processes involved.

An additional mechanism can be the injection of protons into the dielectric leading to positive oxide charges Q_{ox}. These positive charges prevent further protons to be injected. Only when the oxide charges diffuse away from the interface new protons can be injected leading to further degradation [102].

This model is capable of explaining the different susceptibility of n- and p-channel MOSFETs to positive and negative bias stress (Section 6.3.7)

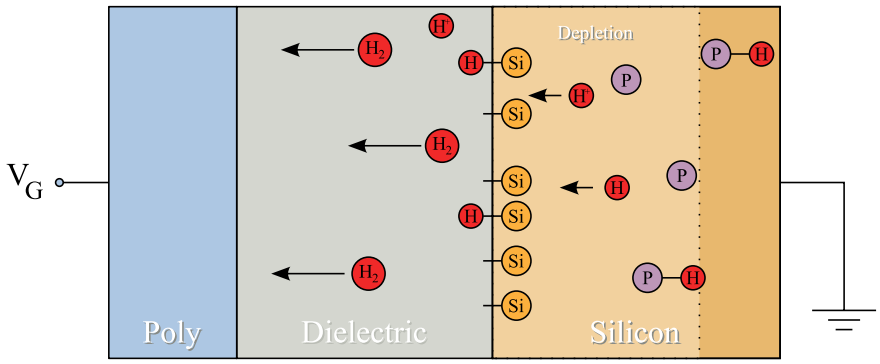


Figure 6.15: Schematic of the NBTI model proposed by Tsetseris *et al.* [102]. P-H bonds in the depletion region of an n-type substrate are drastically weakened at NBTI stress conditions. After dissociation the hydrogen diffuses to the interface and can de-passivate Si-H bonds generating silicon dangling bonds.

- pMOS, negative bias: The P-H dissociation energy is drastically reduced in the depletion region and the positively charged protons are accelerated towards the interface.
- pMOS, positive bias: As no depletion layer is formed the dissociation energy of P-H is very high. Therefore the degradation is drastically suppressed.
- nMOS: In the p-type substrate B-H bonds exist instead of P-H bonds. However, the dissociation activation energy of these complexes is much larger than for P-H complexes, even in a depletion layer, leading to less hydrogen available to degrade the interface.

When assuming that the atomic hydrogen in the bulk diffuses faster than the molecular hydrogen in the dielectric, Tsetseris' model results in a time exponent of $n = 0.25$ [121]. This is not in agreement with recent measurement data (Section 6.3.6) and therefore the model in the present form might be incomplete.

6.6 The New Model for Numerical Simulation of NBTI

From literature it is obvious that no clear consensus about the physical mechanism nor a model which is commonly agreed upon is found up to now. Various research groups report different *key* mechanisms leading to NBTI and therefore propose different models which sometimes completely contradict each other [85, 86, 102, 110, 122]. Is it purely the interface degradation, are oxide traps or oxide charges involved, is hole trapping important, are Si-H or also Si-O bonds broken et cetera.

In this section a model is proposed which was implemented into the numerical device simulator Minimos-NT. Figure 6.16 gives an overview of the involved processes. The model is able to

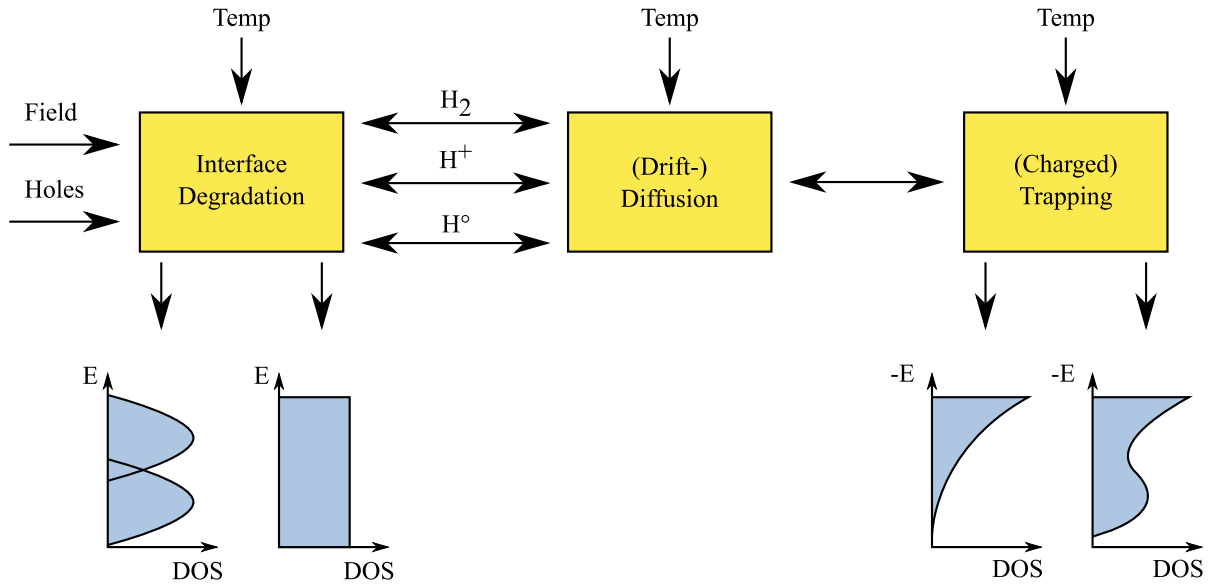


Figure 6.16: Schematic illustration of the model as implemented in Minimos-NT giving an overview of the involved processes.

achieve excellent agreement with measurement data on thick, pure SiO_2 dielectrics of high-voltage devices from our industry partner (Section 7.1) for both, the stress and the relaxation phase of NBTI and for different temperatures.

As the model comprises of many models and physical assumptions described in different parts of this work, I will collect all important equations in this section for the sake of clarity.

6.6.1 Interface Degradation

The first mechanism is the interface degradation as proposed in the reaction-diffusion model (Section 6.4). It is described in form of a balance equation between generation of electrically active interface traps (N_{it}) due to a forward term, and interface trap annealing as a result of a backward term as follows

$$\frac{\partial N_{\mathrm{it}}(t)}{\partial t} = \underbrace{k_{\mathrm{f}}(N_0 - N_{\mathrm{it}}(t))}_{\text{generation}} - \underbrace{k_{\mathrm{r}} N_{\mathrm{it}}(t) N_{\mathrm{X}}(0,t)^{1/a}}_{\text{annealing}} . \tag{6.31}$$

The important parameters in this equation are the forward reaction rate, k_{f}, the sum of passivated and broken interface states N_0, the reverse reaction rate k_{r}, the amount of the mobile species N_{X} directly available at the interface, and the kinetics exponent, a, giving the type of mobile species.

Assuming electrically neutral hydrogen, H^0, or hydrogen protons, H^+, the kinetics exponent a equals 1. In this case, for every generated N_{it} a mobile hydrogen, N_{X}, is released. To assume molecular hydrogen, H_2, as the mobile species the kinetics exponent equals 2.

The amount of interface states N_0 is strongly dependent on the wafer orientation and the process technology. It can greatly differ for different Si/SiO_2 interfaces. Reasonable values are often in the range of 10^{12}–$10^{14}\,cm^{-2}$ as described in Section 3.1. For reliable device operation it is, of course, of highest interest to keep this number as low as possible.

The forward reaction rate is dependent on the local electric field, E_{ox}, and the local hole concentration

$$k_f = k_{f,0}\,\frac{p_s}{p_{ref}}\,\exp\left(\frac{E_{ox}}{E_{ref}}\right)\exp\left(-\frac{E_{aF}}{k_B T}\right). \tag{6.32}$$

Here, k_B is the Boltzmann constant and T the temperature. The prefactor $k_{f,0}$, the reference hole concentration p_{ref}, and the reference electric field E_{ref} are model calibration parameters. The temperature dependence is introduced by an Arrhenius' law of the activation energy E_{aF}, which is also a calibration parameter. The influence of the electric field in this equation is exponential and change in the field has therefore a major impact on the forward reaction rate while the hole concentration has only little impact in agreement with experimental observations [101].

The reverse reaction rate is also introduced as Arrhenius activated with the activation energy E_{aR} and the prefactor $k_{r,0}$

$$k_r = k_{r,0}\,\exp\left(-\frac{E_{aR}}{k_B T}\right). \tag{6.33}$$

As described in Section 3.1, the energy distribution of electrically active interface traps N_{it} can be very important. When the Fermi-level is not very close to one of the band edges, the charge state of these traps differs depending on their density-of-states (DOS). The density of N_{it} is given as the integral of D_{it} across the whole band-gap

$$N_{it} = \int_{band-gap} D_{it}\,dE\,, \tag{6.34}$$

or, when considering the amphoteric nature of the traps with their double peaks (Section 3.1.1)

$$N_{it} = \int_{band-gap/2} D_{it}\,dE\,. \tag{6.35}$$

Several models for different DOS are available in Minimos-NT, where the most plausible according to literature are two peaks of Gaussian form as found in Section 3.1.2.

6.6.2 Drift-Diffusion

The next mechanism in the degradation process is the (drift-) diffusion of the hydrogen related species N_X from the Si/SiO_2 interface into the dielectric.

The convection of the electrically neutral species H^0 and H_2 is a purely diffusive process due to the concentration gradient

$$\frac{\partial N_X(x,t)}{\partial t} = D\exp\left(-\frac{E_{aD}}{k_B T}\right)\vec{\nabla}^2 N_X(x,t)\,. \tag{6.36}$$

Here, D is the diffusion constant. The temperature dependence was introduced using Arrhenius' law with the activation energy E_{aD}.

For the case of proton transport, H^+, an additional drift term has to be added to (6.36). In the typical NBT stress condition where a negative voltage is applied to the gate contact, this leads to quick removal of the positively charged protons from the interface. As a consequence more interface traps can break and the degradation is increased.

6.6.3 Trapping

To account for the dispersive transport theory as described in Section 6.4.3, the trapping and de-trapping of the mobile species in the dielectric is included in this model.

At each trap center in the dielectric a balance equation is solved to account for the trapping and de-trapping effects. The trap occupancy $\rho(E_{\mathrm{t}})$ is considered at each trap energy level E_{t} as

$$\frac{\partial \rho(E_{\mathrm{t}})}{\partial t} = \underbrace{\nu_0 \frac{N_{\mathrm{c}}}{N_{\mathrm{c0}}} \Big(g(E_{\mathrm{t}}) - \rho(E_{\mathrm{t}}) \Big)}_{\text{capture}} - \underbrace{\nu_0 \exp\left(\frac{E_{\mathrm{t}}}{\mathrm{k_B} T} \right) \rho(E_{\mathrm{t}})}_{\text{release}} . \tag{6.37}$$

Here, N_{c0} is the hydrogen effective density-of-states, ν_0 the release rate coefficient, N_{c} the amount of available, uncaptured hydrogen and $g(E_{\mathrm{t}})$ the trap density-of-states at this energy level. In the capture part of this equation the filling of traps is accounted for $(g(E_{\mathrm{t}}) - \rho(E_{\mathrm{t}}))$ so that the maximum possible amount of trapped hydrogen at each trap level equals the DOS. The release of hydrogen is dependent on the "depth" of the trap. Deeper traps have an exponential decrease in their de-trapping probability.

The trap energy E_{t} is zero at the mobility edge and below zero at the trap levels with more negative values for deeper traps. Therefore, with decreasing temperature the trap release rate decreases $(\exp(E_{\mathrm{t}}/\mathrm{k_B}T))$ and de-trapping becomes more unlikely. The trapping probability does not change in this model.

Within the model the charge state of trapped hydrogen can be chosen to model the formation of positively charged trapped protons, as described in [107].

The DOS of the trap levels can be selected from the model. Two plausible densities are schematically illustrated on the right hand side of Figure 6.16. An exponential distribution with the highest DOS at the band edge [88] or with an additional Gaussian peak in deeper states forming deep and therefore "slow" traps [106–108].

Parameter	Unit	Description
N_0	cm^{-2}	Total density of interface states (el. active and inactive)
a	1	Kinetics exponent automatically set according to diffusing species
$k_{\mathrm{f},0}$	s^{-1}	Forward reaction rate prefactor
p_{ref}	cm^{-3}	Reference hole concentration
E_{ref}	$\mathrm{V\ cm}^{-1}$	Reference electric field
E_{aF}	eV	Forward reaction activation energy
$k_{\mathrm{r},0}$	$\mathrm{cm}^3\ \mathrm{s}^{-1}$	Reverse reaction rate prefactor
E_{aR}	eV	Reverse reaction activation energy
D	$\mathrm{cm}^2\ \mathrm{s}^{-1}$	Diffusion coefficient
E_{aD}	eV	Diffusion activation energy
$N_{\mathrm{c}0}$	cm^{-3}	Hydrogen effective density-of-states
ν_0	s^{-1}	Trapping/de-trapping frequency

Table 6.1: NBTI model parameters.

Chapter 7

Case Studies

IN this chapter the effect of negative bias temperature instability on device and circuit performance is investigated. For this purpose the multi-dimensional device- and circuit simulator Minimos-NT [44] is used. As part of this work it has been extended by new models and capabilities to simulate and visualize the degradation mechanisms and their effects.

The calibration task has been performed using the gate oxide of a high voltage MOSFET as reference. The device has been stressed with different temperatures allowing the calibration of the full set of NBTI model parameters. To investigate the degradation at circuit level a CMOS inverter, a 6 transistor SRAM cell, and a five stage ring oscillator are simulated including their susceptibility to NBTI based degradation.

7.1 Power MOS Devices

Since the invention of semiconductor devices based on the bipolar technology, such as the bipolar transistor (BJT) and the thyristor, strong efforts have been made to increase the power handling capability of these devices in order to extend their applications. As the CMOS technology gained importance and process technology in the field of integrated circuits was surpassing the development of bipolar technology used for power devices new concepts were sought. In the 1970's the power MOSFET was introduced commercially and it was now possible to use the steady progress in CMOS technology also for the development of improved power devices such as the combined MOS/bipolar power devices in the 1980's.

The advantages of power devices based on MOS technology compared to bipolar technology are manifold:

- High input impedance due to the insulated gate contact.
- High switching speed as in contrast to BJTs. The n-channel power MOSFET is operating with electron transport which is inherently faster than the combined electron and hole transport the BJTs rely on.

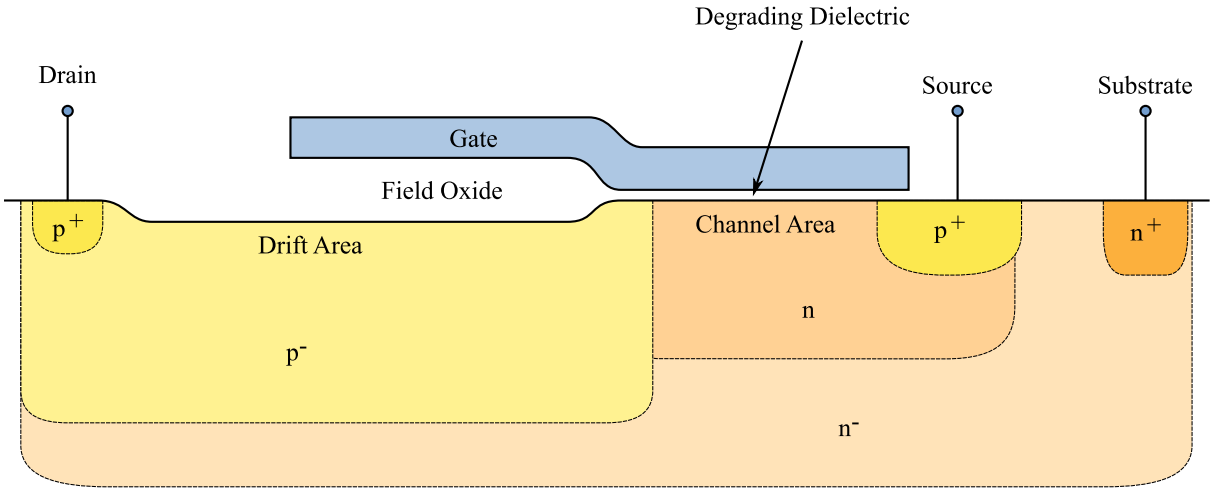


Figure 7.1: A p-channel lateral double-diffused MOS transistor for high-voltage applications. It is processed using CMOS technology and can, thus, be integrated in low-voltage CMOS circuitry.

