DISSERTATION

Comprehensive Physical Modeling of Hot-Carrier Induced Degradation

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

Ivan Starkov

Quellenstrasse 171/27
A-1100 Wien, Österreich

Matrikelnummer 0828601
geboren am 30. März 1983 in Leningrad

Wien, im Januar 2013

Kurzfassung

Transistoren fast aller Technologie-Knoten leiden unter Degradierung durch heiße Ladungsträger (HCD), die mit dem Aufbau von Defekten an oder nahe der Silizium/Siliziumdioxid-Grenzfläche verbunden ist. Dieses nachteilige Phänomen ist seit mehr als vier Jahrzehnten bekannt, seither sind zahlreiche Modellierungsversuche unternommen worden. Die Degradierung durch heiße Ladungsträger zu modellieren ist kompliziert, da diese drei verschiedene, aber stark verbundene Aspekte umfasst. Tatsächlich brechen die Ladungsträger, die mit der Si/SiO$_2$ Grenzfläche interagieren die Si $-$ H Bindungen, wodurch Grenzflächendefekte generiert werden. Dementsprechend müssen die mikroskopischen Mechanismen für die Generierung von Defekten richtig beschrieben werden. Die Information darüber, wie effizient diese Ladungsträger den Bindungs-Dissoziations Prozess auslösen, erfolgt durch eine gründliche Ladungsträger-Transport-Behandlung. Weiters können jene generierten Defekte Ladungsträger einfangen, wodurch die Elektrostatik des Transistors verändert wird. Darüber hinaus wirken sie als zusätzliche Streuzentren und setzen die Ladungsträgerbeweglichkeit herab. Daher ist ein weiterer wichtiger Teilvorgang in dem heißen Ladungsträger-Degradierungs-Modellierungsparadigma die Simulation der Eigenschaften von degradierten Bauelementen. Aufgrund der komplizierten Natur fehlt noch immer ein umfassendes Physikalisches Modell, daher sind die meisten bestehenden Modelle empirisch.

Die Hauptaufgabe dieser Arbeit ist es, ein Physikalisches Modell für die Degradierung durch heiße Ladungsträger vorzuschlagen, welches in der Lage ist, diese Degradierung in Metall-Oxid-Halbleiter Feldeffekttransistoren (MOSFET) mit unterschiedlichen Kanallängen unter Verwendung eines einzigen Satzes von physikalischen Parametern zu beschreiben. Der entwickelte Ansatz berücksichtigt nicht nur die Schäden, die durch Kanal-Elektronen erzeugt werden, sondern auch die Schäden durch Stoßionisation sekundär produzierter Kanal-Löcher. Obwohl der Beitrag der Löcher zu der gesamten Defektgenerierung im Vergleich zu den Elektronen kleiner ist, ist ihr Einfluss auf den linearen Transistor-Strom vergleichbar mit dem der Elektronen. Der Grund für diesen Trend sind durch Löcher erzeugte Defekte nahe dem Drain-Anschluss, wodurch das elektrische Verhalten des Bauelementes stärker beeinflusst wird.

Das Modell umfasst drei Hauptkomponenten: ein Ladungsträger-Transport-Modul, ein Modul zur Modellierung der mikroskopischen Mechanismen der Defekt-Generierung und ein Modul für die Simulation der Eigenschaften degradierter Bauelemente. Das Ladungsträger-Transport-Modul berechnet eine Reihe von Ladungsträgerenergie Verteilungsfunktionen (DF) an einer beliebigen Position in dem MOSFET für eine bestimmte Bauelemente Architektur und für bestimmte Stress/Betriebsbedingungen. Für die Berechnung der DF wird der Vollband-Monte Carlo (MC) Bauelemente Simulator MONJU eingesetzt. Anschließend wird die Information über die Ladungsträger DF verwendet, um für jeden Stress Zeitpunkt Grenzflächenzustandsdichte-Lateral-Profile zu erzeugen. Diese Profile werden dann in den Schaltungs-und Bauelementsimulator MiniMOS-NT geladen, welcher die elektrischen Charakteristiken des degradierten Bauelementes berechnet. Da die MC Methode sehr zeitaufwendig ist, wird der Ansatz verfolgt, das genaue Ladungsträgertransport-Modul durch eine vereinfachte Behandlung zu ersetzen, welches bedeutet, dass ein Vergleich zwischen den verschiedenen Implementierungen des Ladungsträgertransport Moduls des Modells durchgeführt wird. Hierfür werden Monte-Carlo, Hydrodynamische (HD) und Drift-Diffusions (DD) Ansätze für die Lösung der Boltzmann Transportgleichung angewandt. Eine Diskrepanz zwischen experimentellen Ergebnissen und Simulationen, die unter Einsatz dieses vereinfachten Ansatzes eingetreten ist, wird aufgezeigt und erklärt.

Wie bereits erwähnt, umfasst das entwickelte Physikalische Modell für HCD drei Hauptmodule. Jedes Modul basiert auf bestimmten Annahmen, die als potenzielle Fehlerquellen fungieren, und beinhaltet eine bestimmte Anzahl von anzupassenden Parametern. Mit anderen Worten, aufgrund der komplizierten Struktur des Modells, sollten mögliche Fehler durch geeignete Auswertung der Schnittstellen zwischen den Modulen überprüft werden. Obwohl eine gute Darstellung der degradierten Bauelementeeigenschaften erreicht werden könnte, wird im folgenden versucht, das mikroskopische Modell des Defektaufbaus zu überprüfen, indem man die Technik des Ladungspumpen (CP) für eine bestimmte Bauelemente Architektur anwendet und mit dem Modell vergleicht. Es ist zu betonen, dass die Algorithmen für die Extraktion von den Grenzflächenzustandsprofilen, in Abhängigkeit von der Position entlang der Bauelemente Grenzfläche aus den CP Daten, auf bestimmten Annahmen basieren und somit auch potenziell ungenau sein können. Darüber hinaus werden eine umfassende Analyse und ein Vergleich der einzelnen Extraktionstechniken, mit deren Hilfe die räumliche Verteilung der durch heiße Ladungsträger induzierten Grenzflächen- und Oxiddefekte bestimmt wird, durchgeführt. Die Vorteile und Grenzen der Charakterisierungsalgorithmen, verbessert durch ein neues einfaches, kompaktes Modell für die lokale Oxidkapazität, werden diskutiert. Es wird gezeigt, dass durch Ignorieren der räumlichen Veränderung der Oxidkapazität ein falsches Ergebnis erzeugt wird, welches zu einer mehrdeutigen Interpretation von HCD führt. Daher wird eine behutsame Gewinnung des anfänglichen Grenzflächenzustandsdichte-Profils für einen vorgestressten MOSFET vorgenommen. Die Auswirkung der bereits vorhandenen Bauelemente Grenzflächenzustände auf die nachfolgenden Grenzflächen-Defektprofile wird im Rahmen der prädiktiven HCD Modellierung aufgezeigt. Ein gründlicher Vergleich zwischen simulierten Grenzflächenzustandsprofilen und solchen, die aus CP Daten extrahiert wurden, wird durchgeführt.

Basierend auf einer rigorosen TCAD Version eines Physikalische Modells für HCD wird ein analytisches Modell, das sich in geeigneter Weise dem Träger-Beschleunigungs-Integral (AI) nähert, entwickelt. Unter Verwendung eines solchen Ansatzes ist man in der Lage, die Degradierung des linearen Transistor-Stroms zu repräsentieren. Einer der wichtigsten Vorteile dieses analytischen Ansatzes ist, dass dieser auf einem Physikalische TCAD-Modell, anstatt auf einer empirischen Anpassung zu experimentellen Daten, beruht. Das Modell stellt auch die bei relativ langen Stresszeiten beobachtete Sättigung der HCD dar. Die Flexibilität des resultierenden Ausdrucks ermöglicht es, diesen Ansatz zu verwenden, während man die Auswirkungen von schwankenden Parametern der Bauelemente Topologie auf HCD berücksichtigt. In diesem Fall führt das zeitaufwendige MC basierte Transport-Modul zu extrem hohem Rechenaufwand.

Abstract

Transistors of nearly all technology nodes suffer from hot-carrier degradation (HCD) which is associated with the buildup of defects at or near the silicon/silicon dioxide interface. This detrimental phenomenon has been known for more than four decades and numerous modeling attempts have been undertaken. However, hot-carrier degradation is rather complicated to model because it includes three different but strongly connected aspects. In fact, carriers interacting with the Si/SiO_2 interface break $Si-H$ bonds, thereby generating traps and thus the microscopic mechanisms for defect creation must be properly described. The information on how efficiently these carriers trigger the bond dissociation process is provided by a thorough carrier transport treatment. Furthermore, these generated traps can capture carriers and thus distort the electrostatics of the transistor and degrade the carrier mobility. Due to the complicated nature, a comprehensive physics-based model is still missing and most existing HCD models are empirical.

The main task of this work is to design a physics-based model for hot-carrier degradation, which is able to represent HCD observed in metal-oxide-semiconductor field-effect-transistors (MOSFETs) with different channel length using a single set of physical parameters. The developed approach considers not only the damage produced by channel electrons but also by secondary channel holes generated by impact ionization. Although the contribution of the holes to the total defect creation is smaller compared to that of electrons, their impact on the linear drain current is comparable with the electron one. The reason behind this trend is that hole-induced traps are shifted towards the source, thereby more severely affecting the device behavior.

The model includes three main modules: a carrier transport module, a module for modeling of microscopic mechanisms of defect creation and a module for the simulation of the characteristics of degraded devices. The carrier transport module calculates a set of carrier energy distribution functions (DFs) at any position in the MOSFET for a particular device architecture and stress/operating conditions. For DF calculation a full-band Monte Carlo (MC) device simulator MONJU is employed. Then the information regarding the carrier DF is used to generate interface state density profiles. These profiles are loaded into a circuit and device simulator MiniMOS-NT, which calculates the characteristics of the degraded device. Since the MC method is very time-consuming the precise carrier transport module is substituted by a simplified treatment, i.e. a comparison between different realizations of the model carrier transport module is carried out. Namely Monte Carlo, hydrodynamic (HD) and drift-diffusion (DD) schemes for the solution of the Boltzmann transport equation are employed. A discrepancy between experimental results and simulations, which occurred while employing this simplified approach, is shown and explained.

As previously stated, the developed physics-based model for HCD includes three main modules. Each module is based on some assumptions acting as potential sources of error and includes a certain number of fitting parameters. In other words, due to the complicated structure of the model, these errors should be screened by properly evaluating the interfaces between the modules. Therefore, although a good representation of the degraded device characteristics could be achieved, one attempts to verify the microscopic model for the defect build-up in the following by employing the charge pumping (CP) technique for a particular device architecture. It is worth emphasizing that the algorithms for the extraction of the interface state profiles versus position along the device interface from the CP data are based on some assumptions and thus could be potentially inaccurate as well. Therefore, an exhaustive analysis and comparison of extraction techniques of the hot-carrier induced interface and oxide trap spatial distributions is performed. The advantages and limitations of the characterization algorithms improved by a new simple compact model for local oxide capacitance are discussed. It is demonstrated that by ignoring the spatial variation of oxide capacitance, a spurious result is produced, leading to an ambiguous picture of HCD. Additionally, a careful extraction of the initial interface state density profile for a pre-stressed MOSFET is undertaken. The impact of pre-existing interface states on the subsequent interface trap profile evolution is demonstrated in the context of predictive HCD modeling. A thorough comparison between simulated interface state profiles and those extracted from CP data is carried out.

Based on a rigorous technology computer-aided design (TCAD) version of a physics-based model for HCD, an analytical model, which suitably approximates the carrier acceleration integrals (AIs), is developed. Using such an approach one is able to represent the linear drain current degradation. One of the main advantages of this analytical approach is that it is based on a physics-based TCAD model rather than on an empirical fit to experimental data. The model also represents the saturation of HCD observed at relatively long stress times. The flexibility of the resulting expression allows us to employ this approach while considering the impact of fluctuating parameters of device topology on HCD. In this case the time-consuming MC based transport module would otherwise lead to extremely high computational costs.

Education is a weapon whose effects depend on
who holds it in his hands and at whom it is aimed.

Iosif Dzhugashvili, Interview with H. G. Wells (September 1937)

Contents

List of Figures

List of Tables

List of Abbreviations

AC	alternating current
AI	acceleration integral
BTE	Boltzmann's transport equation
BTI	bias temperature instability
CMOS	complementary metal-oxide- semiconductor
CP	charge pumping
DC	direct current
DD	drift-diffusion
DEMOS	drain extended metal-oxide-semiconductor
DF	distribution function
DOS	density-of-states
HC	hot carrier
HCD	hot-carrier degradation
HD	hydrodynamic
HV	high voltage
LDMOS	lateral diffused metal-oxide-semiconductor
MC	Monte Carlo
MOS	metal-oxide-semiconductor
MOSFET	metal-oxide-semiconductor field-effect-transistor
MP	multiple particle
MR	multiple refresh
NBTI	negative bias temperature instability
NSC	non-self-consistent
SC	self-consistent
SHE	spherical harmonics expansion
SP	single particle
STI	shallow trench isolation
STM	scanning tunneling microscope
TCAD	technology computer-aided design
WCC	worst-case conditions

List of Symbols

Symbol	**Unit**	**Description**
C_{ox}	$\mathrm{F\ m^{-2}}$	Oxide capacitance
E	eV	Carrier energy
E_{emi}	eV	Emission process barrier high
E_{pass}	eV	Passivation process barrier high
$E_{\mathrm{th,SP(MP)}}$	eV	Threshold energy of SP(MP)-process
f	Hz	Frequency of gate pulse
$f(E)$	$\mathrm{eV^{-1}}$	Distribution function
F	$\mathrm{V\ m^{-1}}$	Electric field
$g(E)$	$\mathrm{m^{-3}eV^{-1}}$	Density-of-states
g_{m}	$\mathrm{A\ V^{-1}}$	Transconductance
I_{dlin}	A	Linear drain current
I_{dlin0}	A	Initial linear drain current
ΔI_{dlin}	A	Linear drain current shift
I_{ds}	A	Source-Drain current
I_{sub}	A	Substrate current
I_{d}	A	Drain current
I_{s}	A	Source current
I_{g}	A	Gate current
I_{cp}	A	Charge pumping current
$I_{\mathrm{cp,it}}$	A	Charge pumping current induced by border traps
$I_{\mathrm{cp,it}}$	A	Charge pumping current induced by interface states
$I_{\mathrm{cp,max}}$	A	Maximum value of charge pumping current
$I_{\mathrm{SP(MP)}}$	$\mathrm{eV^{-1}s^{-1}}$	Carrier acceleration integral for SP(MP)-process
L_{gate}	m	Gate mask length
L_{eff}	m	Effective channel length contributing to charge pumping current
L_{c}	m	Position of device center
L_{ch}	m	Channel length
ΔL	m	Difference between nominal and effective channel lengths
μ	$\mathrm{m^2\ V^{-1}\ s^{-1}}$	Mobility
n	$\mathrm{m^{-3}}$	Electron density
n_0	$\mathrm{m^{-2}}$	Concentration of passivated $\mathrm{Si-H}$ bonds
$\nu_{\mathrm{SP(MP)}}$	eV	Attempt rate of SP(MP)-process
$\nu_{\mathrm{emi(pass)}}$	$\mathrm{s^{-1}}$	Attempt frequency for emission(passivation) process
N_{d}	$\mathrm{m^{-3}}$	Donor concentration
N_{dd}	$\mathrm{m^{-3}}$	Differential doping density
N_{bt}	$\mathrm{m^{-2}}$	Number of border traps per area

N_{it}	m^{-2}	Number of interface states per area
$N_{\mathrm{it},0}$	m^{-2}	Number of interface traps per area of pre-stressed device
ΔN_{it}	m^{-2}	Generated interface state density
N_{ot}	m^{-2}	Number of oxide traps per area
ΔN_{ot}	m^{-2}	Generated oxide trap density
$N_{\mathrm{SP(MP)}}$	m^{-2}	SP(MP)-related interface state density
p	m^{-3}	Hole density
$P_{\mathrm{emi(pass)}}$	$\mathrm{eV}^{-1}\mathrm{s}^{-1}$	Emission(passivation) process rate
$P_{\mathrm{u(d)}}$	$\mathrm{eV}^{-1}\mathrm{s}^{-1}$	Phonon excitation(decay) rate
Q_{ot}	$\mathrm{C\ m}^{-2}$	Oxide charge per area
Q_{it}	$\mathrm{C\ m}^{-2}$	Interface charge per area
Q_{bt}	$\mathrm{C\ m}^{-2}$	Border charge per area
$\sigma_{\mathrm{surface}}$	$\mathrm{C\ m}^{-2}$	Surface charge density
$\sigma_{\mathrm{SP(MP)}}$	m^{2}	SP(MP)-process related capture cross section
τ_E	s	Energy relaxation time
$\tau_{e(h)}$	s	Time constant for electron(hole) trapping
t	s	Stress time
t_{ox}	m	Oxide thickness
t_{n}	m	Nominal oxide thickness
T_{n}	K	Electron temperature
T_{L}	K	Lattice temperature
U	V	Potential
v	$\mathrm{m\ s}^{-1}$	Carrier velocity
V_{A}	V	Amplitude of applied gate pulse string
V_{g}	V	Gate voltage
V_{gs}	V	Gate-source voltage
V_{ds}	V	Drain-source voltage
V_{gl}	V	Low level of applied gate pulse string
V_{gh}	V	High level of applied gate pulse string
V_{fb}	V	Flatband voltage for fresh device
$V_{\mathrm{fb,s}}$	V	Flatband voltage for stressed device
V_{th}	V	Threshold voltage for fresh device
$V_{\mathrm{th,s}}$	V	Threshold voltage for stressed device
ΔV_{th}	V	Threshold voltage shift
V_{sub}	V	Substrate voltage
x	m	Coordinate along device interface
x_{max}	m	Maximum tunneling distance
w_{e}	s^{-1}	Phonon frequency
W	m	Gate width

Physical Constant

ε_0	$=$	$8.854187818 \cdot 10^{12}\mathrm{F\ m}^{-1}$	Vacuum permittivity
$\varepsilon_{\mathrm{ox}}$	$=$	$3.9 \cdot \varepsilon_0$	Silicon oxide permittivity
h	$=$	$1.054571726 \cdot 10^{34}\mathrm{J\ s}$	Planck constant
$\hbar$	$=$	$h/2\pi$	Reduced Planck constant
$\mathrm{k_B}$	$=$	$1.380648813 \cdot 10^{23}\ \mathrm{J\ K}^{-1}$	Boltzmann constant
q	$=$	$1.602176565 \cdot 10^{19}\ \mathrm{C}$	Elementary charge

Chemical Symbols

D	Deuterium
H^*	Mobile hydrogen
$\mathrm{Si-D}$	Silicon deuterium
$\mathrm{Si-H}$	Silicon hydrogen
$\mathrm{SiO_2}$	Silicon dioxide
SiON	Silicon oxynitride

1 Introduction

The aggressive scaling of the fundamental building block of complementary metal-oxide- semiconductor (CMOS) technology, the MOSFET, has not only resulted in the shrinking of device dimensions, but also in the reduction of the gate oxide thickness and the power supply voltage [1]. However, in the 80s the power supply could not scale down at the same rate as the channel length, leading to an increased electric field in the channel. As a result, one of the most important concerns in the field of transistor reliability, the hot carrier (HC) induced degradation has been of high significance [2–6]. Being accelerated by the electric field, the carriers can gain a sufficiently high energy and thus become "hot". These hot carriers impinge on the silicon/silicon dioxide interface and can cause damage via breaking $\mathrm{Si}-\mathrm{H}$ bonds, thereby generating interface states able to capture carriers and thus become charged. These additional charges introduced into the system are distributed along the channel and perturb the electrostatics of the device resulting, for instance, in a shift of the threshold voltage. Furthermore, they act as additional scattering centers, thereby degrading the mobility and, as a result, the transconductance and linear drain current. This phenomenon is called hot-carrier degradation. It has been known for more than four decades and numerous modeling attempts have been undertaken. However, the exact location of the defects, as well as their temporal buildup during stress, is rarely studied. Moreover, although a considerable number of HCD models have been proposed, most of them are empirical/phenomenological and only a limited number of physics-based models exists.