- As the forward voltage drop in BJTs *decreases* with increasing temperature they suffer from thermal runaway. Using parallel devices would lead to the concentration of the current on a single device. Power MOSFETs have an *increasing* forward voltage drop with increasing temperatures and the current is automatically distributed evenly between parallel devices.
- They can be integrated with low voltage circuitry in a monolithic way.
- Larger safe operating area.

7.1.1 LDMOSFET

For the monolithic integration with low voltage circuitry the lateral double-diffused MOSFET (LDMOSFET) is the preferred device. Figure 7.1 gives a schematic of the device structure.

The channel of the LDMOSFET is formed by two lateral diffusion steps. First the channel doping is implanted in the source region. A following lateral diffusion step moves the dopants under the gate electrode. Then the source doping is implanted in the source region. By means of a second diffusion step the effective channel length can be adjusted. The big advantage is that the channel length is not directly dependent on the feature size of the process technology.

The geometry used for device simulation in this section has been generated using process simulation tools. The maximum voltages for this device are specified as $V_g = -25\,\mathrm{V}$ and $V_d = -50\,\mathrm{V}$ with the other terminals grounded. Figure 7.2 depicts the distribution of the electrostatic potential within the device at these conditions. It can be seen that the major voltage drop between source and drain occurs in the drift area of the device. The typical NBT stress conditions are with zero drain voltage $V_d = 0\,\mathrm{V}$. At this bias condition, the potential at the Si/SiO_2 interface

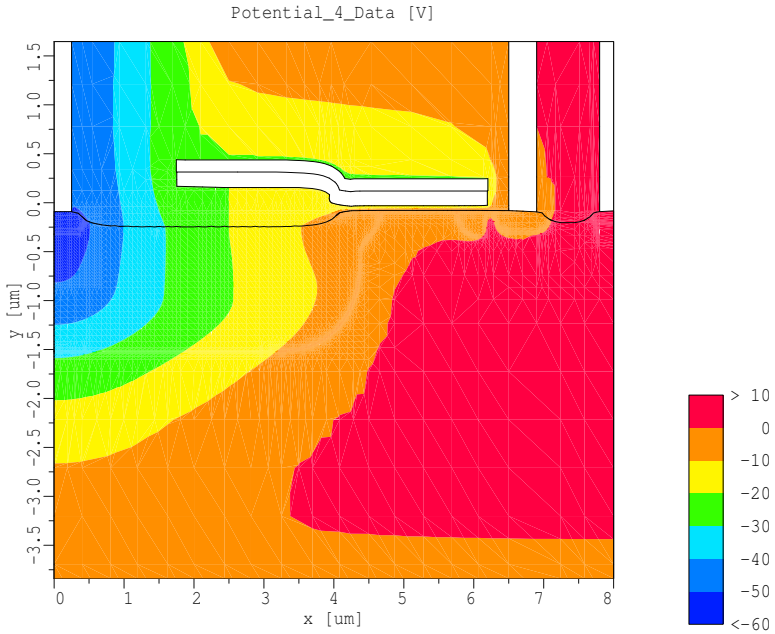


Figure 7.2: Simulation result showing the distribution of the electrostatic potential across a p-channel LDMOSFET. The transistor is turned on with $V_g = -25\,V$ and $V_d = -50\,V$. The major part of the drain to source voltage drop is consumed by the drift area of the device.

is nearly constant throughout the device (Figure 7.3).

Figure 7.4 gives the hole current density of the turned on transistor. In the channel area the current path is very close to the gate oxide, while in the drift region the current flow is deep in the substrate. This shows very well that the state of the silicon/dielectric interface is of highest importance in the gate region, while degradation in the thick oxide above the drift region will have less impact.

Figure 7.5 shows the electric field across the gate dielectric at NBT stress conditions. The field is very homogeneous.

Figure 7.6 gives one-dimensional plots of the electric field and the hole concentration at the Si/SiO_2 interface. These quantities are of key interest for the NBTI model as proposed in Section 6.6. Because of their even distribution it is reasonable to model gate dielectric degradation due to field dependent degradation mechanisms in one spatial dimension.

7.1.2 Parameter Extraction

The results from Section 7.1.1 propose very uniform conditions along the gate dielectric of LDMOSFETs at NBT stress conditions. The electric field, the hole concentration, and of course the oxide thickness are almost constant in both lateral dimensions. It is therefore possible to

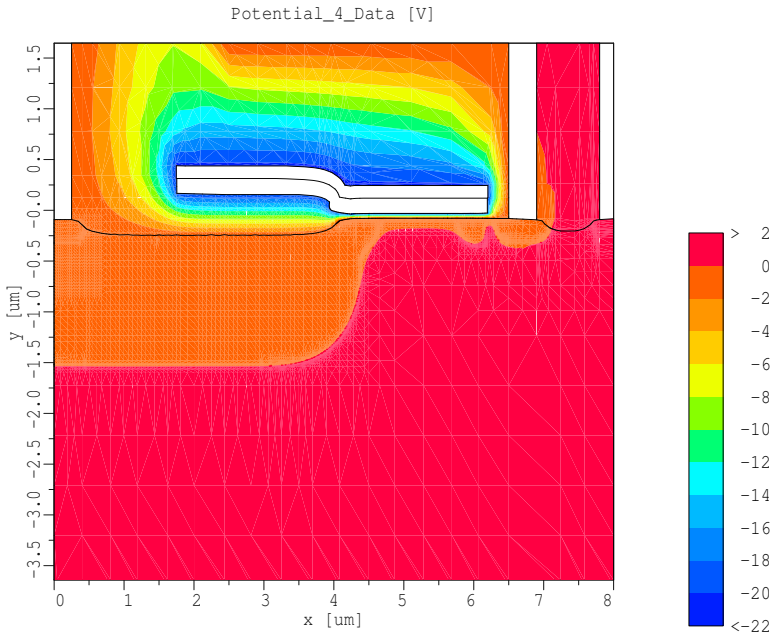


Figure 7.3: The same simulation as in Figure 7.2 but with $V_\mathrm{d} = 0\,\mathrm{V}$. This is the typical NBT stress situation.

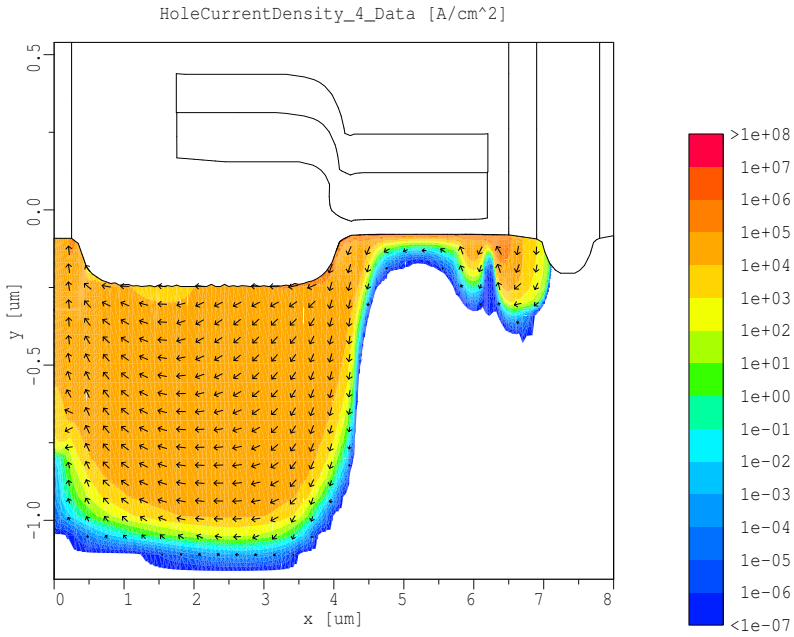


Figure 7.4: The hole current density across the device. Only in the channel area the current density has high values at the silicon/dielectric interface.

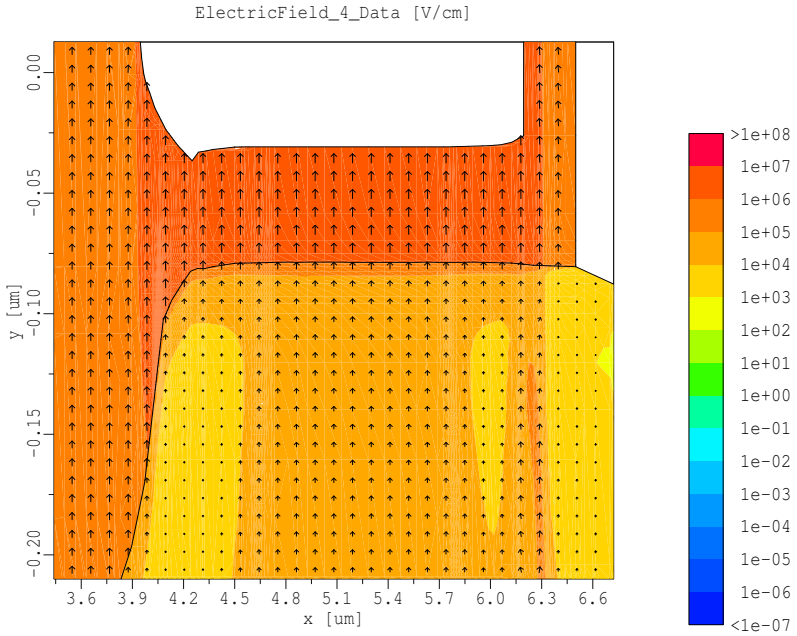
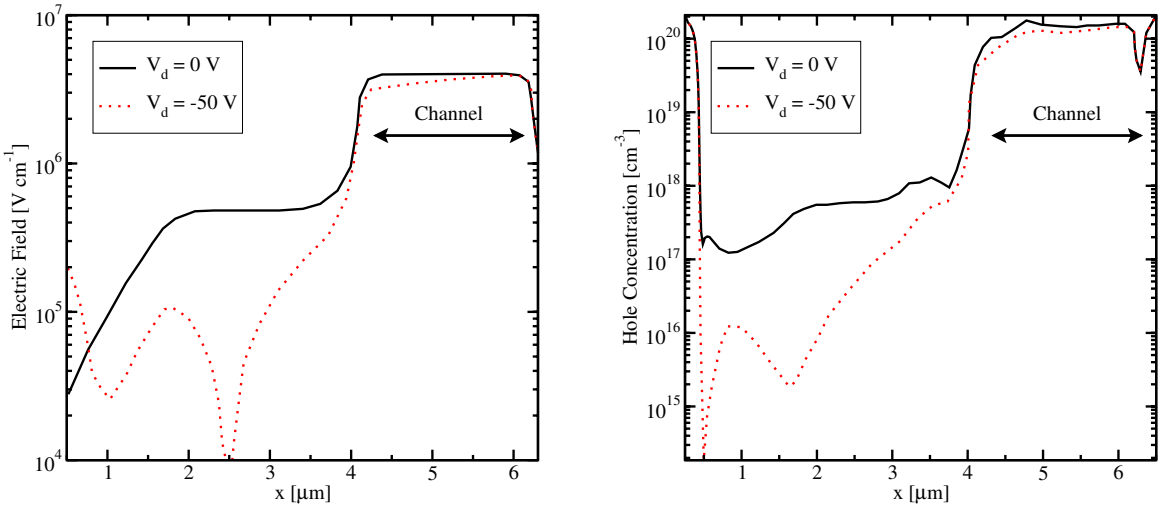


Figure 7.5: The electric field across the gate dielectric is very homogeneous at NBT stress conditions. It is therefore reasonable to model field dependent degradation mechanisms one-dimensionally.



(a) Electric field

(b) Hole concentration

Figure 7.6: The magnitude of the (a) electric field and the (b) hole concentration at the Si/SiO_2 interface. At NBT stress conditions ($V_d = 0\,V$) both quantities are very uniform in the channel area.

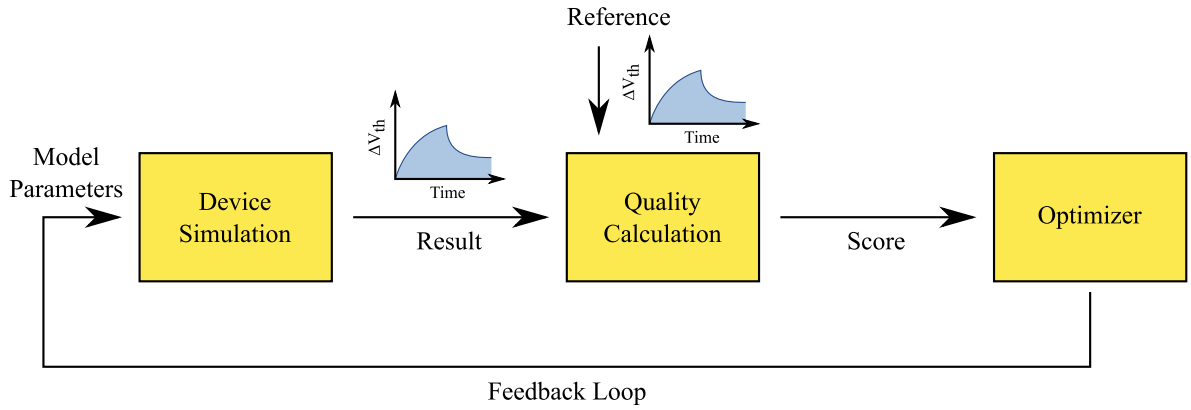


Figure 7.7: Procedure of a calibration process. The simulation results are compared with reference measurement data and a new parameter set is calculated for the next simulation run.

calibrate the NBTI model using a simplified device geometry.

For calibration of the NBTI model parameters a conventional p-channel transistor with large gate area has been processed. The oxide of this calibration structure is the same as in the original LDMOSFET. For simulation an identical geometry has been generated.

The procedure to extract the model parameters is schematically depicted in Figure 7.7. After setting up the device geometry an initial set of parameters must be specified. For each parameter a range has to be provided in which it can be adjusted by the optimizer. The duration of the whole optimization process is strongly dependent on these values, so care should be taken to find initial values as close as possible to the optimal calibration point.

In the next step the optimization loop can be started. The first device simulation is performed using the initial set of parameters. The resulting output is evaluated by comparing it to the measurement data. This comparison must result in a single number, the score of the simulation. It is calculated by comparing the simulation output to the measurement data. To generate a single number from the output curves a "least square fitting" criterion was used. Here, the sum of the squares of the offsets is calculated resulting in a score for the fit. A lower score means less offset between measurement and simulation and, thus, a better calibration.

From the score the optimizer generates a new set of model parameters which are then used for the next simulation run. A new score is obtained and this loop continues until the optimizer finds a minimum.

This calibration was first performed for the threshold voltage shift at 175 °C. To calibrate the temperature dependence a second calibration run was performed for 100 °C using the Arrhenius activation energies E_{aF}, E_{aR}, and E_{aD} as calibration parameters. Table 7.1 gives the NBTI model parameters found for the given device structure. Their definition is given in Section 6.6.