1.1 Historical Background

The first successful attempt to model HCD was the so-called "lucky-electron" model proposed in 1979 by Hu [5,7]. This concept is based on the following assumptions: an electron characterized by an energy high enough to overcome the potential barrier at the interface impinges onto the interface without collision. That is, without energy loss and without being scattered back into the channel and being emitted into the SiO_2 conduction band thereby producing a defect. The "lucky electron" model claims that the threshold of HCD is 3.7eV, however, hot-carrier stresses performed at $V_{\mathrm{ds}} < 3\mathrm{V}$ demonstrated that device aging can also occur at lower voltages [8]. As a consequence, this approach fails for short-channel devices with lower operating voltage. However, due to its simplicity, the model still remains one of the most popular approaches.

An empirical extension of the "lucky electron" model was proposed in 1983 by Takeda and Suzuki [4, 9]. This simple time dependent model expresses the transconductance degradation and/or threshold voltage shift ΔV_{th} by a time power law t^n. The exponent and proportionality coefficients are fitting parameters adjusted independently for a particular device architecture. The advantage of such an approach is that it allows for an easy extrapolation of the device life-time from accelerated hot-carrier stress conditions to real operation biases. However, the applicability of the model is rather limited as demonstrated by investigations employing lightly doped drain structures where the saturation of degradation after a certain value has been observed (see [10] and references therein). Although inaccurate for describing hot-carrier degradation, the Takeda model inspired a number of fitting models. These models try to represent device parameter degradation employing some combinations of time exponents. Among them are the Goo model based on the "lucky electron" concept, which can capture saturation of degradation [10], the Dreesen model [11, 12], which follows the same strategy but was adapted for lightly doped drain MOSFETs and is able to successfully represent the I_{dlin} degradation in the range of ΔI_{dlin}=0.02%...10%.

Other extensions of the Hu concept have been proposed by Woltjer [13, 14] and by Mistry *et al.* only 10 years later. [15, 16]. In contrast to the "lucky electron" model, which deals with interface trap generation under maximum substrate current conditions but fails at other stress conditions, the Woltjer model considers the oxide field as significant for the creation of interface states. As a result, a field-driven correction is incorporated into the "lucky electron" model. This extension gives description of the degradation behavior of devices with various dimensions and oxide thicknesses. Mistry and co-workers reported that a single degradation mechanism is not sufficient for proper degradation modeling and three different modes of damage were proposed: at low V_{gs} creation of interface states and oxide neutral electron traps occurs, while for mid and high V_{gs} only interface state build-up and oxide electron traps contribute. All of them are present during direct current (DC)-stress and each of them can dominate the alternating current (AC)-stress life-time [16]. However, the life-times predicted by this model were rather inaccurate and thus only of limited applicability. Moreover, the general shortcoming of these approaches is that starting from a certain node and beyond, the field-driven paradigm and related modeling approaches, such as extensions of the "lucky electron" model, should be substituted by energy-driven concepts [8, 17, 18].

The idea that two (or several) competing degradation mechanism are required to describe the overall degradation has been further extended in 2003 by Moens *et al.* in order to capture degradation in lateral diffused metal-oxide-semiconductor (LDMOS) transistors [19–22]. In a series of papers Moens demonstrated that for high-voltage devices one should consider defect build-up in different transistor sections, namely in the channel, accumulation, and bird's beak regions. As a result, different components of the damage are characterized by different time exponents, which explains the different slopes of parameter degradation. The dynamic behavior was reported to be determined by a hole trapping/detrapping processes [19, 21].

1.2 Main Physics-Based Concepts for HCD Modeling

All models described above have been developed for the description of HCD observed in a particular class of devices. As such, they are empirical or phenomenological, at the best. However, a proper description of HCD may only be possible when the physical picture is accurately understood and captured by the model. There are five main physics-based concepts for hot-carrier degradation modeling elaborated so far:

- the approach presented by Hess and co-authors [23, 24];
- the empirical extension of the Hess model to make it suitable for TCAD device simulators in the work of Penzin *et al.* [25];
- the extension of the reaction-diffusion framework proposed by Kufluoglu and Alam [26,27];
- the energy-driven paradigm by Rauch and LaRosa [28, 29];
- the Bravaix model based on the Hess approach [30, 31].

The most important breakthrough in HCD modeling is due to Hess who introduced the interplay between a single- and a multiple-carrier mechanism for $\mathrm{Si-H}$ bond-breakage. Since these mechanisms are related to the fractions of "hot" and "colder" carriers, the idea that the matter is controlled by the carrier energy DF was first acknowledged [32]. Notwithstanding the fact that the model is able to explain such a key feature of HCD such as the hydrogen/deuterium isotope effect [33], the link between the device microscopic picture of the defect build-up and degradation of device characteristics is missing. An attempt to link these levels has been undertaken in the Penzin model [25]. In fact, this work (as the successor of the Hess work) presents a phenomenological treatment for HCD modeling. Another approach is the extension of the reaction-diffusion framework of the negative bias temperature instability (NBTI) in order to capture HCD [26,27]. This implies, however, that once stress is removed, full recovery should be observable within reasonable times. In reality, the recovery of HCD is very slow, thus suggesting that HCD is a reaction-limited process [34]. One more strategy for HCD modeling proposed by Rauch and LaRosa is called "energy-driven paradigm" [28, 29]. For channel lengths less than 180nm, HCD was shown to be controlled by the single "knee" energy. This energy is related to the stress bias. Therefore, instead of operating with coordinate-dependent quantities (electric field, dynamic temperature, DF, etc) only a single bias-dependent parameter is considered. A combination of the Hess and Rauch approaches was proposed by Bravaix *et al.* [30, 31]. In this concept the interaction between the single- and multiple-carrier mechanisms for $\mathrm{Si-H}$ bond-breakage has been considered. However, the crucial point is that the information about the carrier DF is substituted by some empirical factors. In spite of a certain success of all these approaches the main problem is that they capture only a fragment of the whole HCD mosaic. Therefore, the hierarchical ladder connecting the microscopic level of defect creation and the device simulation level is still not fully understood.

To summarize, over the last decades hot-carrier degradation modeling has evolved from simple empirical models to a more detailed understanding of the microscopic physics involving single particle (SP) and multiple particle (MP) processes. A detailed description of the physics requires knowledge of the carrier energy distribution function which can only be obtained from a solution of the Boltzmann transport equation (BTE). Most models in use today employ simplified solutions based on the average energy or, even more dramatic, the electric field, while in the ultimate simplification attempts are made to capture the physics using closed analytic expressions. Although computationally more efficient, these approaches are inevitably inaccurate, even though their limitations might not be that obvious when a limited range of bias conditions, temperatures, and channel-lengths is investigated. Therefore, after describing the main features of hot-carrier degradation, one proceeds to the detailed analysis of the existing physics-based HCD models finishing with the presentation and validation of a detailed model.

1.3 Characteristic Features of HCD

Although the detrimental phenomenon of hot-carrier degradation has been known for more than four decades, it remains one of the most significant concerns in transistor reliability. Since during this period of time several generations of MOSFETs have been in production, the characteristic features of HCD, their understanding, and the modeling approaches also reflect these trends. For instance, in the 80s, the device dimensions were reduced rather quickly, accompanied by a slower scaling of the transistor power supply. This tendency led to high electric fields in the metal-oxide-semiconductor (MOS) transistor channel, which accelerated carriers up to energies high enough to directly trigger a $\mathrm{Si}-\mathrm{H}$ bond-breakage process by a solitary carrier, which was then considered "hot" [2,4,5]. Such a situation required specific measures in order to suppress carrier heating. Among them was the demand that the supply voltage should scale faster than device dimensions [35–38] in addition to requirements for doping profiles and device geometry, which for instance resulted in lightly doped drain structures [30,39].

In particular, even though in the 0.25μm node hot-carrier degradation could be rather dramatic, its importance was expected to reduce drastically for coming nodes [2]. The physical reason behind this expectation was that the source-drain voltage V_{ds} had already been scaled down to $1-1.5$V while the threshold energy required for triggering the $\mathrm{Si}-\mathrm{H}$ bond dissociation process is about $\sim 3.0-3.5$eV. Therefore, it was expected that the carrier would not be heated up to energies sufficient enough for the $\mathrm{Si}-\mathrm{H}$ bond-breakage, resulting in a suppression of HCD. Overall, a complete absence of HCD was expected for extremely-scaled devices [24,30,40,41].

1.3.1 Different Nature of HCD Mechanisms

In reality, however, even ultra-scaled modern MOSFETs can show severe HCD [30,31,41]. This was first demonstrated for gate lengths less than than 0.2μm and supply voltages below 1.0V by Mizuno *et al.* [42]. The authors related this finding to an energy exchange mechanism populating the "hot" fraction of carrier ensemble. Here "hot" means that these carriers have energies substantially higher than the lattice temperature. Possible mechanisms responsible for such an energy gain include impact ionization [43], Auger recombination [44], electron-phonon [45], and electron-electron scattering [17,18,46].

Note that electron-electron scattering is of particular importance for nano-scale devices [30,41]. Particularly for these devices the situation is even more complicated because the dominant mechanism for $\mathrm{Si}-\mathrm{H}$ bond-breakage changes from a single-carrier to a multiple-carrier mechanism [24,30,40,41]. For example, in a long-channel or high-voltage device carriers striking on the interface are already rather hot and are able to trigger silicon-hydrogen bond rupture by a single collision, which is referred to as the single-particle mechanism. In contrast, such extremely hot carriers do not exist in a sufficient quantity for scaled devices. Rather, several particles subsequently bombard a bond, thereby exciting and eventually rupturing it, which is referred to as the multiple-particle process. However, these two scenarios are just limiting cases and in a particular device geometry under certain operating/stress conditions a superposition of these two mechanisms has to be expected.

The most important consequence of the interplay between SP- and MP-carrier processes is the change of the worst-case condition of hot-carrier degradation: traditionally, the worst-case of HCD occurred at $V_{\mathrm{gs}} \cong (0.4-0.5)V_{\mathrm{ds}}$, corresponding to the maximum substrate current or – in other words – to the largest impact ionization rate [8, 47–49]. However, this is not always the case even for long-channel devices; for example, in high-voltage p-MOSFETs the worst-case conditions are observed at the maximum gate current and no empirical law exists for this case [50–52]. This regime corresponds to the situation where the average carrier energy is maximum, that is, the carrier ensemble includes a substantial fraction of particles with energies high enough to induce the bond dissociation following a single impact to launch the SP-mechanism.

1.3.2 HCD in High-Voltage Devices

While the history of HCD in scaled MOSFETs has been rather dramatic, the situation around high voltage (HV) devices is equally challenging, i.e. this concern was and is one of the most severe problems in devices of this class. At present, the LDMOS and drain extended metal-oxide-semiconductor (DEMOS) are the most promising high-voltage devices commonly employed in display drivers, automobile and smart-power applications, because of their compatibility with the standard CMOS process [19, 22, 47, 51, 53–63]. The post-stress behavior of both CMOS and high-voltage devices displays many similarities:

- the HC damage is associated with the breakage of $\mathrm{Si-H}$ bonds in the vicinity of the interface resulting in interface state generation and bulk oxide charge build-up;
- the strong localization of degradation [56–58, 63–67];
- temperature effect, i.e. that HCD becomes less pronounced at higher temperatures [68–73]. In extremely-scaled devices HCD becomes more significant at elevated temperatures due to the dominant role of electron-electron scattering [41, 74–76].

The behavior of high-voltage devices under HC stress features many additional peculiarities. One of the most severe issues is the Kirk effect [58–62, 77, 78] resulting in a secondary substrate current I_{sub} peak at high gate voltages V_{gs} (above ~30V) and, hence, in a strong degradation of the ON-state resistance [58–61]. The second one is the considerable HCD recovery [21,60,79–81] vs. a weak one observed in CMOS devices [82–84]). Due to this recovery the AC device lifetime is much higher than that in the DC regime. The degradation suppression under accelerated temperatures is still an ambiguous issue for high-voltage devices. In fact, the parasitic effect of self-heating (often strongly localized) drastically complicates the high-voltage device behavior under HC stress [19, 73, 79]. Additionally, the HCD picture for p-channel devices (no matter whether low- or high-voltage transistors are meant) is contaminated by an interplay with NBTI which shows the opposite temperature behavior [76, 85–87].

1.3.3 HCD in Scaled Devices

In contrast to HV devices, the operating voltages of scaled devices are such that a single carrier is unlikely to reach energies sufficiently large to trigger an SP-process. The process of energy exchange between carriers is of a stochastic nature and therefore one may expect that a certain fraction of particles – however small – may still obtain a relatively large energy. Regardless, although particles which are able to launch the SP-mechanism are in principle present, their relative number is rather small and, hence, the MP-process becomes dominant [40,88]. Contrary to the SP-mechanism, the individual carriers contributing to the MP-mechanism require only a relatively low energy. However, a large number of those carriers is needed. Thus, the carrier flux rather than the single-carrier energy becomes important in this case. The maximum carrier flux is obtained at $V_{\mathrm{ds}} = V_{\mathrm{gs}}$ for both scaled n- and p-MOSFETs [89–92], which now becomes the region of maximum HCD.

It is worth mentioning that even in the case of ultra-short devices a certain fraction of "hot" carriers exists because the high-energy tail of the carrier distribution function is populated for instance by the electron-electron scattering process [29, 41]. Therefore, the SP-mechanism will still contribute in these devices. Also, thermalized, that is, "cold", particles still exist even in the case of high-voltage devices, thereby also leading to HCD by the MP-process. To conclude, in a real device under real operating/stress conditions, the interplay between the SP- and MP-modes of bond-breakage must considered and is controlled by the way carriers are distributed over energy, that is, by the carrier DF.

1.3.4 Strong Localization

Another characteristic feature of HCD is its strong localization near the pinch-off region (or the drain end of the gate), near the area where the electric field peaks [2, 41, 56–58, 63, 65]. Such a peculiarity is again related to carriers heating up to energies required to launch the bond-breakage process. Since the driving force of this acceleration is the electric field, for the sake of simplicity it is often assumed that the maximum of the interface state generation rate corresponds to the electric field peak. However, it has been long understood that the DF can follow changes in the electric field only with a certain delay [93]. Therefore, in order to improve over the electric field approximation, such quantities as the carrier temperature have been used to estimate the location of the maximum damage. However, as it was demonstrated in [94, 95], the maxima of different quantities are observed at different positions and therefore the N_{it} peak never directly coincides with that of the electric field. Moreover, Zaka *et al.* showed that different simplified treatments of carrier transport employing the drift-diffusion, energy-transport, and spherical harmonics expansion methods (keeping only the 0^{th} and 1^{st} order polynomials) lead to spurious description of hot-carrier injection [96]. As a result, the spherical harmonics expansion method for BTE solution with a higher expansion number of the stochastic Monte Carlo-based solver must be used. This finding is very important because the $\mathrm{Si}-\mathrm{H}$ bond-breakage process is described by an energy-dependent reaction cross section [30, 98, 99]. Hence, it is important to know the magnitude of the carrier fraction which corresponds to the given energy.

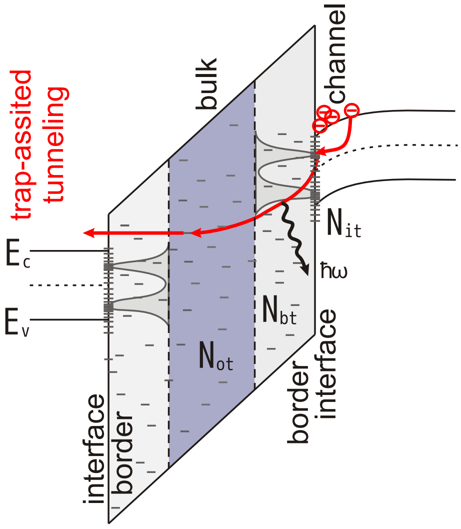


Figure 1.1: Interface, border and oxide bulk traps. Trap-assisted tunneling is represented by red arrows.

1.3.5 Interface, Border and Oxide Traps

One of the essential contributions to HCD is the build-up of interface states with the corresponding concentration N_{it} dependent on the coordinate along the interface. These traps when charged perturb the electrostatics of the transistor (resulting e.g. in the threshold voltage shift ΔV_{th}) and act as additional scattering centers thus degrading the mobility, transconductance and drain current. Another important component of the damage is related to the build-up of bulk oxide charge [60, 66, 91, 100–104]. These traps are distributed spatially and over energy and thus have dramatically different times characterizing their charge communication (by tunneling and/or thermal activation) with Si. Depending on the length of this characteristic time, the states are conditionally separated on the border and oxide traps with densities N_{bt} and N_{ot} [41, 105–107] (see Figure 1.1).

In high-voltage devices the superposition of the interface states generation and oxide charge trapping accompanied by a specific field partition (compared to the conventional CMOS), Kirk effect (resulting in field redistribution) and mirror charge leads to a rather sophisticated stress behavior. For instance, the HC stress may alleviate the Kirk effect [61]. The situation is made more complicated because not only the magnitude of $N_{\mathrm{it}}/N_{\mathrm{ot}}$ is important but its spacial location is also to be considered. Thus, if in a conventional MOSFET the degradation of I_{dlin} is determined by the total interface state density, in the case of LDMOS the strongest impact on I_{dlin} is contributed by states located in the drain-side spacer region [59].

While the generation of interface states during HC stress appears universally acknowledged, a considerable controversy exists regarding the build-up of bulk oxide charge. The detrimental impact of N_{ot} on the device performance has been repeatedly reported in literature for practically every generation of CMOS devices and for various types of HC injection [2, 11–16, 108–110]. At the same time, for old CMOS technology nodes the concept that HCD is governed primarily by the interface traps localized near the drain region has been reported [5, 89, 90]. Therefore, for nodes smaller than 0.13μm and beyond even under the worst-case conditions occurring when $V_{\mathrm{ds}} = V_{\mathrm{gs}}$, the substrate current I_{sub} is shifted from its peak, and N_{it} still dominates

HCD [89, 90]. This idea was independently confirmed by Momose *et al.* [111] who applied HC stresses to MOSFETs with oxide thickness varying in the range of 1.5-3.8nm where direct tunneling between the gate and channel occurs. The authors claimed that in these thin gate oxides the trapped charge immediately tunnels away and cannot be stored in the bulk of the SiO_2 film. Mahapatra *et al.* in a series of works devoted to the determination of the N_{it} and N_{ot} spatial distributions by means of the charge-pumping technique have suggested that in various regimes the trapped charge does not play a significant role and the degradation is dominated by N_{it} creation [66, 102]. Finally, Polishchuk *et al.* investigated the reliability of MOSFET with an effective channel length of 100nm and an oxide thickness of 1.4nm. They have identified the dominant component of the degradation as N_{it} creation in spite the observed bulk trapping [112].