Parameter	Value	Unit
N_0	1.18×10^{13}	cm^{-2}
a	2	1
$k_{\mathrm{f},0}$	8.88×10^{-3}	s^{-1}
p_{ref}	1×10^{20}	cm^{-3}
E_{ref}	3×10^{6}	$\mathrm{V\ cm}^{-1}$
E_{aF}	0.2238	eV
$k_{\mathrm{r},0}$	3.3449×10^{-4}	$\mathrm{cm}^{3}\ \mathrm{s}^{-1}$
E_{aR}	-0.0938	eV
D	1.8165×10^{-10}	$\mathrm{cm}^{2}\ \mathrm{s}^{-1}$
E_{aD}	9.2608×10^{-3}	eV
$N_{\mathrm{c}0}$	6.9226×10^{26}	cm^{-3}
ν_0	1.0491×10^{12}	s^{-1}

Table 7.1: Calibration results.

7.1.3 Measurements and Simulation Results

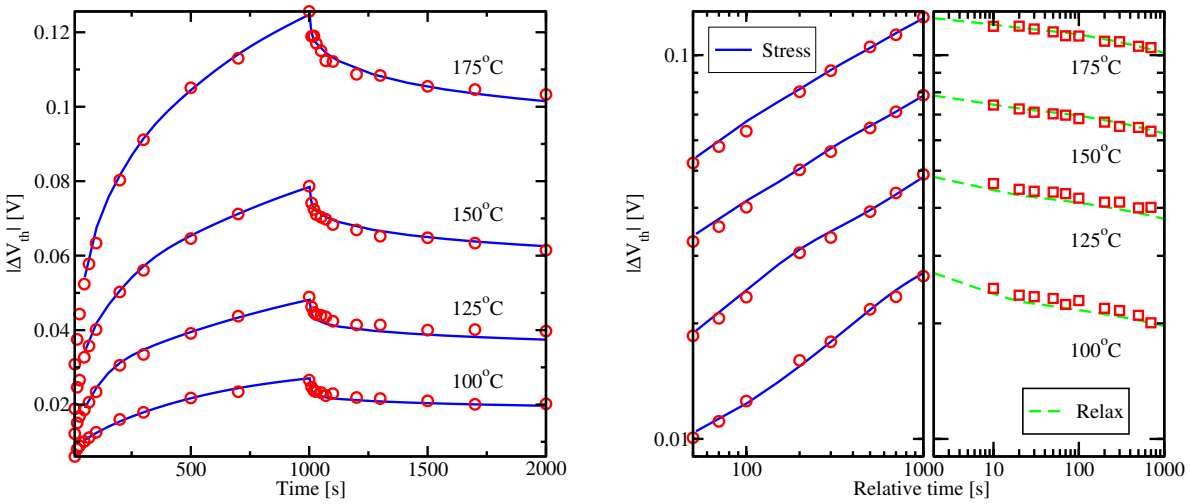
The device was stressed with $V_{\mathrm{g}} = -25\,\mathrm{V}$ for 1000 seconds and in the following the gate stress was relaxed for another 1000 seconds. This procedure was performed at 100, 125, 150, and 175 °C. To determine the threshold voltage shift ΔV_{th}, the stress was interrupted for two seconds at each measurement point to perform a gate voltage sweep from 0 V to −2 V (Section 6.3.1). During this period relaxation can be observed. Thus, it is crucial to include the recovery process in the simulation.

For the comparison of measurements to simulation results the complete dynamics of degradation and annealing during the measurement intervals have to be taken into account. The simulation run was therefore set-up to exactly mimic the measurement. Here, also the threshold voltage was extracted using a $I_{\mathrm{d}}/V_{\mathrm{g}}$ sweep to allow relaxation in the same way as in the measurements.

The calibration using the extended model (Section 6.6) shows excellent agreement with measurement data for a wide range of temperatures (Figure 7.8), which can be achieved using a single set of model parameters (Table 7.1).

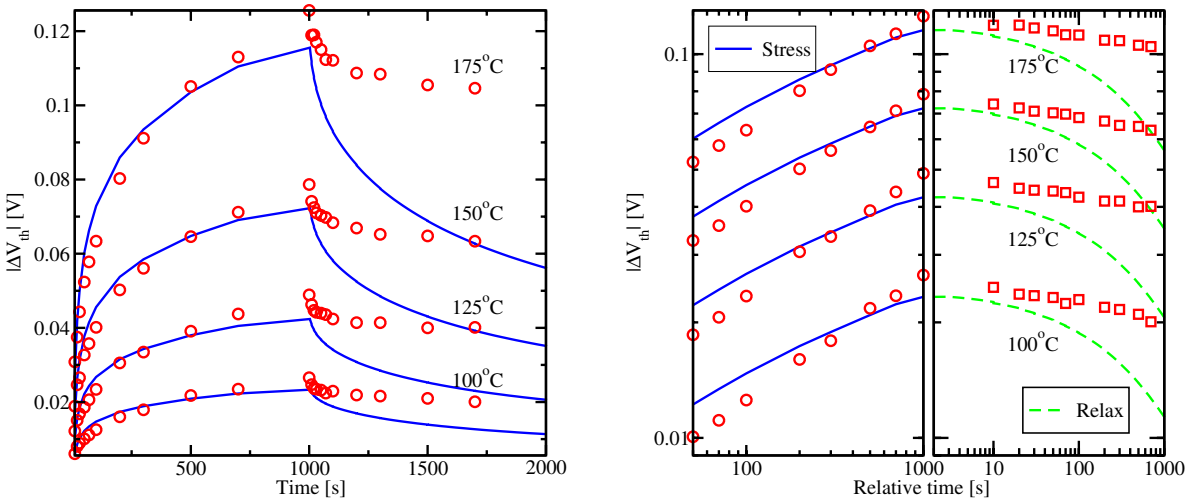
The pure reaction-diffusion model, in contrast, cannot reproduce the measurement data, as shown in Figure 7.9. In the stress phase a power law slope of $n = 0.23$ under-estimates the measured slope of $n = 0.31$. In the relaxation phase the amount of annealing is completely over-estimated.

The inability to calibrate the RD model to the measurement data emphasizes the importance to correctly model the dispersive hydrogen transport in the dielectric. Only the extended model (Section 6.6) is capable of correctly reproducing the data.



(a) Linear scale (b) Logarithmic scale

Figure 7.8: Result of the calibration process. Very good agreement is found with measurement data using the extended model (Section 6.6). The slope of $n = 0.31$ during the stress phase is very well matched with the trap-controlled transport model and the agreement in the relaxation phase is excellent.



(a) Linear scale (b) Logarithmic scale

Figure 7.9: The pure reaction-diffusion model without dispersion. This model obtains a slope of $n = 0.23$ and completely fails to reproduce the measurement data. Especially the annealing in the relaxation phase is drastically over-estimated.

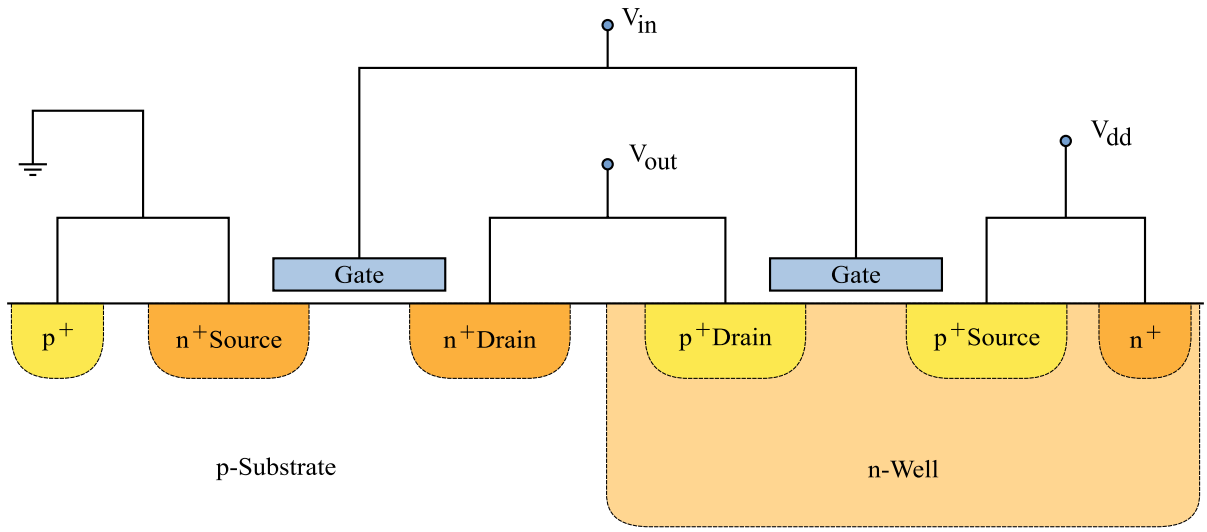


Figure 7.10: Schematic of a CMOS inverter as processed on a p-type silicon substrate. The effect of NBTI mainly impacts the p-channel MOSFET (right hand side transistor).

7.2 CMOS Inverter

For the investigation of circuit-level degradation a CMOS (complementary MOS) inverter is analyzed. A major advantage of CMOS technology is the ability to easily combine complementary transistors, n-channel and p-channel, on a single substrate. The CMOS inverter consists of the two transistor types which are processed and connected, as seen schematically in Figure 7.10.

The p-channel MOSFET relies on an n-type substrate. As commonly p-type wafers are used for processing, an additional n-type well implant is necessary. In this well, which is a deep region of n-type doping, the p-channel MOSFET is placed. As the p-substrate and the n-well junction is reverse biased, no significant current flows between these regions and the two transistors are isolated.

The output current of the p-channel MOSFET is typically much lower than the current of an n-channel MOSFET with similar dimensions and dopings. This is due to the lower carrier mobility of holes compared to electrons. As the characteristics of the complementary transistors should be as equal as possible, the width of the p-channel MOSFET is typically made larger to compensate the difference. In our example device the necessary geometry factor is 3.5 to obtain equal drain currents for equal gate biases.

Figure 7.11 gives the schematic of the CMOS inverter circuit. It can be seen that the gates are at the same bias V_{in} which means that they are always in a complementary state. When V_{in} is high, $V_{\text{in}} \approx V_{\text{dd}}$, the voltage between gate and substrate of the nMOS transistor is also approximately V_{dd} and the transistor is in on-state. The gate-substrate bias at the pMOS on the other side is nearly zero and the transistor is turned off. The output voltage V_{out} is pulled to ground, which is the low state. When the input voltage is in a high-state, the complementary

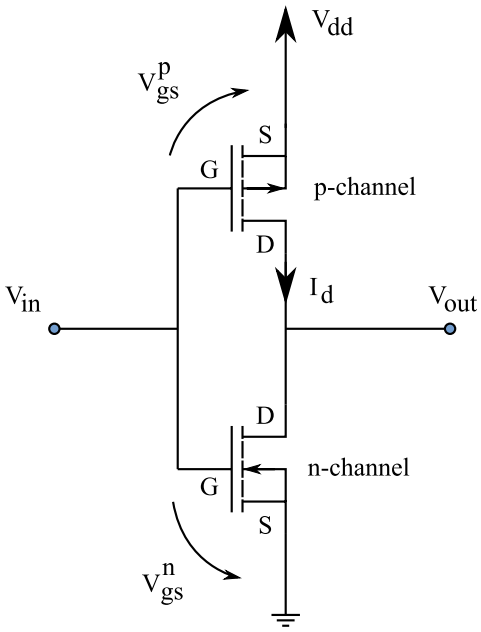


Figure 7.11: Schematic of a CMOS inverter circuit. In the stationary case the circuit does not consume any power when assuming perfect devices without leakage current. NBT stress is imposed on the p-channel device at $V_{\text{in}} = V_{\text{low}}$.

situation occurs and the pMOSFET is turned on while the nMOSFET is turned off. The output voltage is therefore pulled to V_{dd} which is the high-state. It is important to note that in both states, high and low, no static current flows through the inverter. This is of course only valid when assuming ideal devices with zero off- and leakage-currents.

Considering negative bias temperature instability, the worst stress conditions are imposed on the p-channel MOSFET at $V_{\text{in}} = V_{\text{low}}$. At this bias condition the pMOSFET is turned on, with approximately the same potential at the source and the drain $V_{\text{gs}} = V_{\text{gd}} = V_{\text{dd}}$ and negative gate to substrate voltage $V_{\text{gsub}} = -V_{\text{dd}}$.

7.2.1 Voltage Transfer Characteristics

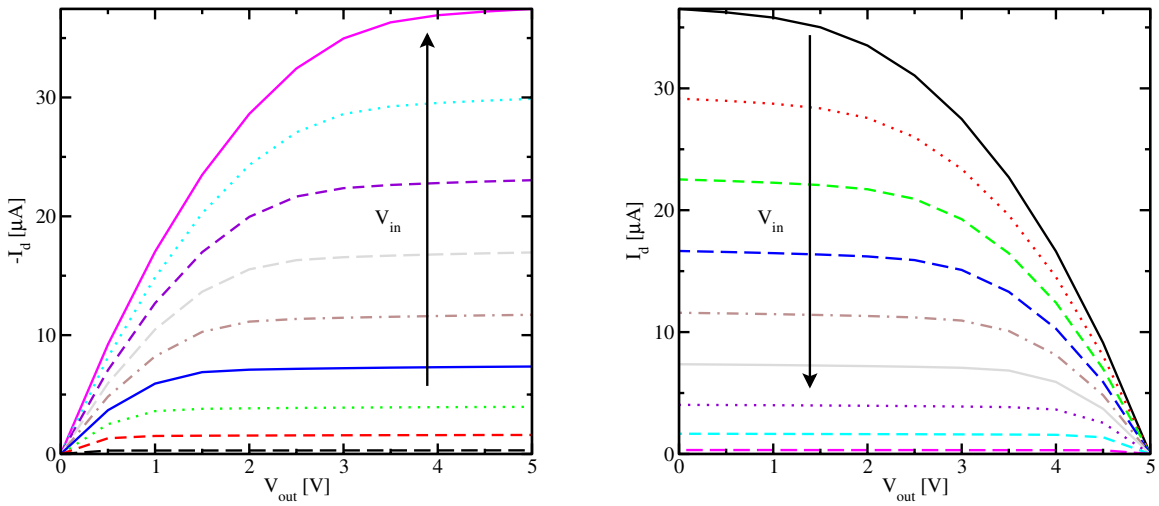
The voltage transfer characteristic (VTC) gives the response of the inverter circuit, V_{out}, to specific input voltages, V_{in}. It is a figure of merit for the static behavior of the inverter.

The gate-source voltage V_{gs} of the n-channel MOSFET is equal to V_{in} while the gate-source voltage of the p-channel MOSFET calculates as

$$V_{\text{gs}}^{\text{p}} = V_{\text{in}} - V_{\text{dd}}\,, \tag{7.1}$$

and the drain-source voltage V_{ds}^{p} of the pMOSFET can be expressed as

$$V_{\text{ds}}^{\text{p}} = V_{\text{ds}}^{\text{n}} - V_{\text{dd}}\,. \tag{7.2}$$



(a) n-channel MOSFET (b) p-channel MOSFET

Figure 7.12: Output characteristics of both transistors up to $V_{\text{in}} = V_{\text{dd}} = 5\,\text{V}$. The resulting drain currents in the inverter circuit must be equal for each V_{in}.

Looking at the output characteristics of the two transistors (Figure 7.12), and considering that the drain currents, I_{d}, of both transistors must be equal, the voltage transfer characteristic can be extracted, as seen on Figure 7.13. From this figure it is obvious that a shift of the output characteristics of one transistor can have big impact especially around the turn-over point of the VTC.

The mixed-mode of Minimos-NT allows to simulate the whole circuit while the device characteristics for each device are solved using the semiconductor device equations. Thus, the degradation of the p-channel MOSFET due to negative bias temperature instability can be accounted for in the circuit simulation.

7.2.2 Steady State Degradation

The voltage transfer characteristics of the unstressed inverter can be seen in Figure 7.14. The transition from the on to the off state is very well aligned around $V_{\text{dd}}/2$.

NBT stress has its highest impact on the p-channel MOSFET during low input $V_{\text{in}} = V_{\text{low}}$. At this condition the transistor has a gate to substrate voltage of approximately $-V_{\text{dd}}$. When the circuit is additionally subject to thermal stress, then the threshold voltage of the p-channel transistor is degraded. As the n-channel device has a much lower susceptibility to this type of stress (Section 6.3.7), the circuit loses its symmetry. As shown in Figure 7.15, is the switching point of the output potential moved to a lower input voltage. An interface trap density $N_{\text{it}} = 10^{12}\,\text{cm}^{-2}$, which is already a severely damaged interface (Chapter 3), reduces the switching point by more than 1 V.

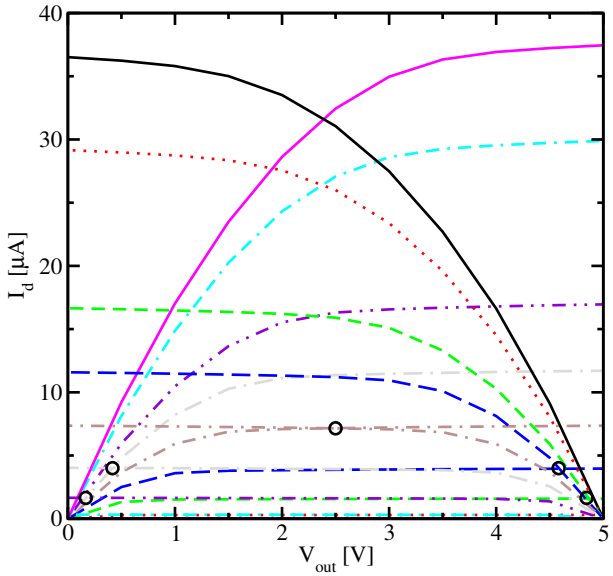


Figure 7.13: Extraction of the voltage transfer characteristics of a CMOS inverter. The drain currents of both transistors must be equal. Therefore, the intersection of the output characteristics of both transistors for each input voltage V_{in} give the output voltage V_{out}. The circles mark five points of the voltage transfer characteristics.

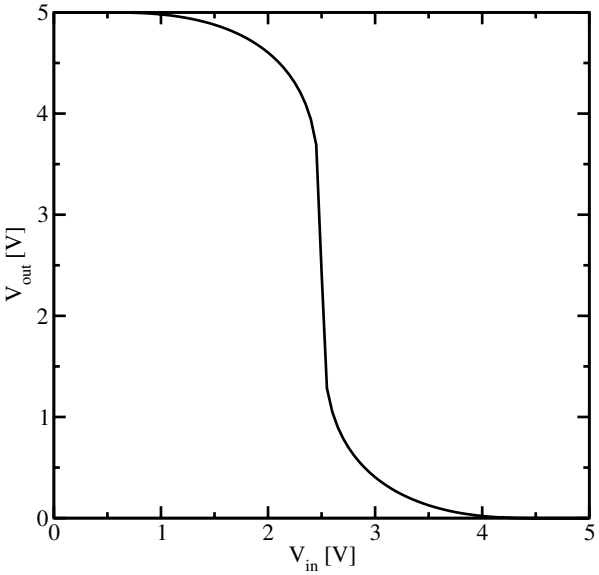


Figure 7.14: Voltage transfer characteristics of the CMOS inverter without degradation. The transition from $V_{\text{out}} = V_{\text{high}}$ to $V_{\text{out}} = V_{\text{low}}$ is symmetric and very well centered around $V_{\text{high}}/2$.

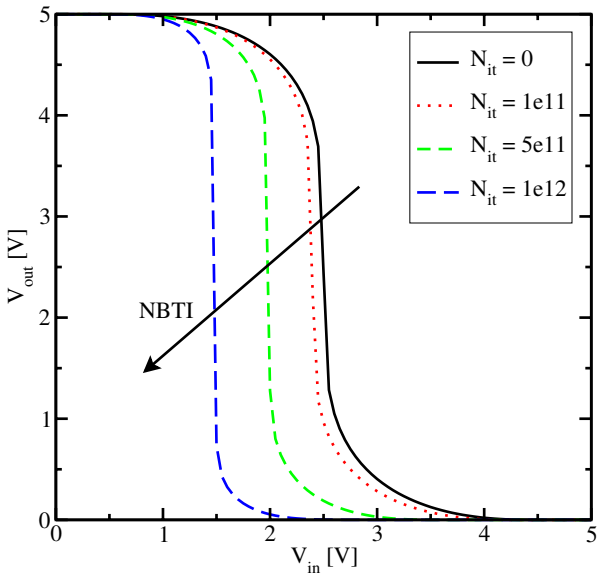


Figure 7.15: Degrading voltage transfer characteristics due to NBTI. Only the threshold voltage of the p-channel MOSFET shifts. The CMOS inverter does not switch symmetrically anymore and the switching point shifts to lower input voltages V_{in}.

7.2.3 Transient Behavior

Not only in the stationary case does the degradation influence the circuit performance. Transient simulations show (Figure 7.16 and 7.17) that the switching behavior of a circuit comprising a degraded p-channel MOSFET is different. This must be kept in mind when designing timing-critical CMOS circuits.

At high input, $V_{\text{in}} = V_{\text{high}}$, the p-channel MOSFET is turned off and the n-channel device turned on, pulling the output voltage to ground, $V_{\text{out}} = V_{\text{low}}$. Switching to $V_{\text{in}} = V_{\text{low}}$ turns off the n-channel device and on the p-channel device. The switching speed depends on the magnitude of the gate overdrive, $V_{\text{gsub}} - V_{\text{th}}$. An NBT degraded pMOS transistor has a lower (more negative) threshold voltage, therefore a lower gate overdrive and is turned on slower. The result is a slower CMOS inverter when turning the output $V_{\text{out}} = V_{\text{high}}$, as seen in Figure 7.16.

The opposite case, turning the inverter from $V_{\text{in}} = V_{\text{low}}$ to $V_{\text{in}} = V_{\text{high}}$ is completely different, as seen in Figure 7.17. Here, the p-channel transistor is switched from on to off. In this case the gate overdrive equals the magnitude of the threshold voltage. The degraded device, with its more negative V_{th} is driven into stronger inversion and can, thus, be turned off more quickly.

7.3 6T SRAM Cell

Static random access memory (SRAM) can retain its stored information as long as power is supplied. This is in contrast to dynamic RAM (DRAM) where periodic refreshes are necessary

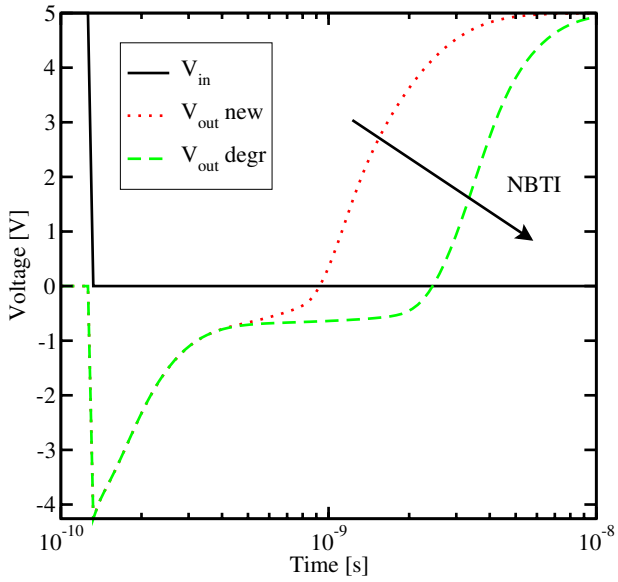


Figure 7.16: Transient switching-on behavior of the CMOS inverter. Because of the threshold voltage shift of the p-channel device, the degraded circuit needs more time to reach $V_{\text{out}} = V_{\text{high}}$.

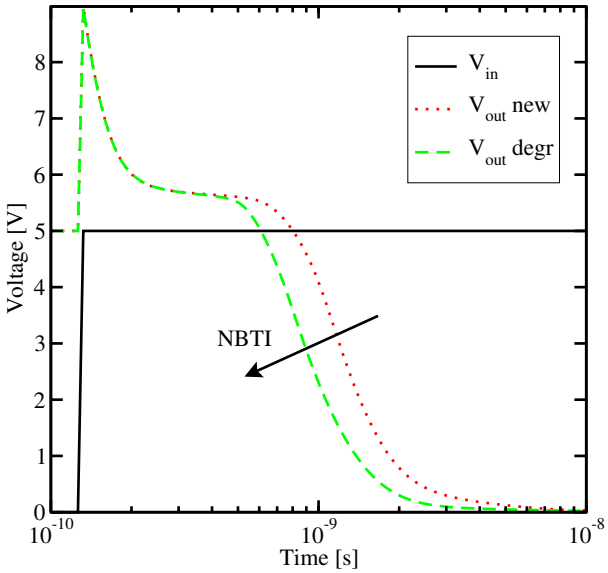


Figure 7.17: Transient switching-off behavior of the CMOS inverter. When switching the input from low to high state, the degraded circuit even outperforms the fresh circuit. This is due to the p-channel device turning off at lower gate voltage as the threshold voltage is shifted to a more negative voltage.

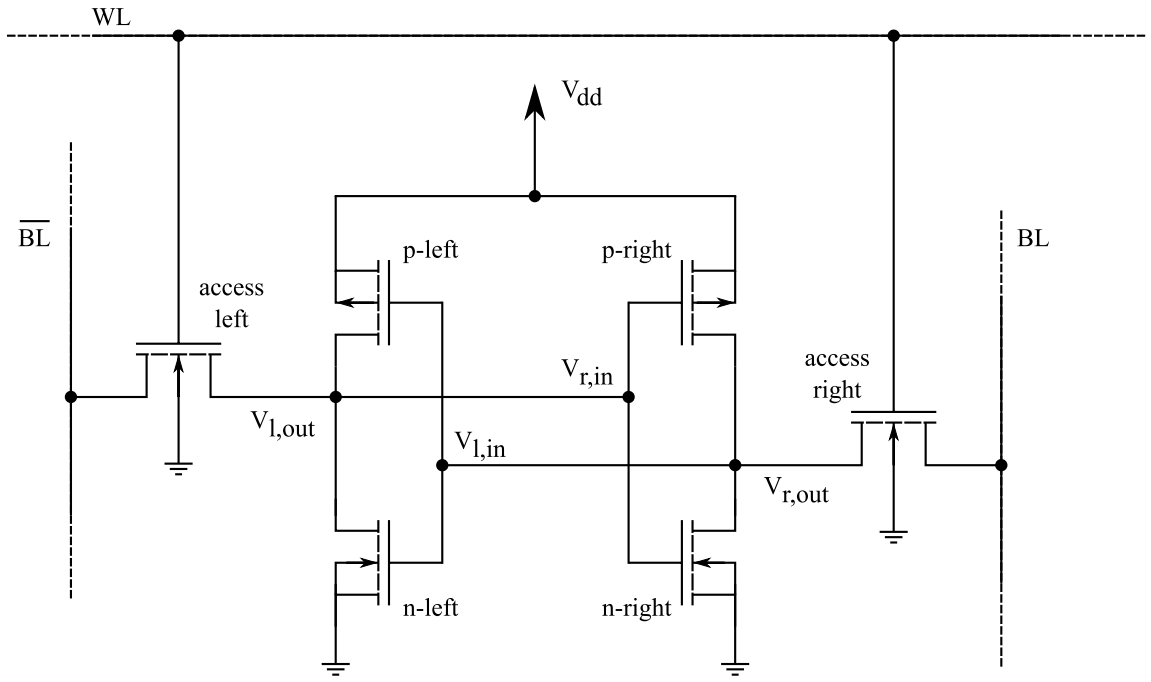


Figure 7.18: Circuit of a 6 transistor SRAM cell. It consists of two CMOS inverters and two access MOSFETs. NBT stress mainly affects the p-channel transistors.

or non-volatile memory where no power needs to be supplied for data retention, as for example flash memory. The term "random access" means that in an array of SRAM cells each cell can be read or written in any order, no matter which cell was last accessed.

The structure of a 6 transistor SRAM cell, storing one bit of information, can be seen in Figure 7.18. The core of the cell is formed by two CMOS inverters, where the output potential of each inverter V_{out} is fed as input into the other V_{in}. This feedback loop stabilizes the inverters to their respective state.

The access transistors and the word and bit lines, WL and BL, are used to read and write from or to the cell. In standby mode the word line is low, turning the access transistors off. In this state the inverters are in complementary state. When the p-channel MOSFET of the left inverter is turned on, the potential $V_{\text{l,out}}$ is high and the p-channel MOSFET of inverter two is turned off, $V_{\text{r,out}}$ is low.

To write information the data is imposed on the bit line and the inverse data on the inverse bit line, $\overline{\text{BL}}$. Then the access transistors are turned on by setting the word line to high. As the driver of the bit lines is much stronger it can assert the inverter transistors. As soon as the information is stored in the inverters, the access transistors can be turned off and the information in the inverter is preserved.

For reading the word line is turned on to activate the access transistors while the information is sensed at the bit lines.

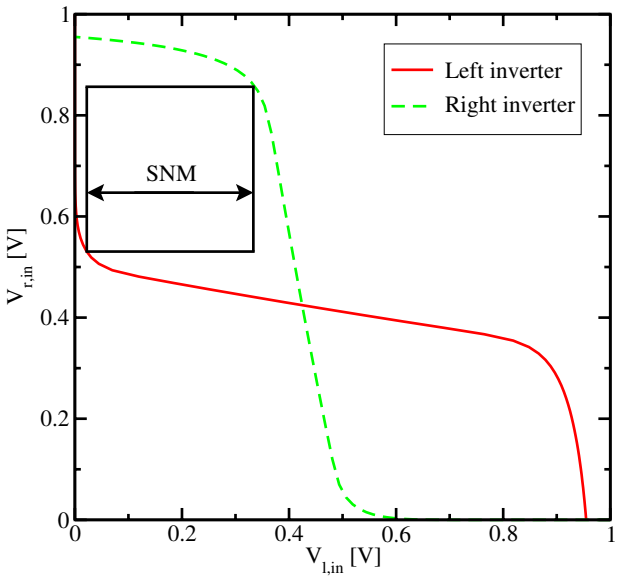


Figure 7.19: Static noise margin (SNM) of the unstressed SRAM cell. The window in the "butterfly" curve illustrates the hardiness against DC noise.

7.3.1 Static Noise Margin

A key figure of merit for an SRAM cell is its static noise margin (SNM). It can be extracted by nesting the largest possible square in the two voltage transfer curves (VTC) of the involved CMOS inverters, as seen in Figure 7.19. The SNM is defined as the side-length of the square, given in volts. When an external DC noise is larger than the SNM, the state of the SRAM cell can change and data is lost.

7.3.2 Steady State Degradation

The effect of NBTI impacts one or both p-channel MOSFETs in the SRAM cell, depending on the charge state during temperature stress. The result is a degraded VTC and, therefore, a degraded SNM.