For modern ultra-scaled CMOS devices, the charge trapping in the oxide bulk becomes an issue again. Various research groups reported its presence as a drastically unfavorable factor deteriorating the performance of different types of devices with different gate oxides [113–117]. As for high-voltage devices, N_{ot} build-up is a commonly observed HCD component which is responsible for many peculiarities in the transformation of the transistor characteristics. [19, 21, 53, 66, 91, 100, 101, 103, 104]. The kinetics of charge trapping in the bulk of the dielectrica has been the subject of intensive research for more than four decades, i.e. starting from pioneering works by Nicollian *et al.* [118]. Further investigations resulted in a first-order trapping model based on the assumption that electron trapping is a stochastic process and disregarding Coulomb repulsion due to already trapped carriers and detrapping processes [119, 120]. The extension of the approach in order to cover these vital features was developed by Wolters *et al.* [100, 121] and then updated by Woltjer by linking the model to HC experimental data [101]. Therefore, the framework of this model includes injection, capture and emission. Such a sequence is typical for another intimate by related phenomenon, for trap-assisted tunneling, which is responsible for the stress-induced leakage current, a manifestation of HCD and NBTI [122–129]. For the present level of knowledge trap-assisted tunneling is an inelastic tunneling process assisted by phonon emission [130, 131]. The multiphonon based models for gate oxide leakage have been proposed and verified not only for the stress-induced leakage current observed in degraded dielectric layers but also for tunnel transport through oxides and SiO_2/high-k gate stacks [132–135]. This work is focused on the modeling of HCD phenomena using a reproduction of interface state density profile of a degraded device. The investigation of N_{bt} and N_{ot} defect contributions as well as their interplay with N_{it} is outside the scope of this work.

1.3.6 Temperature Behavior

To make the picture complete one should pay attention to the temperature behavior of HCD. Contrary to NBTI, which is made more severe at higher temperatures (see [137, 138]), hot-carrier induced damage usually becomes less pronounced at elevated temperatures [68–73]. Note that this traditional tendency is typical only for (relatively) long-channel devices while for ultra-scaled MOSFETs HCD becomes more significant at higher temperatures due to the dominant role of electron-electron scattering and its impact on the carrier distribution function [41, 74–76]. The essential features of hot-carrier induced degradation unequivocally demonstrate that the matter

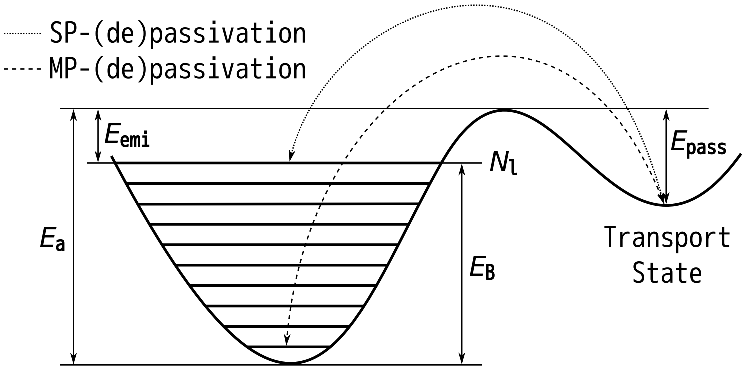


Figure 1.2: Two competing processes of Si – H bond-breakage: the single- and multiple-carrier mechanisms. The bond is interpreted as a truncated harmonic oscillator.

is controlled by the carrier distribution function. The DF allows us to judge how efficiently the carriers interact with the bonds or – in other words – how strong the bond dissociation reactions are. As a result, a comprehensive physics-based HCD model is expected to rely on consistent consideration of the microscopic mechanisms of defect creation and the carrier DF. In turn, for the calculation of the carrier distribution function a carrier transport module must be incorporated into the model.

1.4 Analysis and Comparison of Existing Physics-Based HCD Models

1.4.1 Hess Model

The main breakthrough in the area of HCD modeling associated with the Hess concept was the introduction of two competing mechanisms for Si – H bond-breakage, namely the single- and multiple-carrier processes, see Figure 1.2 [23, 32, 99]. A single-particle process is due to the interaction of a high-energy solitary carrier with the bond. During this interaction energy is transfered to the bond followed by its dissociation. Due to the large disparity of the electron mass and the mass of the hydrogen nucleus, the most probable way to deliver such an energy is via excitation of one of the bonding electrons to an antibonding state. As a consequence, a repulsive force acting on the H atom is induced followed by the release of hydrogen. The desorption rate of this process is [32]

$$R_{\mathrm{SP}} \sim \int\limits_{E_{\mathrm{th}}}^{\infty} I(E)P(E)\sigma(E)\mathrm{d}E, \tag{1.1}$$

where $I(E)$ is the flux of carriers with energies in the range of $[E; E{+}\mathrm{d}E]$, $\sigma(E)$ energy-dependent Keldysh-like reaction cross section, $P(E)$ the desorption probability, while the integration starts from the threshold energy E_{th}.

The first success of the theory was achieved when hydrogen/deuterium desorption induced by subsequent bombardment by several ("cold") carriers from the tip of a scanning tunneling microscope (STM) was investigated on hydrogen- and deuterium-passivated Si surfaces [98, 139–142].

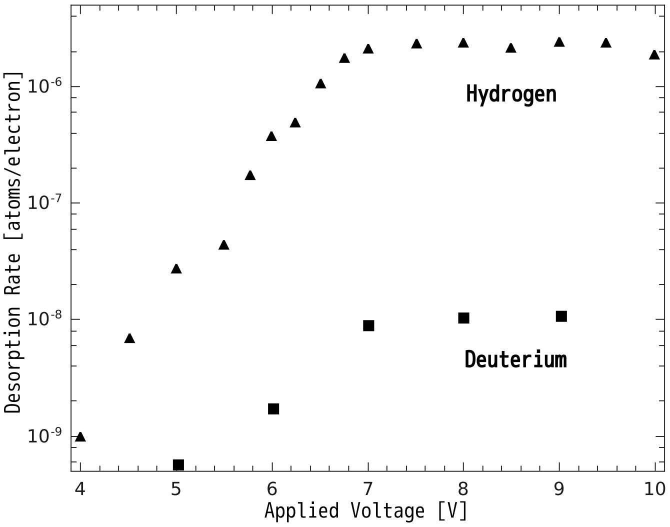


Figure 1.3: Disparity between H and D desorption rates induced by electrons tunneling from the STM tip on the passivated Si surface (data from [140]).

These experiments showed that the D-passivated surfaces are much more resistant with respect to electron bombardment compared to hydrogenated ones. In other words, substantially higher densities of STM currents are required to release the same amount of D atoms vs. H atoms. The difference in depassivation rates (Figure 1.3) may be more than two orders of magnitude at high voltages, which gave rise to the name "giant isotope effect". The similarities between the dangling bonds at surfaces and interfaces lead to the application of the theory to H-passivated Si/SiO_2 interfaces subjected to HC stress [88, 143–146].

This giant isotope effect was explained by the concept of multivibrational mode excitation by linking to the excitation of the phonon modes to a cascade of subsequent bombardments by interfacial carriers. It is assumed that the $Si-H$ bond can be considered as a truncated harmonic oscillator characterized by a system of levels in the corresponding quantum well and sketched Figure 1.2. The interface bombardment by carriers leads to the multivibrational mode excitation accompanied also by the phonon mode decay (corresponding rates are designated as P_u and P_d). The bond excitation is eventually terminated when the last bonded level N_l is reached. Being settled in this level the H ion can overcome the potential barrier separating the last level N_l and the transport state thereby breaking the bond and becoming mobile. The reciprocal process when the hydrogen ion is transferred from the transport to the bonded state is linked to the bond passivation. This scenario corresponds to the multiple-carrier mechanism.

To obtain an expression for the phonon excitation and decay rates (P_u, P_d), the formalism described in [32, 99] is applied. The electron flux can induce either phonon absorption (that is, bond heating) or phonon emission (related to the multivibrational mode decay). These absorption and emission rates are

$$\begin{aligned} P_\mathrm{ab} &= j_\mathrm{d}\sigma_\mathrm{ab}\exp\left(-\frac{\hbar\omega}{\mathrm{k_B}T_\mathrm{L}}\right) \\ P_\mathrm{emi} &= j_\mathrm{d}\sigma_\mathrm{emi}\left[\exp\left(-\frac{\hbar\omega}{\mathrm{k_B}T_\mathrm{L}}\right)-1\right], \end{aligned} \tag{1.2}$$

where j_d is the electron flux and $\sigma_\mathrm{ab}/\sigma_\mathrm{emi}$ are phonon absorption/emission reaction cross sections, Since the reaction cross section is energy-dependent, the electron flux is to be expressed over energies employing the carrier DF. Summarizing all these considerations one obtains the following expressions for (P_u, P_d) [99]

$$\begin{aligned} P_\mathrm{d} &\sim \int\limits_{E_\mathrm{th}}^{\infty} I(E)\sigma_\mathrm{ab}(E)[1 - f_\mathrm{ph}(E - \hbar\omega)]\mathrm{d}E, \\ P_\mathrm{u} &\sim \int\limits_{E_\mathrm{th}}^{\infty} I(E)\sigma_\mathrm{emi}(E)[1 - f_\mathrm{ph}(E + \hbar\omega)]\mathrm{d}E, \end{aligned} \tag{1.3}$$

where $I(E)$ is the carrier impact frequency on the surface per unit area per unit energy, $\hbar\omega$ phonon energy and phonon occupation numbers entering the expressions as $f_\mathrm{ph}(E)$. Finally, using (1.3), one obtains the bond-breakage rate corresponding to the MP-process as

$$R_\mathrm{MP} = \left(\frac{E_\mathrm{B}}{\hbar\omega} + 1\right)\left[P_\mathrm{d} + \exp\left(-\frac{\hbar\omega}{\mathrm{k_B}T_\mathrm{L}}\right)\right]\left[\frac{P_\mathrm{u} + \omega_\mathrm{e}}{P_\mathrm{d} + \exp(-\hbar\omega/\mathrm{k_B}T_\mathrm{L})}\right]^{-E_\mathrm{B}/\hbar\omega}, \tag{1.4}$$

with E_B being the energy of the last bonded level in the quantum well (Figure 1.2) and the phonon reciprocal life-time ω_e; $\mathrm{k_B}$ and T_L are the Boltzmann constant and the lattice temperature, respectively.

It is worth emphasizing that the particle flux differential $I(E)$ entering formulae (1.1,1.3) requires knowledge of the carrier DF. Thus, one of the main conclusions of the works by the group of Hess is the idea that for a proper description of HCD, the carrier energy distribution function is required. Another important achievement of this concept is that the isotope effect is essentially explained because different vibrational properties of Si − H and Si − D lead to different parameters of the corresponding quantum wells (see Figure 1.2), that is, to different positions of the last level E_B and phonon life-time.

Another characteristic feature of the Hess model is the assumption that the activation energy E_a for the Si − H bond-breakage rate is statistically distributed, see Figure 1.2. This assumption is supported by ab initio calculations using density functional theory [99,147]. As a consequence, the dispersion of E_a leads to different power-law slopes during degradation, see Figure 1.4 [23,24,40]. This is essential as simple first-order kinetics of Si − H bond-breakage with a single-valued activation energy lead to an exponential transition between the bonded and broken states within about a single decade in time. However, experimental observations demonstrate a double-power law of degradation

$$N_\mathrm{it} \sim \frac{p_1}{1 + (t/\tau_1)^{-\alpha_1}} + \frac{p_2}{1 + (t/\tau_2)^{-\alpha_2}}, \tag{1.5}$$

where τ_1/τ_2 are characteristic times and α_1/α_2 are two different sublinear time slopes ($\sim 1/2$). This time evolution has been explained assuming that two different types of traps (realized with the probabilities p_1 and p_2) contribute to HCD. These traps are similarly distributed and can be fit by the derivative of the Fermi-Dirac function with different mean values $E_\mathrm{am,1}/E_\mathrm{am,2}$ and

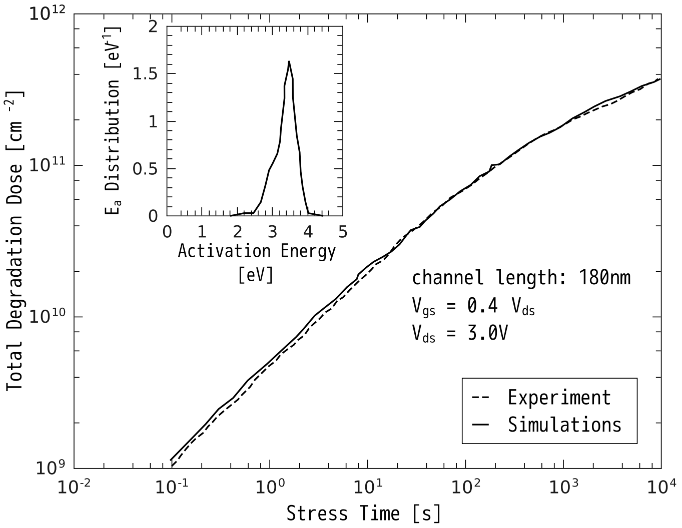


Figure 1.4: The total degradation dose (cumulative N_{it}) as a function of stress time: experiment vs. theory obtained for a 180nm device under worst-case stress conditions, i.e. $V_{\mathrm{gs}} = 0.4V_{\mathrm{ds}}$. Inset: distribution of Si – H bond-breakage activation energy. The data are borrowed from [24].

standard deviations $\sigma_{\mathrm{a},1}/\sigma_{\mathrm{a},2}$ [148], see Figure 1.4, inset

$$E_{\mathrm{a},1/2} \propto \frac{1}{\sigma_{\mathrm{a},1/2}} \frac{\exp\left(\dfrac{E_{\mathrm{am},1/2} - E_{\mathrm{a},1/2}}{\sigma_{\mathrm{a},1/2}}\right)}{\left[1 + \exp\left(\dfrac{E_{\mathrm{am},1/2} - E_{\mathrm{a},1/2}}{\sigma_{\mathrm{a},1/2}}\right)\right]^2}. \tag{1.6}$$

Despite the significant progress due to the work of Hess *et al.*, the interface traps are considered on a microscopic level and remain unconnected to the device level. For instance, the device life-time is estimated as the time when the concentration N_{it} reaches a certain level. Also, the degradation of such parameters as transconductance, linear drain current and so forth, is not really addressed. Furthermore, although the necessity for the evaluation of the carrier DF is acknowledged, this information has not been incorporated into the approach. As a result, the model operates with a static N_{it}, thereby not considering its distributed nature and that the details in the N_{it} distribution follows the features found in the DF.

1.4.2 Penzin Model

The Hess model was simplified for TCAD device simulations by Penzin [25] by employing phenomenological approximation. The model omits the microscopic level of defect generation (with the interplay of the SP- and MP-mechanisms as the essential attribute) but operates already on the device level. The bond rupture process is described by a kinetic equation for the passivated bond concentration n

$$\frac{\mathrm{d}n}{\mathrm{d}t} = -kn + \gamma(N_0 - n), \tag{1.7}$$

where k is the forward (depassivation) reaction rate while γ is the backward (passivation) rate and N_0 the total concentration of both "virgin" and broken bonds. The forward reaction rate has the following structure: $k = k_0 \exp(-E_\mathrm{a}/\mathrm{k_B}T_\mathrm{L})k_\mathrm{H}$ with the attempt frequency k_0 and k_H being the hot-carrier acceleration factor. This term is controlled by the "local hot carrier current" [25] I_HC:

$$k_\mathrm{H} = 1 + \delta_\mathrm{HC}|I_\mathrm{HC}|^{\rho_\mathrm{HC}}, \tag{1.8}$$

where δ_HC and ρ_HC are fitting parameters.

An important peculiarity of the Penzin approach is that the activation energy of bond dissociation depends on the hydrogen density and the transversal component of the electric field. The Si/SiO$_2$ interface (and its vicinity) is considered as a capacitor. The released hydrogen is assumed to be charged as well as the remaining dangling bonds. As a results, an additional electric field related to these charges is introduced into the system. This field prevents subsequent hydrogen ions from leaving the system and thus the potential barrier separating bonded and transport states is increased:

$$\begin{aligned} E_\mathrm{a} &= E_\mathrm{a}^0 + \delta|F|^\rho + \beta \mathrm{k_B} T_\mathrm{L} \ln \frac{N_0 - n}{N_0 - n^{(0)}}, \\ \beta &= 1 + \beta_\perp F_\perp, \end{aligned} \tag{1.9}$$

with E_a^0 being the activation energy in the absence of mobile H and $n^{(0)}$ the preexisting mobile hydrogen concentration. Since the system is considered as a capacitor, removal of charge from the capacitor is related to an additional energy required to compensate the change of the electric field. This energy is proportional to the capacitor electric field, which is in this case the normal (to the interface) component $F_\perp$ entering the expression. Additionally, the external electric field F can stretch or squeeze the bond, thereby changing the activation energy which is controlled by the term $\delta|F|^\rho$.

Similarly to the approach of Hess, the model employs a distribution of the activation energy and thus is able to represent the sublinear slope of degradation; Figure 1.5 demonstrates a reasonable agreement between experiment and theory. Although the model attempts to capture the carrier transport, this issue still remains vague. Formula (1.8) includes the questionable acceleration factor related to the "local hot carrier current". This criterion may be based for instance on the carrier temperature, which is related to the average energy of the distribution function (compare with the expressions (1.1) and (1.3)). As an adjacent problem, the information about the N_it profile is hardly achievable. Moreover, in spite of the efforts to link the kinetics of the trap generation and the device characteristics, one is not aware of a rigorous comparision against experimental device characteristics. Instead the soundness of the model is only proven by representing the experimental value of some cumulative N_it (Figure 1.5), but such a representation has already been obtained within the Hess approach, see Figure 1.4.