Figure 7.20 shows the impact of NBTI on an SRAM cell. Here, only one of the p-channel MOSFETs is degraded. This is the typical situation when the information stored in the cell does not change frequently. An interface trap density of $N_{\mathrm{it}} = 10^{11}\,\mathrm{cm}^{-2}$ is enough to drastically reduce the SNM. At $N_{\mathrm{it}} = 3 \times 10^{11}\,\mathrm{cm}^{-2}$ the window completely disappears and the SRAM cell loses its functionality.

Assumed degradation of both p-channel MOSFETs results in shift of the VTCs of both inverters (Figure 7.21). This leads to a complete degeneration of the original VTCs.

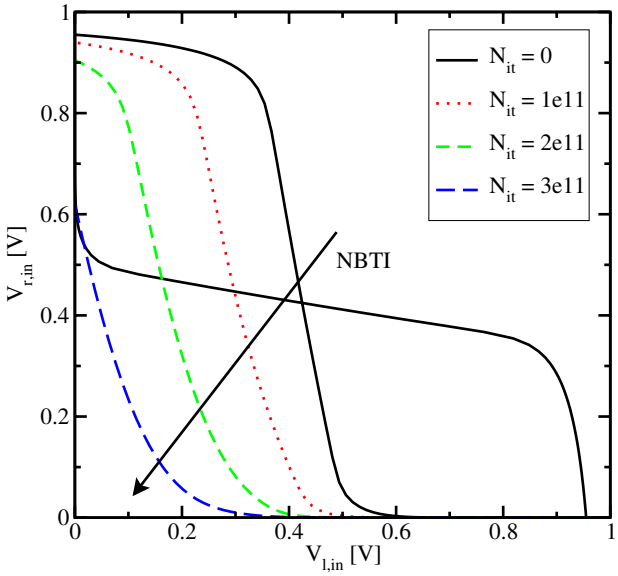


Figure 7.20: Degradation of one p-channel MOSFET in an SRAM cell. An interface trap density $N_{\mathrm{it}} = 3 \times 10^{11}\,\mathrm{cm}^{-2}$ causes read failure even at zero noise.

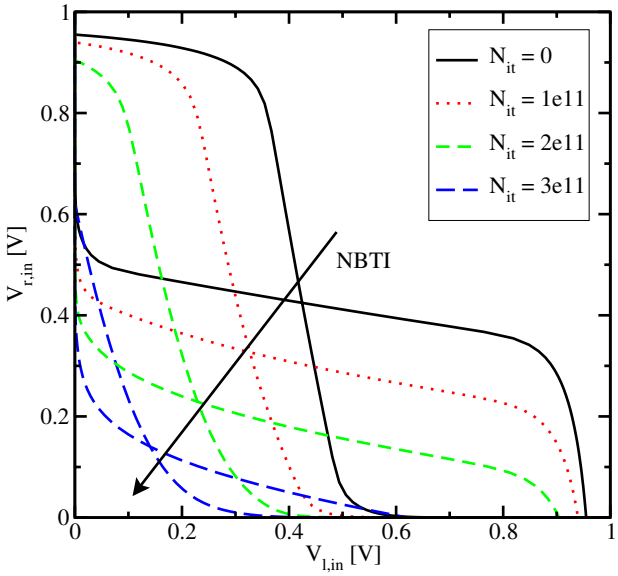


Figure 7.21: Degradation of both p-channel MOSFET in an SRAM cell. The original "butterfly" curve is completely degenerated.

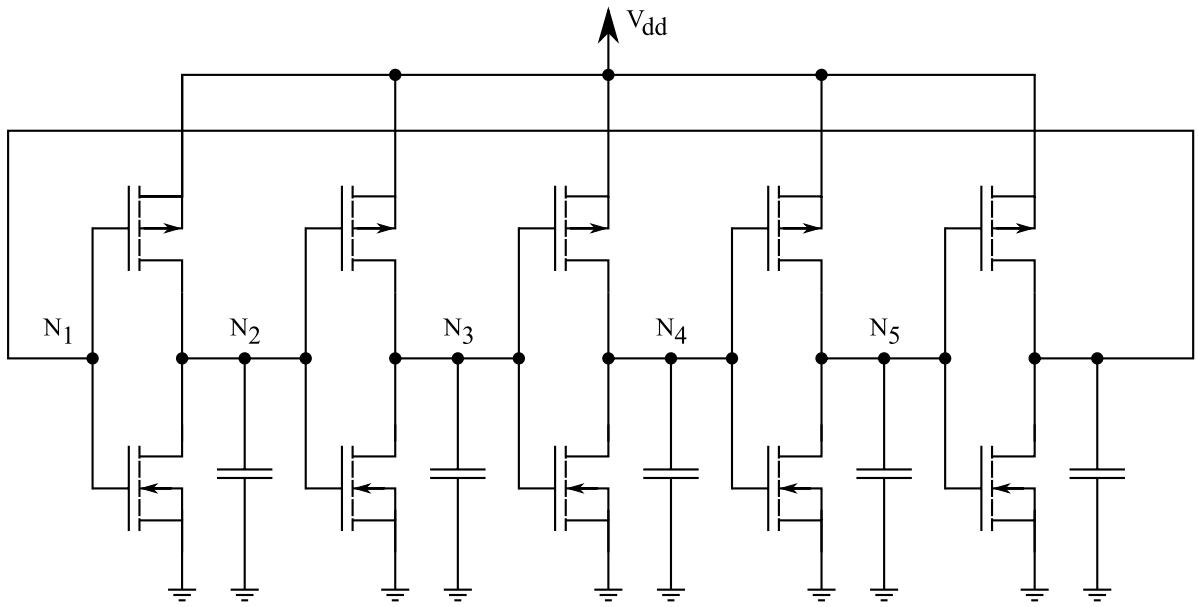


Figure 7.22: Circuit of a five stage ring oscillator. It consists of five CMOS inverters with coupled in- and outputs. NBT stress mainly affects the p-channel transistors.

7.4 Ring Oscillator

In this section a five stage CMOS ring oscillator, as depicted in Figure 7.22, is investigated with regard to NBTI degradation. A ring oscillator comprises of an odd number of CMOS inverters. The output of each inverter is used as input for the next one. The last output is fed back to the first inverter. Because of the delay time of each stage the whole circuit spontaneously starts oscillating at a certain frequency. The frequency f depends on the number of stages n and the delay time of the inverters τ as follows

$$f = \frac{1}{2n\tau}\,. \tag{7.3}$$

In the simulation the initial condition for the node voltages V_{N1}–V_{N5} has to be defined. In the first time step the voltages at node N1, N3, and N4 are forced to $V = 5\,\mathrm{V}$ and the others to $V = 0\,\mathrm{V}$. The voltages quickly reach their oscillation voltages.

Figure 7.23 shows the transient oscillation of the inverter. This non-degraded circuit oscillates with a frequency of $f = 17.8\,\mathrm{MHz}$. Using (7.3) the delay time of the inverters calculates as $\tau = 5.6\,\mathrm{ns}$.

7.4.1 Frequency Degradation

Negative bias temperature instability mainly affects the p-channel transistors in the inverter circuits. Their level of degradation is approximately equal because of the identical stress conditions in the oscillating circuit.

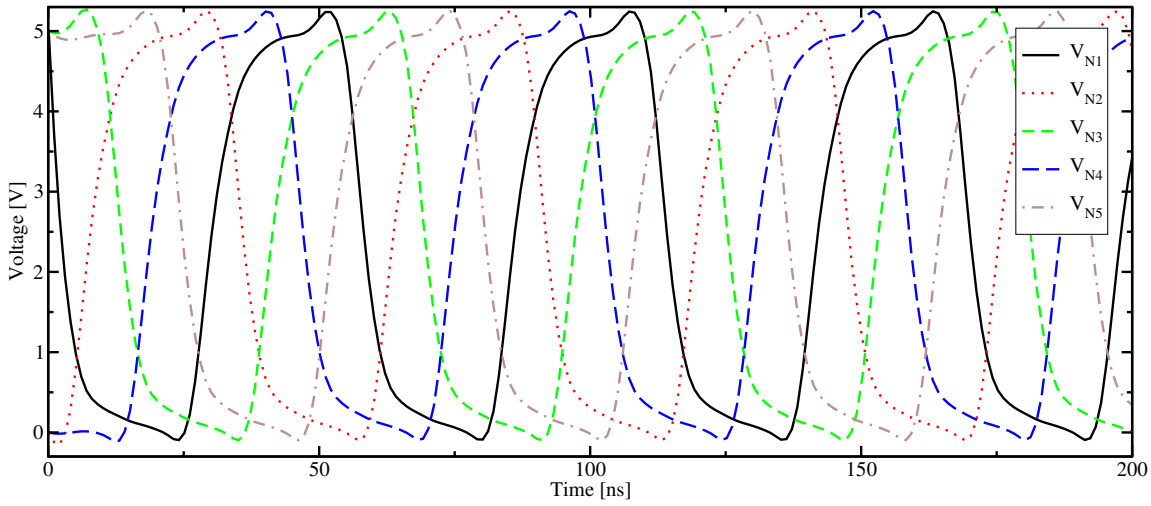


Figure 7.23: Transient oscillation in the simulation run. The circuit oscillates with a frequency of $f = 17.8\,\mathrm{MHz}$.

NBTI leads to an increased absolute threshold voltage, $|V_{\mathrm{th}}|$. This, in turn, reduces the gate overdrive required to turn on the p-channel transistors. When turning the transistor off, on the other hand side, the overdrive is increased and this process is therefore performed faster, as shown in Section 7.2.3. Still, as the lower turn-on speed dominates, the inverter's delay time increases, $\tau + \Delta\tau$,

$$f = \frac{1}{2n(\tau + \Delta\tau)}\,, \tag{7.4}$$

reducing the oscillation frequency.

Figure 7.24 depicts the oscillation voltage of the unstressed circuit and after degradation due to NBTI. A clear reduction of the oscillation frequency can be observed. With a very large interface degradation $N_{\mathrm{it}} = 1 \times 10^{12}$ the oscillation frequency is reduced from $f = 17.8\,\mathrm{MHz}$ to $f = 9.5\,\mathrm{MHz}$. The delay time of the five inverters is increased by $\Delta\tau = 4.9\,\mathrm{ns}$.

The evolution of the frequency with NBTI induced interface trap generation can be seen in Figure 7.25. Increasing degradation leads to a reduced oscillation frequency of the ring oscillator, as predicted by (7.4).

7.4.2 Transient Degradation

To evaluate the effect of long time NBT stress on the ring oscillator's frequency degradation a transient NBTI simulation of a single p-channel MOSFET has been performed. Figure 7.26 shows the resulting interface trap density N_{it} at constant voltage and constant temperature stress for 10 years using the new model. When assuming a frequency degradation of 5 % to be within the circuits design rules, Figure 7.26(a) shows very well that at temperatures above 100° C and with a stress voltage of $V_{\mathrm{g}} = -25\,\mathrm{V}$ the degradation exceeds this limit within 10

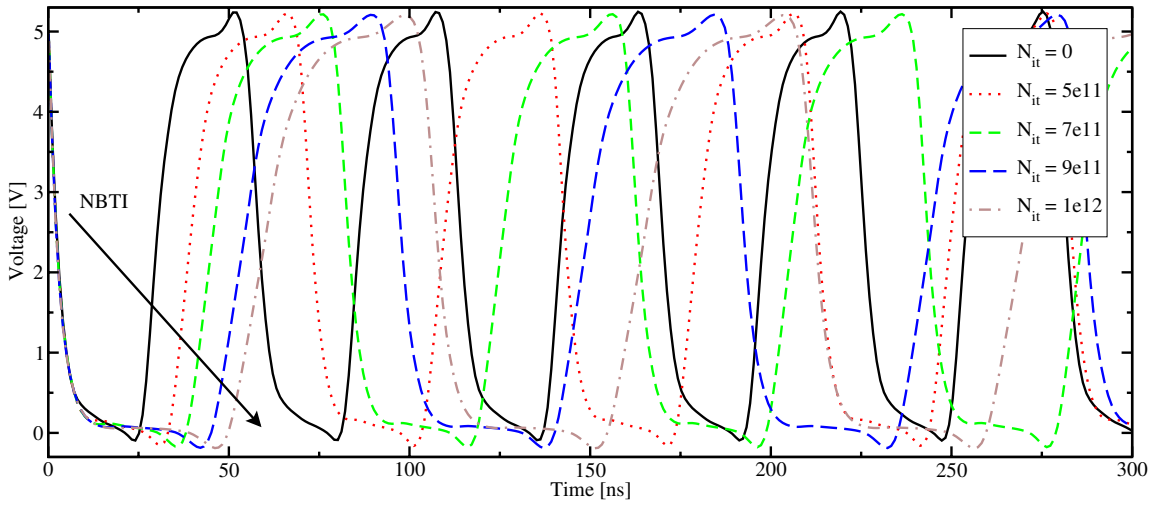
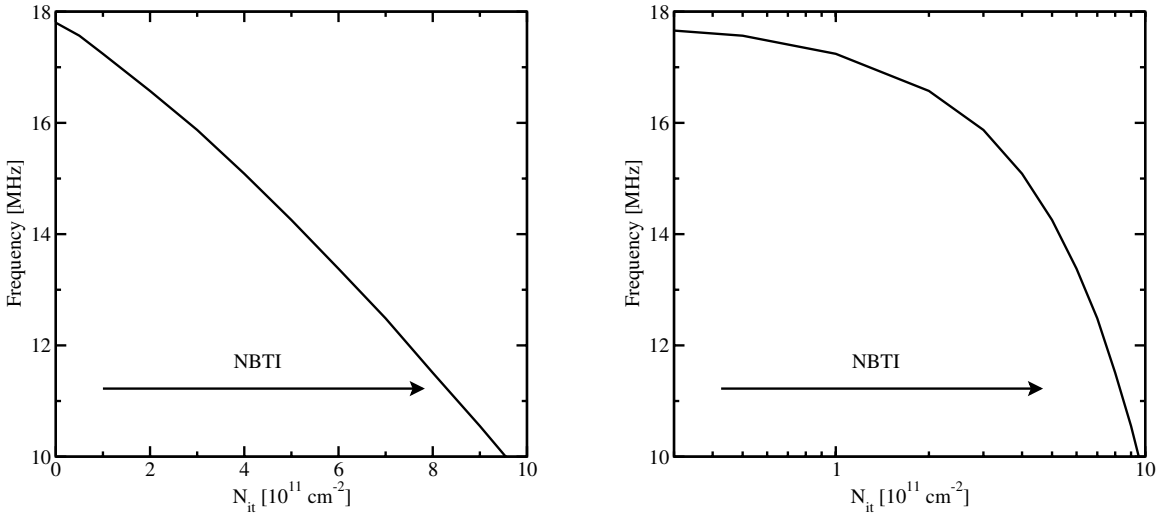
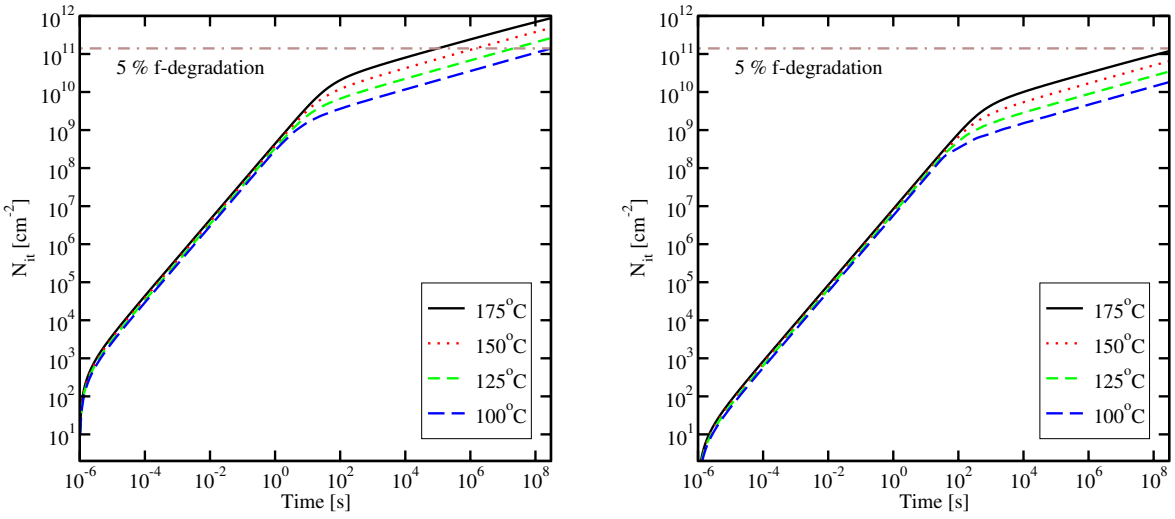


Figure 7.24: Influence of NBTI on a CMOS ring oscillator. Because of the threshold voltage shift due to NBT stress the circuit's oscillation frequency is reduced.



(a) Linear scale (b) Logarithmic scale

Figure 7.25: Frequency degradation versus N_{it}. As predicted by (7.4) the frequency of the ring oscillator is drastically reduced with increasing NBTI degradation.



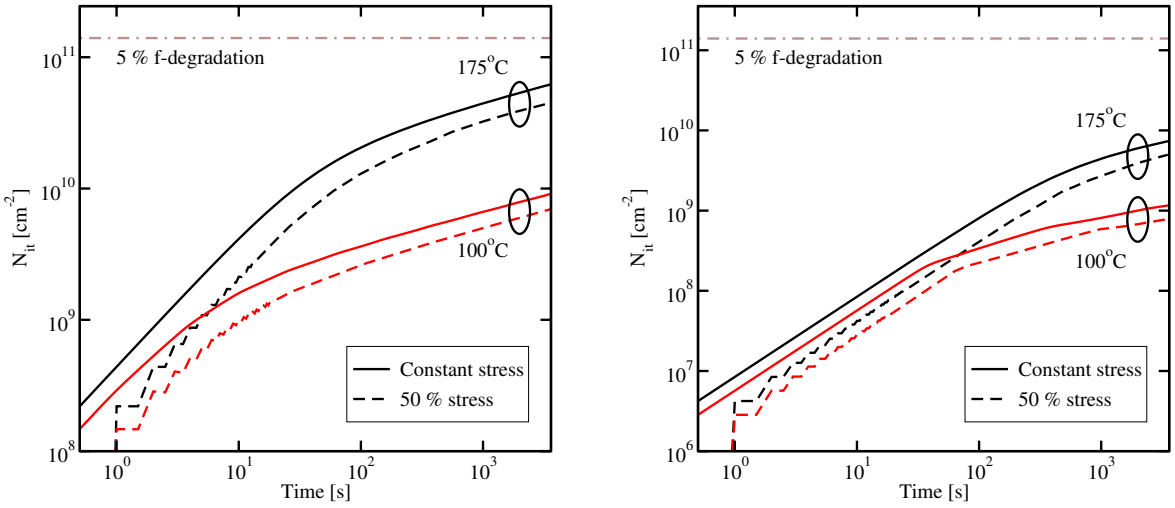
(a) $V_g = -25\,V$ (b) $V_g = -5\,V$

Figure 7.26: p-channel MOSFET degradation simulation at constant NBT stress for 10 years using the new model. A 5 % frequency degradation of the ring oscillator is observed at $N_{it} = 1.4 \times 10^{11}\,cm^{-2}$. Only at (a) 25 V gate stress and more than 100°C the degradation threshold is reached within 10 years. At regular operating voltage (b) the frequency degradation stays within 5 % in 10 years.

years (3×10^8 s). At regular operating voltage with $V_g = -5\,V$ the degradation limit is not reached.

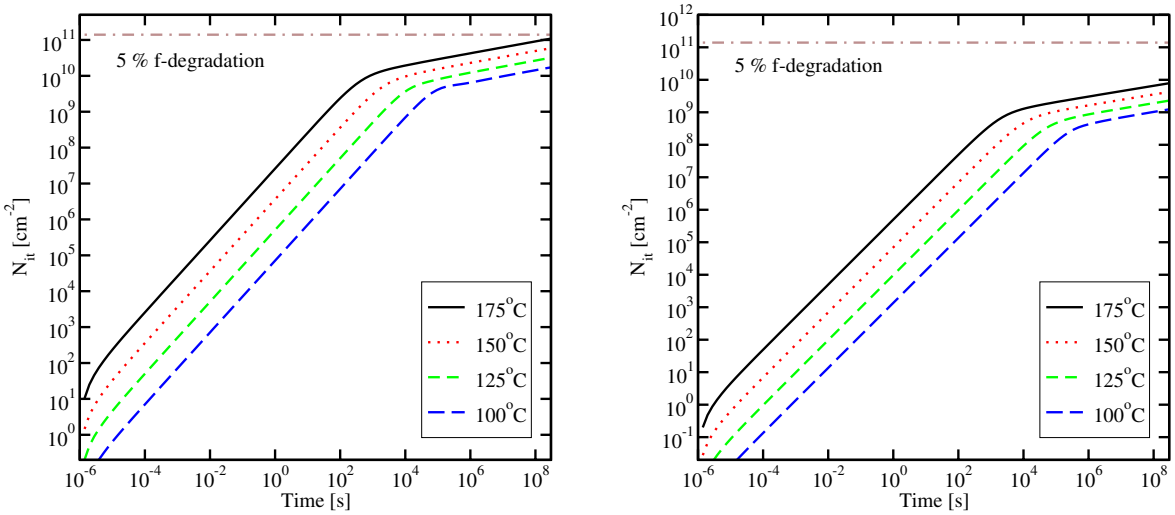
It has to be considered, though, that for these simulations constant stress was assumed. As shown in Chapter 6 NBTI has a recovery effect when the stress conditions are removed. Therefore, Figure 7.27 gives the resulting degradation at periodic stress with 1 Hz oscillation frequency for 1 hour. It is not possible to simulate the whole 10 years of lifetime, as for each second at least two simulations have to be performed, but the trend is clear. The overall degradation is reduced for periodic stress voltages and stays below the degradation at permanent stress conditions. This enhances the lifetime as defined above considerably.

The same simulations have been performed using the standard reaction-diffusion model (Figure 7.28 and 7.29). The estimated degradation after 10 years stress is clearly different and even lies beyond the 5 % border for all temperatures and voltages. This result emphasizes how important the use of a correct model and the right model parameters is for long time predictions.



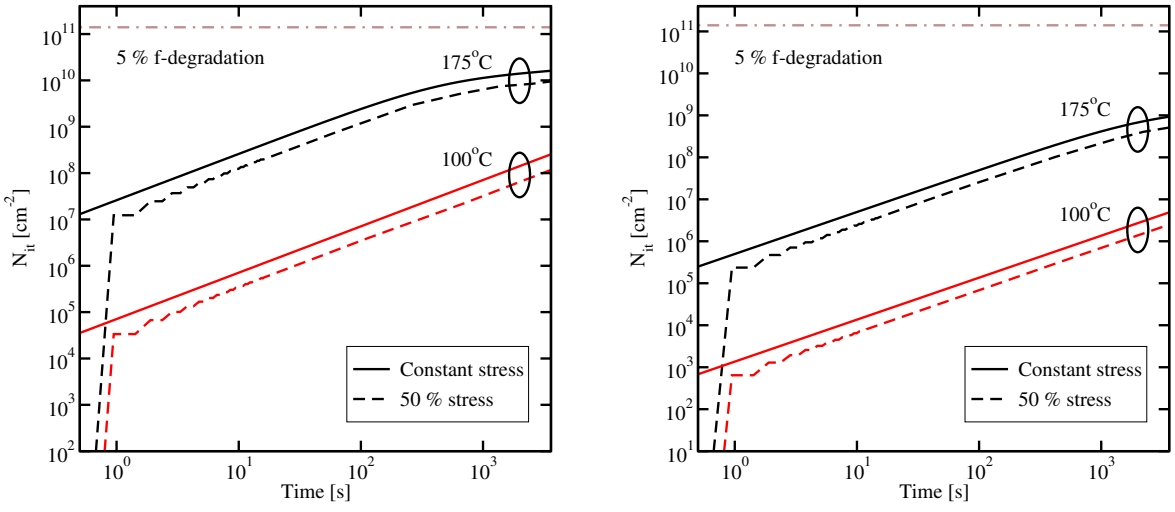
(a) $V_g = -25\,V$ (b) $V_g = -5\,V$

Figure 7.27: p-channel MOSFET degradation simulation at dynamic NBT stress for 1 hour using the new model. The stress voltage oscillates with 1 Hz and a duty cycle of 50 % (dashed lines) and with constant stress (solid lines) have been used. The level of degradation with oscillating stress is always below constant stress and increases the lifetime.



(a) $V_g = -25\,V$ (b) $V_g = -5\,V$

Figure 7.28: p-channel MOSFET degradation simulation at constant NBT stress for 10 years using the standard reaction-diffusion model. The simulation shows how crucial the use of the correct model is for long time predictions. The R-D model predicts that the ring oscillator would hardly reach the 5 % degradation within 10 years even at the harshest stress conditions.



(a) $V_g = -25\,V$ (b) $V_g = -5\,V$

Figure 7.29: p-channel MOSFET degradation simulation at dynamic NBT stress for 1 hour using the reaction-diffusion model.

Chapter 8

Summary and Conclusions

TECHNOLOGY computer aided design (TCAD) has evolved into an excellent means to aid the development of semiconductor device fabrication processes. Therefore TCAD has gained tremendously in importance and foundries widely use it to improve not only their processes but also their understanding of the physical quantities inside the devices, which measurements cannot capture. The foundation of device simulation are the basic semiconductor equations and they have proven to provide accurate results for models discussed in this work, where the carrier gas can be assumed to be in quasi-equilibrium with the lattice. For degradation mechanisms like hot carrier degradation the device equations have to be replaced with more sophisticated models which can also account for non-equilibrium dynamics.

In this thesis a look beyond the basic semiconductor equations in the context of reliability issues is taken because not only the performance of unstressed devices is important. Due to the constant geometric shrinking and the use of new materials, many degradation mechanisms which were formerly of minor importance, are now coming to the forefront of interest. In this work, two of those degradation mechanisms, namely the dielectric leakage currents of highly degraded dielectrics and negative bias temperature instability (NBTI), have been investigated. Through implementing the models in a numerical device and circuit simulator and coupling them to the semiconductor device equations, self-consistent simulations become possible. The device engineer can not only evaluate the normal operation of an improved or newly designed device concept, he can also investigate the susceptibility to those two degradation mechanisms in the same simulation run. By having insight into the device and the physical quantities causing the degradation he can optimize the structure not only for fast but also for reliable operation.

Especially NBTI, where the main focus of this work lies upon, has obtained tremendous scientific and industrial interest in recent years. As the exact physical mechanisms responsible for the degradation are still not completely identified and researchers are doing their best to promote the understanding, considerable progress has been made and new insights are gained nearly every month. Thus, it might be interesting to have another look at the topic in one or two years and further improve the available models.

Bibliography

[1] G. E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics*, vol. 38, no. 8, pp. 114–117, 1965.

[2] J. C. Maxwell, "A Dynamical Theory of the Electromagnetic Field," *Royal Society Transactions*, vol. CLV, 1864.

[3] W. VanRoosbroeck, "Theory of Flow of Electrons and Holes in Germanium and Other Semiconductors," *Bell Syst.Techn.J.*, vol. 29, pp. 560–607, 1950.

[4] T. Grasser, W. Gös, V. Sverdlov, and B. Kaczer, "The Universality of NBTI Relaxation and its Implications for Modeling and Characterization," in *Proc. Intl.Rel.Phys.Symp.*, pp. 1–13, 2007.

[5] D. K. Schroder, *Semiconductor Material and Device Characterization, 3rd Edition*. Wiley Interscience, 2006.

[6] R. A. B. Devine, J.-L. Autran, W. L. Warren, K. L. Vanheusdan, and J.-C. Rostaing, "Interfacial Hardness Enhancement in Deuterium Annealed 0.25 mu m Channel Metal Oxide Semiconductor Transistors," *Appl.Phys.Lett.*, vol. 70, no. 22, pp. 2999–3001, 1997.

[7] W. Shockley and W. T. Read, "Statistics of the Recombinations of Holes and Electrons," *Phys.Rev.*, vol. 87, no. 5, pp. 835–842, 1952.

[8] R. N. Hall, "Electron-Hole Recombination in Germanium," *Phys.Rev.*, vol. 87, no. 2, p. 387, 1952.

[9] R. Entner, A. Gehring, T. Grasser, and S. Selberherr, "A Comparison of Quantum Correction Models for the Three-Dimensional Simulation of FinFET Structures," in *Proc. IEEE International Spring Seminar on Electronics Technology*, (Sophia, Bulgaria), pp. 114–117, 2004. (best poster).

[10] G. Paasch and H. Übensee, "A Modified Local Density Approximation – Electron Density in Inversion Layers," *Phys.Stat.Sol B*, vol. 113, pp. 165–178, 1982.

[11] W. Hänsch, T. Vogelsang, R. Kircher, and M. Orlowski, "Carrier Transport Near the Si/SiO2 Interface of a MOSFET," *Solid-State Electron.*, vol. 32, no. 10, pp. 839–849, 1989.

[12] M. J. van Dort, P. H. Woerlee, and A. J. Walker, "A Simple Model for Quantisation Effects in Heavily-Doped Silicon MOSFETs at Inversion Conditions," *Solid-State Electron.*, vol. 37, no. 3, pp. 411–414, 1994.

[13] M. J. van Dort, P. H. Woerlee, A. J. Walker, C. A. H. Juffermans, and H. Lifka, "Influence of High Substrate Doping Levels on the Threshold Voltage and the Mobility of Deep-Submicrometer MOSFETs," *IEEE Trans.Electron Devices*, vol. 39, no. 4, pp. 932–938, 1992.

[14] S. Selberherr, "MOS Device Modeling at 77K," *IEEE Trans.Electron Devices*, vol. 36, no. 8, pp. 1464–1474, 1989.

[15] F. J. Grunthaner, P. J. Grunthaner, R. P. Vasquez, B. F. Lewis, J. Maserjian, and A. Madhukar, "Local Atomic and Electronic Structure of Oxide/GaAs and SiO_2/Si Interfaces Using High-Resolution XPS," *J.Vac.Sci.Technol.*, vol. 16, pp. 1443–1453, 1979.

[16] A. H. Edwards, "Interaction of H and H_2 with the Silicon Dangling Orbital at the $\langle 111 \rangle$ Si/SiO_2 Interface," *Phys.Rev.B*, vol. 44, no. 4, pp. 1832–1838, 1991.

[17] C. R. Helms and E. H. Poindexter, "The Silicon-Silicon Dioxide System: Its Microstructure and Imperfections," *Reports on Progress in Physics*, vol. 57, no. 8, pp. 791–852, 1994.

[18] J. P. Campbell, P. M. Lenahan, A. T. Krishnan, and S. Krishnan, "Observations of NBTI-Induced Atomic Scale Defects," *IEEE Trans.Dev.Mat.Rel.*, vol. 6, no. 2, pp. 117–122, 2006.

[19] J. Campbell, P. Lenahan, A. Krishnan, and S. Krishnan, "Direct Observation of the Structure of Defect Centers Involved in the Negative Bias Temperature Instability," *Appl.Phys.Lett.*, vol. 87, no. 20, pp. 1–3, 2005.