1.4.3 Reaction-Diffusion Framework

Another approach, which focused on the physical picture behind hot-carrier degradation was developed by the group of Alam [26,149]. The assumption was that NBTI and HCD are related

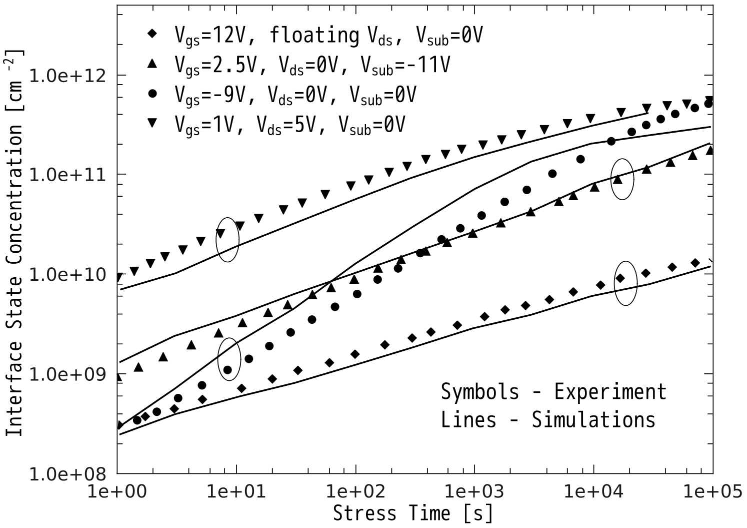


Figure 1.5: The interface state concentration N_{it}, simulation vs. experiment. An n-MOSFET with a gate length of 0.35μm and an oxide thickness of 6.5nm was subjected to hot-carrier stress at (1): V_{gs} = -9V, $V_{\mathrm{ds}} = V_{\mathrm{sub}}$ = 0V (V_{sub} is the substrate voltage); (2): V_{gs} = 12V, V_{sub} = 0V and floating source and drain; (3): V_{gs} = 1V, V_{ds} = 0V, V_{sub} = -11V; (4): V_{gs} = 2.5V, V_{ds} = 5V, V_{sub} = 0V. Data from [25].

to the breakage of silicon-hydrogen bonds, differing only in the driving force triggering this dissociation. Therefore, both phenomena are to be coupled within the same modeling framework. The authors claimed that since NBTI is just the breakage of $\mathrm{Si-H}$ bonds followed by hydrogen release and diffusion, NBTI and HCD are to be united within the reaction-diffusion concept.

Experimental observations demonstrated that time signatures of NBTI and HCD have different power-law slopes, i.e. the former one can be approximated by a $t^{1/4}$ law while the latter one better obeys a $t^{1/2}$ dependence, see Figure 1.6 and [26, 149]. The reaction-diffusion framework includes the following stages (are depicted in Figure 1.7 [149, 150]):

1. Creation of interface states via breaking $\mathrm{Si-H}$ bonds. This stage is reaction-limited and described by a t^1 dependence.
2. Hydrogen diffusion begins to take over with no more interface states created: $N_{\mathrm{it}} \sim t^0$.
3. Diffusion-limited phase with $t^{1/4}$ behavior.
4. Hydrogen diffuses away with unlimited diffusion velocity resulting in the 1/2 degradation time slope, i.e. $N_{\mathrm{it}} \sim t^{1/2}$.
5. Finally, saturation occurs when all the “virgin” $\mathrm{Si-H}$ bonds are depassivated: $N_{\mathrm{it}} \sim t^0$.

Therefore, it was assumed that NBTI is diffusion-limited, which explains its $t^{1/4}$ behavior while HCD is controlled by the 4th phase. However, this scenario presumes that in the case of HCD a transition from $t^{1/4}$ to $t^{1/2}$ is to be observed but the authors of [149] claim that in practice no experimental evidence of such a transition is known. Instead, they suggest that the difference in time slopes is related to the circumstance that NBTI is a 1D problem while HCD is a 2D phenomenon due to the non-uniform N_{it} distribution over the lateral coordinate. Since the $\mathrm{Si-H}$ bond-breakage event generates one mobile hydrogen and one interface trap one writes N_{it} =

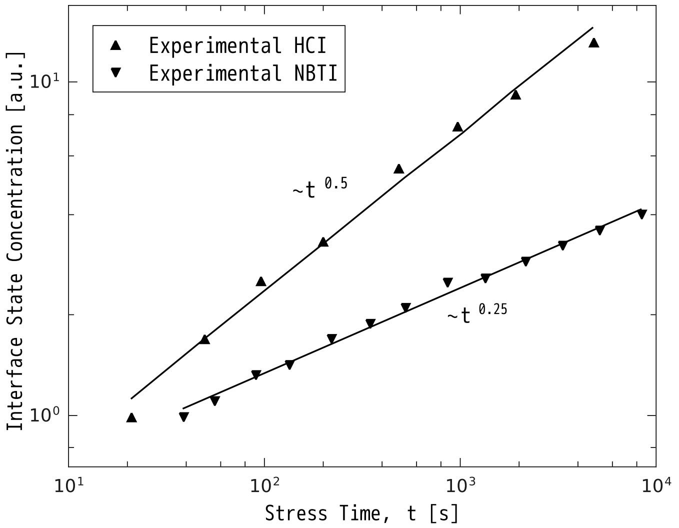


Figure 1.6: Different time slopes of hot-carrier induced degradation and NBTI. The data are borrowed from [26].

$\int N_\mathrm{H}(r,t)\mathrm{d}^3r$, ($N_\mathrm{H}(r,t)$ is the coordinate-dependent hydrogen concentration). The diffusion front moves as $(D_\mathrm{H}t)^{1/2}$ and thus NBTI- and HCD-related N_it are

$$\begin{aligned} N_\mathrm{it}^\mathrm{(NBTI)} &= (1/A_\mathrm{d}) \int\limits_0^{(D_\mathrm{H}t)^{1/2}} N_\mathrm{H}^{(0)} \left[1 - \frac{r}{(D_\mathrm{H}t)^{1/2}}\right] A_\mathrm{d}\mathrm{d}r = \frac{1}{2} N_\mathrm{H}^{(0)} (D_\mathrm{H}t)^{1/2}, \\ N_\mathrm{it}^\mathrm{(HCD)} &= \frac{\pi}{2A_\mathrm{d}} \int\limits_0^{(D_\mathrm{H}t)^{1/2}} N_\mathrm{H}^{(0)} \left[1 - \frac{r}{(D_\mathrm{H}t)^{1/2}}\right] r\mathrm{d}r = \frac{\pi}{12A_\mathrm{d}} N_\mathrm{H}^{(0)} (D_\mathrm{H}t), \end{aligned} \tag{1.10}$$

where D_H is the hydrogen diffusivity, A_d the area of the degraded spot and $N_\mathrm{H}^{(0)}$ is the H density at the interface. Assuming that $N_\mathrm{it}N_\mathrm{H}^{(0)} \sim$ constant, one obtains that $N_\mathrm{it}^\mathrm{(NBTI)} \sim (D_\mathrm{H}t)^{1/4}$ and $N_\mathrm{it}^\mathrm{(HCD)} \sim (D_\mathrm{H}t)^{1/2}$.

Despite its ability to explain the different time slopes of NBTI and HCD, this reaction-diffusion model suffers from serious shortcomings. First, within this framework it is assumed that both phenomena are diffusion limited. This implies, however, that once the stress is removed recovery should occur rather quickly. Recent NBTI data suggest, however, that interface state creation is reaction rather than diffusion limited [136,138,151]. Concerning HCD, the recovery is in general rather weak if there is any recovery at all. Second, the model does not rely on carrier transport, that is, it does not consider the driving force behind the trap generation. As a consequence, the N_it distribution and the localized nature of the damage are not addressed. To conclude, the attempt to describe only one essential aspect of HCD – the mechanisms of defect creation – not capturing others has been undertaken.

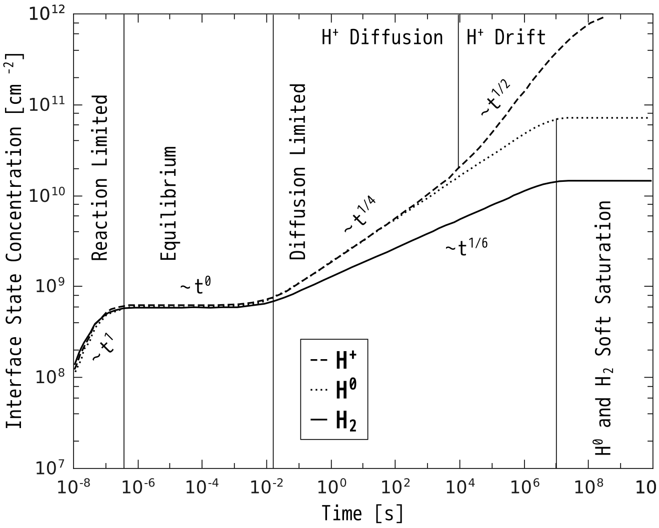


Figure 1.7: The main phases of the reaction-diffusion model applied to NBTI with different time slopes being marked. Data from [150].

1.4.4 The Energy-Driven Paradigm of Rauch and LaRosa

A new paradigm, or underlying concept, of nMOSFET hot-carrier behavior is proposed in works of Rauch and LaRosa [8, 17]. It is believed that the fundamental "driving force" is the available energy, rather than the peak lateral electric field, as in the lucky electron model. Two main issues associated with the approach of Rauch and LaRosa are:

- the increasing impact of electron-electron scattering on HCD at reduced channel lengths [8, 17] and
- the idea that in the case of scaled devices with channel lengths less than 180nm, the driving force of HCD is the carrier energy rather than the electric field [28, 29, 152].

Electron-electron scattering is of special interest in the case of ultra-scaled MOSFETs because in these devices the supply voltage is rather low and therefore the single-carrier mechanisms of $\mathrm{Si-H}$ bond-breakage were expected to be suppressed. This energy exchange mechanism, however, populates the "hot" fraction of the DF and modifies the shape of the DF, that is, results in a pronounced hump in the carrier distribution function, see Figure 1.8. Thus, the high-energy tail of the DF can expand deeper into energy than expected from the supply voltage. As a result, the contribution from the SP-mechanism is increased. Additionally, only electron-electron scattering defines the acceleration of HCD at elevated temperatures, which is pronounced in the case of extremely-scaled MOSFETs [41, 74–76].