[20] P. Lenahan, T. Mishima, J. Jumper, T. Fogarty, and R. Wilkins, "Direct Experimental Evidence for Atomic Scale Structural Changes Involved in the Interface-Trap Transformation Process," *IEEE Trans.Nuclear Science*, vol. 48, no. 6, pp. 2131–2135, 2001.

[21] L.-A. Ragnarsson and P. Lundgren, "Electrical Characterization of P_b Centers in (100) Si/SiO_2 Structures: The Influence of Surface Potential on Passivation During Post Metallization Anneal," *J.Appl.Phys.*, vol. 88, no. 2, pp. 938–942, 2000.

[22] P. Lenahan and J. Conley Jr., "What Can Electron Paramagnetic Resonance Tell Us about the Si/SiO2 System?," *J.Vac.Sci.Technol.B*, vol. 16, no. 4, pp. 2134–2153, 1998.

[23] S. Fujieda, Y. Miura, M. Saitoh, Y. Teraoka, and A. Yoshigoe, "Characterization of Interface Defects Related to Negative-Bias Temperature Instability in Ultrathin Plasma-Nitrided SiON/Si$\langle 100 \rangle$ Systems," *Microelectron.Reliab.*, vol. 45, no. 1, pp. 57–64, 2005.

[24] S. Fujieda, Y. Miura, M. Saitoh, E. Hasegawa, S. Koyama, and K. Ando, "Interface Defects Responsible for Negative-Bias Temperature Instability in Plasma-Nitrided SiON/Si(100) Systems," *Appl.Phys.Lett.*, vol. 82, no. 21, pp. 3677–3679, 2003.

[25] Y. Miura and S. Fujieda, "Nitridation Effects on P_b Center Structures at SiO_2/Si(100) Interfaces," *J.Appl.Phys.*, vol. 95, no. 8, pp. 4096–4101, 2004.

[26] M. Jupina and P. Lenahan, "A Spin Dependent Recombination Study of Radiation Induced Defectsat and Near the Si/SiO_2 Interface," *IEEE Trans.Nuclear Science*, vol. 36, no. 6, pp. 1800–1807, 1989.

[27] P. M. Lenahan and M. A. Jupina, "Spin Dependent Recombination at the Silicon/Silicon Dioxide Interface," *Colloids and Surfaces*, vol. 45, pp. 191–211, 1990.

[28] D. J. Lepine, "Spin-Dependent Recombination on Silicon Surface," *Phys.Rev.B*, vol. 6, no. 2, pp. 436–441, 1972.

[29] K. L. Brower and S. M. Myers, "Chemical Kinetics of Hydrogen and (111) Si-SiO_2 Interface Defects," *Appl.Phys.Lett.*, vol. 57, no. 2, pp. 162–164, 1990.

[30] K. L. Brower, "Dissociation Kinetics of Hydrogen-Passivated (111) Si-SiO_2 Interface Defects," *Phys.Rev.B*, vol. 42, no. 6, pp. 3444–3453, 1990.

[31] J. P. Campbell and P. M. Lenahan, "Density of States of P_{b1} Si/SiO_2 Interface Trap Centers," *Appl.Phys.Lett.*, vol. 80, pp. 1945–1947, 2002.

[32] M. L. Reed and J. D. Plummer, "Chemistry of Si-SiO_2 Interface Trap Annealing," *J.Appl.Phys.*, vol. 63, pp. 5776–5793, June 1988.

[33] K. L. Brower, "Passivation of Paramagnetic Si/SiO_2 Interface States with Molecular Hydrogen," *Appl.Phys.Lett.*, vol. 53, no. 6, pp. 508–510, 1988.

[34] A. Stesmans, "Passivation of P_{b0} and P_{b1} Interface Defects in Thermal (100) Si/SiO_2 with Molecular Hydrogen," *Appl.Phys.Lett.*, vol. 68, no. 15, pp. 2076–2078, 1996.

[35] E. Cartier, J. Stathis, and D. Buchanan, "Passivation and Depassivation of Silicon Dangling Bonds at the Si(111)/SiO_2 Interface by Atomic Hydrogen," *Appl.Phys.Lett.*, vol. 63, no. 11, pp. 1510–1512, 1993.

[36] J. F. Conley, P. M. Lenahan, H. L. Evans, R. K. Lowry, and T. J. Morthorst, "Electron-Spin-Resonance Evidence for an Impurity-Related E'-Like Hole Trapping Defect in Thermally Grown SiO_2 on Si," *J.Appl.Phys.*, vol. 76, no. 12, pp. 8186–8188, 1994.

[37] J. F. Conley and P. M. Lenahan, "Room Temperature Reactions Involving Silicon Dangling Bond Centers and Molecular Hydrogen in Amorphous SiO_2 Thin Films on Silicon," *IEEE Trans.Nuclear Science*, vol. 39, no. 6, pp. 2186–2191, 1992.

[38] Z.-Y. Lu, C. J. Nicklaw, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, "Structure, Properties, and Dynamics of Oxygen Vacancies in Amorphous SiO_2," *Phys.Rev.Lett.*, vol. 89, p. 285505, Dec 2002.

[39] J. S. Brugler and P. G. A. Jespers, "Charge Pumping in MOS Devices," *IEEE Trans.Electron Devices*, vol. 16, no. 3, pp. 297–302, 1969.

[40] G. Groeseneken, H. Maes, N. Beltran, and R. F. De Keersmaecker, "Reliable Approach to Charge-Pumping Measurements in MOS Transistors," *IEEE Trans.Electron Devices*, vol. 31, no. 1, pp. 42–53, 1984.

[41] A. B. M. Elliot, "The Use of Charge Pumping Currents to Measure Surface State Densities in MOS Transistors," *Solid-State Electron.*, vol. 19, pp. 241–247, Mar. 1976.

[42] U. Cilingiroglu, "A General Model for Interface Trap Charge Pumping Effects in MOS-Devices," *Solid-State Electron.*, vol. 28, no. 11, pp. 1127–1141, 1985.

[43] E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology*. Wiley, 1982.

[44] Institut für Mikroelektronik, Technische Universität Wien, Austria, *MINIMOS-NT 2.1 User's Guide*, 2004. `www.iue.tuwien.ac.at/software/`.

[45] S. Zhu, A. Nakajima, T. Ohashi, and H. Miyake, "Interface Trap and Oxide Charge Generation under Negative Bias Temperature Instability of p-Channel Metal-Oxide-Semiconductor Field-Effect Transistors with Ultrathin Plasma-Nitrided SiON Gate Dielectrics," *J.Appl.Phys.*, vol. 98, no. 11, pp. 1–6, 2005.

[46] D. J. Fitzgerald and A. S. Grove, "Surface Recombination in Semiconductors," *IEEE Trans.Electron Devices*, vol. 15, no. 6, pp. 426–427, 1968.

[47] A. Neugroschel, C.-T. Sah, K. M. Han, and M. S. Carroll, "Direct-Current Measurements of Oxide and Interface Traps on Oxidized Silicon," *IEEE Trans.Electron Devices*, vol. 42, no. 9, pp. 1657–1662, 1995.

[48] D. M. Fleetwood, P. S. Winokur, J. R. A. Reber, T. L. Meisenheimer, J. R. Schwank, M. R. Shaneyfelt, and L. C. Riewe, "Effects of Oxide Traps, Interface Traps, and "Border Traps" on Metal-Oxide-Semiconductor Devices," *J.Appl.Phys.*, vol. 73, no. 10, pp. 5058–5074, 1993.

[49] E. Takeda, N. Suzuki, and T. Hagiwara, "Device Performance Degradation to Hot-Carrier Injection at Energies Below the Si-SiO_2 Energy Barrier," in *Proc. Intl. Electron Devices Meeting*, pp. 396–399, 1983.

[50] P. E. Cottrell, R. R. Troutman, and T. H. Ning, "Hot-Electron Emission in n-Channel IGFETs," *IEEE J.Solid-State Circuits*, vol. 14, no. 2, pp. 442–455, 1979.

[51] S. Tam, P. K. Ko, and C. Hu, "Lucky-Electron Model of Channel Hot-Electron Injection in MOSFET's," *IEEE Trans.Electron Devices*, vol. 31, no. 9, pp. 1116–1125, 1984.

[52] S. Kohyama, T. Furuyama, S. Mimurat, and H. Iizuka, "Non-Thermal Carrier Generation in MOS Structures," *Jpn.J.Appl.Phys.*, vol. 19, Suppl. 19-1, pp. 85–92, 1980.

[53] T. H. Ning, "Hot-Electron Emission from Silicon into Silicon Dioxide," *Solid-State Electron.*, vol. 21, no. 1, pp. 273–282, 1978.

[54] D. J. DiMaria, "Defect Generation Under Substrate-Hot-Electron Injection Into Ultrathin Silicon Dioxide Layers," *J.Appl.Phys.*, vol. 86, no. 4, pp. 2100–2109, 1999.

[55] J. H. Stathis and D. J. DiMaria, "Reliability Projection for Ultra-Thin Oxides at Low Voltage," *Proc. Intl. Electron Devices Meeting*, pp. 167–170, 1998.

[56] J. S. Suehle, "Ultrathin Gate Oxide Reliability: Physical Models, Statistics, and Characterization," *IEEE Trans.Electron Devices*, vol. 49, no. 6, pp. 958–971, 2002.

[57] S. Lee, B. Cho, J. Kim, and S. Choi, "Quasi-Breakdown of Ultrathin Gate Oxide Under High Field Stress," *Proc. Intl. Electron Devices Meeting*, pp. 605–608, 1994.

[58] B. E. Weir, P. J. Silverman, D. Monroe, K. S. Krisch, M. A. Alam, G. B. Alers, T. W. Sorsch, G. L. Timp, F. Baumann, C. T. Liu, Y. Ma, and D. Hwang, "Ultra-Thin Gate Dielectrics: They Break Down, but do they Fail?," in *Proc. Intl. Electron Devices Meeting*, pp. 73–76, 1997.

[59] I. C. Chen, S. Holland, K. K. Young, C. Chang, and C. Hu, "Substrate Hole Current and Oxide Breakdown," *Appl.Phys.Lett.*, vol. 49, no. 11, pp. 669–671, 1986.

[60] K. F. Schuegraf and C. Hu, "Metal-Oxide-Semiconductor Field-Effect-Transistor Substrate Current During Fowler–Nordheim Tunneling Stress and Silicon Dioxide Reliability," *J.Appl.Phys.*, vol. 76, no. 6, pp. 3695–3700, 1994.

[61] K. F. Schuegraf and C. Hu, "Effects of Temperature and Defects on Breakdown Lifetime of Thin SiO_2 at Very Low Voltages," *Proc. Intl.Rel.Phys.Symp.*, pp. 126–135, 1994.

[62] E. Avni and J. Shappir, "A Model for Silicon-Oxide Breakdown Under High Field and Current Stress," *J.Appl.Phys.*, vol. 64, no. 2, pp. 743–748, 1988.

[63] J. Suñé, I. Placencia, N. Barniol, E. Farrés, F. Martín, and X. Aymerich, "On the Breakdown Statistics of Very Thin SiO_2 Films," *Thin Solid Films*, vol. 185, pp. 347–362, 1990.

[64] D. J. Dumin, J. R. Maddux, R. S. Scott, and R. Subramoniam, "A Model Relating Wearout to Breakdown in Thin Oxides," *IEEE Trans.Electron Devices*, vol. 41, no. 9, pp. 1570–1580, 1994.

[65] P. P. Apte and K. C. Saraswat, "Modeling Ultrathin Dielectric Breakdown on Correlation of Charge Trap-Generation to Charge-to-Breakdown," *Proc. Intl.Rel.Phys.Symp.*, pp. 136–142, 1994.

[66] R. Degraeve, G. Groeseneken, R. Bellens, J. L. Ogier, M. Depas, P. J. Roussel, and H. E. Maes, "New Insights in the Relation Between Electron Trap Generation and the Statistical Properties of Oxide Breakdown," *IEEE Trans.Electron Devices*, vol. 45, no. 4, pp. 904–911, 1998.

[67] R. Tsu and L. Esaki, "Tunneling in a Finite Superlattice," *Appl.Phys.Lett.*, vol. 22, no. 11, pp. 562–564, 1973.

[68] Khairurrijal, W. Mizubayashi, S. Miyazaki, and M. Hirose, "Analytic Model of Direct Tunnel Current Through Ultrathin Gate Oxides," *J.Appl.Phys.*, vol. 87, no. 6, pp. 3000–3005, 2000.

[69] A. Gehring, *Simulation of Tunneling in Semiconductor Devices*. Dissertation, Technische Universität Wien, 2003. `http://www.iue.tuwien.ac.at/phd/gehring`.

[70] R. Moazzami and C. Hu, "Stress-Induced Current in Thin Silicon Dioxide Films," in *Proc. Intl. Electron Devices Meeting*, pp. 139–142, 1992.

[71] B. Riccó, G. Gozzi, and M. Lanzoni, "Modeling and Simulation of Stress-Induced Leakage Current in Ultrathin SiO2 Films," *IEEE Trans.Electron Devices*, vol. 45, no. 7, pp. 1554–1560, 1998.

[72] M. Herrmann and A. Schenk, "Field and High-Temperature Dependence of the Long Term Charge Loss in Erasable Programmable Read Only Memories: Measurements and Modeling," *J.Appl.Phys.*, vol. 77, no. 9, pp. 4522–4540, 1995.

[73] F. Schuler, R. Degraeve, P. Hendrickx, and D. Wellekens, "Physical Description of Anomalous Charge Loss in Floating Gate Based NVM's and Identification of its Dominant Parameter," in *Intl. Reliability Physics Symposium*, pp. 26–33, 2002.

[74] D. Ielmini, A. S. Spinelli, A. L. Lacaita, and A. Modelli, "Modeling of Anomalous SILC in Flash Memories Based on Tunneling at Multiple Defects," *Solid-State Electron.*, vol. 46, no. 11, pp. 1749–1756, 2002.

[75] D. Ielmini, A. S. Spinelli, A. L. Lacaita, and M. J. van Duuren, "Defect Generation Statistics in Thin Gate Oxides," *IEEE Trans.Electron Devices*, vol. 51, no. 8, pp. 1288–1295, 2004.

[76] L. Larcher, "Statistical Simulation of Leakage Currents in MOS and Flash Memory Devices with a New Multiphonon Trap-Assisted Tunneling Model," *IEEE Trans.Electron Devices*, vol. 50, no. 5, pp. 1246–1253, 2003.

[77] R. Entner, A. Gehring, H. Kosina, T. Grasser, and S. Selberherr, "Impact of Multi-Trap Assisted Tunneling on Gate Leakage of CMOS Memory Devices," in *Proc. NSTI Nanotech*, (Anaheim, California, USA), pp. 45–48, May 2005.

[78] R. Entner, A. Gehring, H. Kosina, T. Grasser, and S. Selberherr, "Modeling of Tunneling Currents for Highly Degraded CMOS Devices," in *Proc. Simulation of Semiconductor Processes and Devices*, (Tokyo, Japan), pp. 219–222, Sept. 2005.

[79] W. J. Chang, M. P. Houng, and Y. H. Wang, "Simulation of Stress-Induced Leakage Current in Silicon Dioxides: A Modified Trap-Assisted Tunneling Model considering Gaussian-Distributed Traps and Electron Energy Loss," *J.Appl.Phys.*, vol. 89, no. 11, pp. 6285–6293, 2001.

[80] A. Gehring, F. Jiménez-Molinos, H. Kosina, A. Palma, F. Gámiz, and S. Selberherr, "Modeling of Retention Time Degradation Due to Inelastic Trap-Assisted Tunneling in EEPROM Devices," *Microelectron.Reliab.*, vol. 43, no. 9-11, pp. 1495–1500, 2003.

[81] A. Palma, A. Godoy, J. A. Jimenez-Tejada, J. E. Carceller, and J. A. Lopez-Villanueva, "Quantum Two-Dimensional Calculation of Time Constants of Random Telegraph Signals in Metal-Oxide-Semiconductor Structures," *Phys.Rev.B*, vol. 56, no. 15, pp. 9565–9574, 1997.

[82] F. Jiménez-Molinos, A. Palma, F. Gámiz, J. Banqueri, and J. A. Lopez-Villanueva, "Physical Model for Trap-Assisted Inelastic Tunneling in Metal-Oxide-Semiconductor Structures," *J.Appl.Phys.*, vol. 90, no. 7, pp. 3396–3404, 2001.

[83] Y. Miura and Y. Matukura, "Investigation of Silicon-Silicon Dioxide Interface Using MOS Structure," *Jpn.J.Appl.Phys.*, vol. 5, p. 180, 1966.

[84] D. K. Schroder and J. A. Babcock, "Negative Bias Temperature Instability: Road to Cross in Deep Submicron Silicon Semiconductor Manufactoring," *J.Appl.Phys.*, vol. 94, no. 1, pp. 1–18, 2003.

[85] M. A. Alam and S. Mahapatra, "A Comprehensive Model of PMOS NBTI Degradation," *Microelectron.Reliab.*, vol. 45, no. 71-81, pp. 71–81, 2005.

[86] V. Huard, M. Denais, and C. Parthasarathy, "NBTI Degradation: From Physical Mechanisms to Modelling," *Microelectron.Reliab.*, vol. 46, no. 1, pp. 1–23, 2006.

[87] J. Stathis and S. Zafar, "The Negative Bias Temperature Instability in MOS Devices: A Review," *Microelectron.Reliab.*, vol. 46, no. 2-4, pp. 270–286, 2006.

[88] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, "Temperature Dependence of the Negative Bias Temperature Instability in the Framework of Dispersive Transport," *Appl.Phys.Lett.*, vol. 86, p. 143506, 2005.

[89] A. Goetzberger and H. Nigh, "Surface Charge After Annealing of Al-SiO_2-Si Structures Under Bias," *Proc.IEEE*, vol. 54, no. 10, p. 1454, 1966.

[90] E. Deal, M. Sklar, A. S. Grove, and E. H. Snow, "Characteristics of the Surface-State Charge (Qss) of Thermally Oxidized Silicon," *J.Electrochem.Soc.*, vol. 114, no. 3, p. 266, 1967.

[91] M. Ershov, S. Saxena, H. Karbasi, S. Winters, S. Minehane, J. Babcock, R. Lindley, P. Clifton, M. Redford, and A. Shibkov, "Dynamic Recovery of Negative Bias Temperature Instability in p-Type Metal-Oxide-Semiconductor Field-Effect Transistors," *Appl.Phys.Lett.*, vol. 83, no. 8, pp. 1647–1649, 2003.

[92] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, "Disorder-Controlled-Kinetics Model for Negative Bias Temperature Instability and Its Experimental Verification," *Proc. Intl.Rel.Phys.Symp.*, pp. 381–387, 2005.

[93] M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, Y. Rey-Tauriac, and N. Revil, "On-the-fly Characterization of NBTI in Ultra-Thin Gate Oxide PMOSFET's," in *Proc. Intl. Electron Devices Meeting*, pp. 109–112, 2004.

[94] M. Denais, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, N. Revil, and A. Bravaix, "New Methodologies of NBTI Characterization Eliminating Recovery Effects," in *Proc. ESSDERC*, pp. 265–268, 2004.

[95] S. Rangan, N. Mielke, and E. Yeh, "Universal Recovery Behavior of Negative Bias Temperature Instability," in *Proc. Intl. Electron Devices Meeting*, pp. 341–344, 2003.

[96] T. Yang, C. Shen, M. Li, C. Ang, C. Zhu, Y.-C. Yeo, G. Samudra, and D.-L. Kwong, "Interface Trap Passivation Effect in NBTI Measurement for p-MOSFET with SiON Gate Dielectric," *IEEE Electron Device Lett.*, vol. 26, no. 10, pp. 758–760, 2005.

[97] T. Yang, M. Li, C. Shen, C. Ang, C. Zhu, Y. Yeo, G. Samudra, S. Rustagi, M. Yu, and D. Kwong, "Fast and Slow Dynamic NBTI Components in p-MOSFET with SiON Dielectric and Their Impact on Device Life-Time and Circuit Application," in *IEEE Symposium on VLSI Technology Digest of Technical Papers*, pp. 92–93, 2005.

[98] D. Varghese, D. Saha, S. Mahapatra, K. Ahmed, F. Nouri, and M. Alam, "On the Dispersive versus Arrhenius Temperature Activation of NBTI Time Evolution in Plasma Nitrided Gate Oxides: Measurements, Theory, and Implications," in *Proc. Intl. Electron Devices Meeting*, pp. 1–4, Dec. 2005.

[99] C. Shen, M.-F. Li, C. E. Foo, T. Yang, G. S. Samudra, and Y.-C. Yeo, "Characterization and Physical Origin of Fast $V_{\rm th}$ Transient in NBTI of pMOSFETs with SiON Dielectric," in *Proc. Intl. Electron Devices Meeting*, 2006.

[100] C. Shen, M.-F. Li, X. P. Wang, Y.-C. Yeo, and D.-L. Kwong, "A Fast Measurement Technique of MOSFET $I_{\rm d}$-$V_{\rm g}$ Characteristics," *IEEE Electron Device Lett.*, vol. 27, no. 1, pp. 55–57, 2006.

[101] Y. Mitani, M. Nagamine, H. Satake, and A. Toriumi, "NBTI Mechanism in Ultra-Thin Gate Dielectric - Nitrogen-Originated Mechanism in SiON," in *Proc. Intl. Electron Devices Meeting*, pp. 509–512, 2002.

[102] L. Tsetseris, X. Zhou, D. Fleetwood, R. Schrimpf, and S. Pantelides, "Physical Mechanisms of Negative-Bias Temperature Instability," *Appl.Phys.Lett.*, vol. 86, no. 14, pp. 1–3, 2005.

[103] K. O. Jeppson and C. M. Svensson, "Negative Bias Stress of MOS Devices at High Electric Fields and Degradation of MNOS Devices," *J.Appl.Phys.*, vol. 48, no. 5, pp. 2004–2014, 1977.

[104] S. Chakravarthi, A. Krishnan, V. Reddy, C. Machala, and S. Krishnan, "A Comprehensive Framework for Predictive Modeling of Negative Bias Temperature Instability," in *Proc. Intl.Rel.Phys.Symp.*, pp. 273–282, 2004.

[105] S. Ogawa and N. Shiono, "Generalized Diffusion-Reaction Model for the Low-Field Charge Build Up Instability at the Si/SiO_2 Interface," *Phys.Rev.B*, vol. 51, no. 7, pp. 4218–4230, 1995.

[106] R. Entner, T. Grasser, H. Enichlmair, and R. Minixhofer, "Negative Bias Temperature Instability Modeling for High-Voltage Oxides at Different Stress Temperatures," in *Proc. Workshop on Dielectrics in Microelectronics*, pp. 96–97, 2006.

[107] R. Entner, T. Grasser, H. Enichlmair, and R. Minixhofer, "Investigation of NBTI Recovery During Measurement," in *San Francisco 2006 MRS Meeting Abstracts*, (San Francisco), pp. 110–111, Apr. 2006.

[108] R. Entner, T. Grasser, H. Enichlmair, and R. Minixhofer, "Influence of Interface and Oxide Traps on Negative Bias Temperature Instability," in *Abstracts IEEE 2006 Silicon Nanoelectronics Workshop*, (Honolulu), pp. 163–164, June 2006.

[109] M. Houssa, M. Aoulaiche, S. D. Gendt, G. Groeseneken, M. Heyns, and A. Stesmans, "Reaction-Dispersive Proton Transport Model for Negative Bias Temperature Instabilities," *Appl.Phys.Lett.*, vol. 86, no. 9, pp. 1–3, 2005.

[110] S. Zafar, "Statistical Mechanics Based Model for Negative Bias Temperature Instability Induced Degradation," *J.Appl.Phys.*, vol. 97, no. 10, pp. 1–9, 2005.

[111] T. Grasser, R. Entner, O. Triebl, H. Enichlmair, and R. Minixhofer, "TCAD Modeling of Negative Bias Temperature Instability," in *Proc. Simulation of Semiconductor Processes and Devices*, (Monterey, USA), pp. 330–333, Sept. 2006.

[112] R. Entner, T. Grasser, O. Triebl, H. Enichlmair, and R. Minixhofer, "Negative Bias Temperature Instability Modeling for High-Voltage Oxides at Different Stress Temperatures," *Microelectron.Reliab.*, 2007, in print.

[113] V. Arkhipov and A. Rudenko, "Drift and Diffusion in Materials with Traps," *Philos.Mag.B*, vol. 45, no. 2, pp. 189–207, 1982.

[114] T. Grasser, W. Gös, and B. Kaczer, "Dispersive Transport and Negative Bias Temperature Instability: Influence of Boundary and Initial Conditions," submitted.

[115] C. Schlünder, R. Brederlow, B. Ankele, W. Gustin, K. Goser, and R. Thewes, "Effects of Inhomogeneous Negative Bias Temperature Stress on p-channel MOSFETs of Analog and RF Circuits," *Microelectron.Reliab.*, vol. 45, no. 1, pp. 39–46, 2005.

[116] G. Kresse and J. Furthmüller, "Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set," *Phys.Rev.B*, vol. 54, no. 16, pp. 11169–11186, 1996.

[117] J. H. Stathis, "Dissociation Kinetics of Hydrogen-Passivated (100) Si/SiO_2 Interface Defects," *J.Appl.Phys.*, vol. 77, no. 12, pp. 6205–6207, 1995.

[118] S. N. Rashkeev, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, "Defect Generation by Hydrogen at the $Si\text{-}SiO_2$ Interface," *Phys.Rev.Lett.*, vol. 87, no. 16, p. 165506, 2001.

[119] L. Tsetseris and S. Pantelides, "Migration, Incorporation, and Passivation Reactions of Molecular Hydrogen at the $Si\text{-}SiO_2$ Interface," *Phys.Rev.B*, vol. 70, no. 24, pp. 1–6, 2004.

[120] J. Zhu, N. M. Johnson, and C. Herring, "Negative-Charge State of Hydrogen in Silicon," *Phys.Rev.B*, vol. 41, no. 17, pp. 12354–12357, 1990.

[121] S. Chakravarthi, A. Krishnan, V. Reddy, and S. Krishnan, "Probing Negative Bias Temperature Instability Using a Continuum Numerical Framework: Physics to Real World Operation," *Microelectron.Reliab.*, 2007, in print.

[122] D. Varghese, S. Mahapatra, and M. Alam, "Hole Energy Dependent Interface Trap Generation in MOSFET Si/SiO_2 Interface," *IEEE Electron Device Lett.*, vol. 26, no. 8, pp. 572–574, 2005.

Own Publications

[1] R. Entner, *Three-Dimensional Device Simulation with MINIMOS-NT Using the Wafer-State-Server*. Diplomarbeit, Technische Universität Wien, 2003.

[2] R. Entner, A. Gehring, T. Grasser, and S. Selberherr, "A Comparison of Quantum Correction Models for the Three-Dimensional Simulation of FinFET Structures," in *Proc. IEEE International Spring Seminar on Electronics Technology*, (Sophia, Bulgaria), pp. 114–117, 2004. (best poster).

[3] R. Kosik, T. Grasser, R. Entner, and K. Dragosits, "On the Highest Order Moment Closure Problem," in *Proc. IEEE International Spring Seminar on Electronics Technology*, (Sophia, Bulgaria), pp. 118–121, 2004.

[4] R. Entner, A. Gehring, H. Kosina, T. Grasser, and S. Selberherr, "Impact of Multi-Trap Assisted Tunneling on Gate Leakage of CMOS Memory Devices," in *Proc. NSTI Nanotech*, (Anaheim, California, USA), pp. 45–48, May 2005.

[5] S. Holzer, C. Hollauer, H. Ceric, S. Wagner, R. Entner, E. Langer, T. Grasser, and S. Selberherr, "Three-Dimensional Transient Electro-Thermal Interconnect Simulation for Stress and Electromigration Analysis," in *Proc. NSTI Nanotech*, (Anaheim, California, USA), pp. 620–623, May 2005.

[6] R. Entner, A. Gehring, H. Kosina, T. Grasser, and S. Selberherr, "Modeling of Tunneling Currents for Highly Degraded CMOS Devices," in *Proc. Simulation of Semiconductor Processes and Devices*, (Tokyo, Japan), pp. 219–222, Sept. 2005.

[7] R. Entner, T. Grasser, H. Enichlmair, and R. Minixhofer, "Investigation of NBTI Recovery During Measurement," in *San Francisco 2006 MRS Meeting Abstracts*, (San Francisco), pp. 110–111, Apr. 2006.

[8] R. Entner, T. Grasser, H. Enichlmair, and R. Minixhofer, "Influence of Interface and Oxide Traps on Negative Bias Temperature Instability," in *Abstracts IEEE 2006 Silicon Nanoelectronics Workshop*, (Honolulu), pp. 163–164, June 2006.

[9] R. Entner, T. Grasser, H. Enichlmair, and R. Minixhofer, "Negative Bias Temperature Instability Modeling for High-Voltage Oxides at Different Stress Temperatures," in *Proc. Workshop on Dielectrics in Microelectronics*, pp. 96–97, 2006.

[10] M. Karner, A. Gehring, M. Wagner, R. Entner, S. Holzer, W. Gös, M. Vasicek, T. Grasser, H. Kosina, and S. Selberherr, "VSP-A Gate Stack Analyzer," in *Proc. Workshop on Dielectrics in Microelectronics*, pp. 101–102, 2006.

[11] T. Grasser, R. Entner, O. Triebl, H. Enichlmair, and R. Minixhofer, "TCAD Modeling of Negative Bias Temperature Instability," in *Proc. Simulation of Semiconductor Processes and Devices*, (Monterey, USA), pp. 330–333, Sept. 2006.

[12] R. Entner, T. Grasser, O. Triebl, H. Enichlmair, and R. Minixhofer, "Negative Bias Temperature Instability Modeling for High-Voltage Oxides at Different Stress Temperatures," *Microelectron.Reliab.*, 2007, in print.

Curriculum Vitae

September – November 2005
Held a position as a visiting researcher in the reliability group of Guido Groeseneken at IMEC in Leuven, Belgium.

May 2004
Received "Best Poster Award for Young Scientist", at the 27th IEEE International Spring Seminar on Electronics Technology.

November 2003
Entered doctoral program at the Institute for Microelectronics, Vienna University of Technology, in cooperation with Austriamicrosystems and Infineon, under the supervision of Prof. Tibor Grasser.

October 2003
Received degree of Diplom–Ingenieur (M.Sc.) in Electrical Engineering from the Vienna University of Technology (with honors).

1994 – 2001
Summer jobs at: Landhotel Bierpeter, AVL–List Graz, EDS Graz, Austriamicrosystems Unterpremstätten.

October 1997
Enrolled in Electrical Engineering at Vienna University of Technology.

1996 – 1997
Compulsory military service.

June 1996
High school graduation (Matura) at the Bundesrealgymnasium Keplerstraße Graz (with honors).

1993 – 1995
Wifi Diplomas: PC–Administrator, System–Betreuer, Software–Entwickler, Certified Netware Administrator.

October 22nd, 1977
Born in Graz, Austria.