The energy-driven paradigm presented by Rauch and LaRosa claims that beyond the 180nm node the driving force of HCD is the energy deposited by carriers, not the maximal electric field in the channel as it was in the "lucky electron model" [5]. Both the impact ionization rate as well as the rate of hot-carrier induced interface state generation is controlled by integrals of the form $\int f(E)S(E)\mathrm{d}E$, where $f(E)$ is the carrier DF and $S(E)$ the reaction cross section; compare this to the formula (1.1) used previously which has the same structure. The DF is a strongly decaying function of energy while $S(E)$ grows as a power-law. Hence, this trade-off results in a maximum

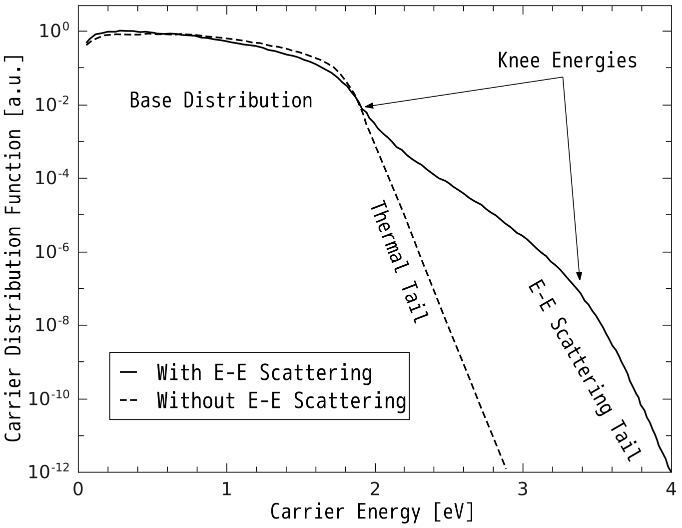


Figure 1.8: The impact of electron-electron scattering on the shape of the carrier energy distribution function. In the former case an additional hump in the distribution function high-energy tail appears. Data from [29].

of the rate pronounced at a certain energy (Figure 1.9) determined according to the criterion $\mathrm{d}(\ln f)/\mathrm{d}E = -\mathrm{d}(\ln S)/\mathrm{d}E$. This energy E_{knee} is called "knee" energy and is a weak function of the applied bias V_{ds}. Therefore, if the maximum of the product $f(E)S(E)$ is sufficiently narrow, it can be approximated by a delta-function and instead of integration in the entire energy range one can only calculate the value of the integrand for this energy. To conclude, the main message of the energy-driven paradigm is that one may avoid time-consuming calculations of the carrier DF substituting it by the empirical parameter. This parameter is proportional to the reaction rate calculated for $E = E_{\mathrm{knee}}$ which is defined by the bias conditions. This dependence will be discussed in the next subsection devoted to the Bravaix model.

Although this paradigm substantially simplifies the treatment of HCD, it suffers from some shortcomings. Indeed, one can see in Figure 1.9 that the maximum of the integrand $f(E)S(E)$ is not necessarily narrow and in the particular case shown by the authors [29] has a width of 1.5 – 2eV. Therefore, the concept of a dominant energy sounds doubtful. Furthermore, such a treatment of HCD does not deal with N_{it} as a distributed quantity and thus one of the main features of HCD – its strong localization – is not captured. Finally, as it was in the Hess approach, the device life-time is estimated by the interface state generation rate. However, it would be more reasonable to define it as the time when the degradation of V_{th} or I_{dlin}, etc., has reached a critical value.

1.4.5 Bravaix Model

The model of Bravaix inherits the main features of both the Hess and the Rauch/LaRosa approaches: the interplay between single- and multiple-carrier mechanisms as well as the idea that the damage is defined by the carrier DF, which is implemented using Rauch/LaRosa's "fashion", that is, calculations of the DF are substituted by operation/stress condition-related empirical factors.

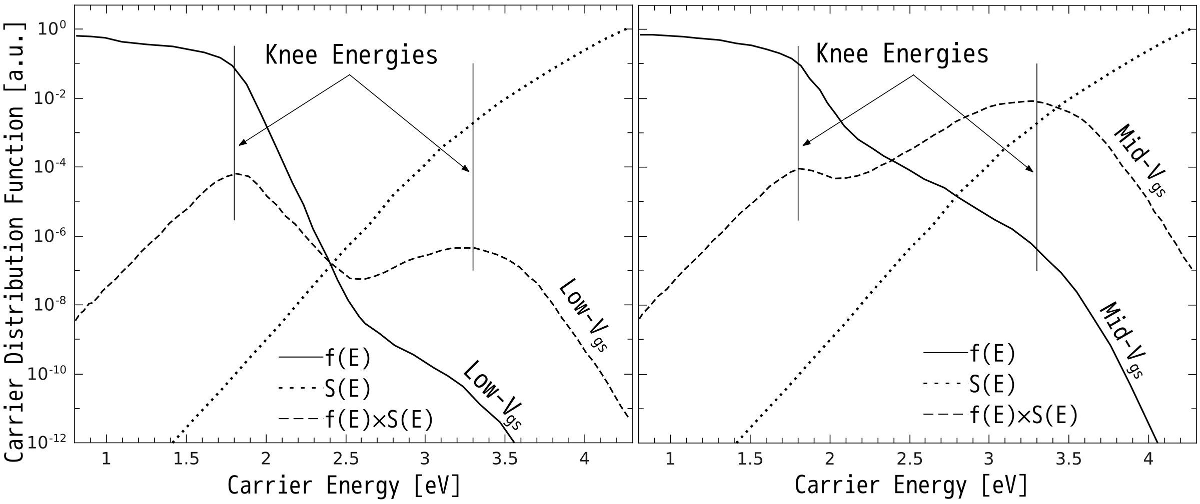


Figure 1.9: Schematic representation of the energy-driven paradigm. Knee energies shift depending on the applied voltage (the data borrowed from [29]).

To describe the MP-process the authors use the formalism of Hess where the $\mathrm{Si-H}$ bond is considered as a truncated harmonic oscillator. Following Hess they employ a system of rate equations to describe the kinetics of the oscillator [30, 32]

$$\begin{aligned}\frac{\mathrm{d}n_0}{\mathrm{d}t} &= P_\mathrm{d}n_1 - P_\mathrm{u}n_0,\\ \frac{\mathrm{d}n_\mathrm{i}}{\mathrm{d}t} &= P_\mathrm{d}(n_\mathrm{i+1} - n_\mathrm{i}) - P_\mathrm{u}(n_\mathrm{i} - n_\mathrm{i-1}),\\ \frac{\mathrm{d}n_{N_\mathrm{l}}}{\mathrm{d}t} &= P_\mathrm{u}n_{N_\mathrm{l}-1} - \lambda_\mathrm{emi}N_\mathrm{it}[\mathrm{H}^*],\end{aligned} \tag{1.11}$$

where $[\mathrm{H}^*]$ is the concentration of the mobile hydrogen and n_i is the occupancy of the the i^th oscillator level. In the last equation corresponding to the last bonded level (labeled as N_l, see Figure 1.2) the terms representing the passivation (i.e. from the transport to the last bonded state) and transition from the N_l to $N_\mathrm{l}-1$ state are omitted. The hydrogen released to the transport state is characterized by the rate $\lambda_\mathrm{emi} = \nu_\mathrm{emi}\exp(-E_\mathrm{emi}/\mathrm{k_B}T_\mathrm{L})$ with E_emi the height of the barrier separating bonded and transport states (see Figure 1.2) and ν_emi the attempt frequency.

Similar to [32], the phonon excitation/decay rates are written in a slightly modified form compared to (1.3)

$$\begin{aligned}P_\mathrm{u} &= \int I_\mathrm{d}\sigma\mathrm{d}E_\mathrm{e} + \omega_\mathrm{e}\exp\left(-\frac{\hbar\omega}{\mathrm{k_B}T_\mathrm{L}}\right),\\ P_\mathrm{d} &= \int I_\mathrm{d}\sigma\mathrm{d}E_\mathrm{e} + \omega_\mathrm{e},\end{aligned} \tag{1.12}$$

with I_d being the source-drain current. Employing the energy-driven paradigm the hot carrier acceleration factor – the first term in (1.12) – substituted by the empirical factor S_MP

Parameters	**Stretching**	**Bending**
E_B, eV	2.5	1.5
$\hbar\omega$, eV	0.25	0.075
w_e, ps^{-1}	1/295	1/10

Table 1.1: The parameters of the stretching and bending vibrational modes of the Si − H bond [30].

$$\begin{aligned} P_\mathrm{u} &= S_\mathrm{MP}(I_\mathrm{e}/\mathrm{q}) + \omega_\mathrm{e} \exp\left(-\frac{\hbar\omega}{\mathrm{k_B}T_\mathrm{L}}\right) \\ P_\mathrm{d} &= S_\mathrm{MP}(I_\mathrm{e}/\mathrm{q}) + \omega_\mathrm{e}. \end{aligned} \tag{1.13}$$

The solution of the system (1.11) for the case of weak bond-breakage rate ($\lambda_\mathrm{emi}t \ll 1$) leads to a square root time dependence of N_it [30]

$$N_\mathrm{it} = \left[N_0 \lambda_\mathrm{emi} \left(\frac{P_\mathrm{u}}{P_\mathrm{d}}\right)^{N_\mathrm{l}}\right] t^{1/2}. \tag{1.14}$$

In addition it was assumed that the bond is predominately situated in the ground state, i.e. $n_0 \approx \sum n_\mathrm{i} \approx N_0$. The MP-related interface state generation rate is

$$R_\mathrm{MP} \sim N_0 \left[\frac{S_\mathrm{MP}(I_\mathrm{d}/\mathrm{q}) + \omega_\mathrm{e}\exp(-\hbar\omega/\mathrm{k_B}T_\mathrm{L})}{S_\mathrm{MP}(I_\mathrm{d}/\mathrm{q}) + \omega_\mathrm{e}}\right]^{E_\mathrm{B}/\hbar\omega} \exp(-E_\mathrm{emi}/\mathrm{k_B}T_\mathrm{L}). \tag{1.15}$$

An important question is the choice of quantities (such as E_B, E_emi, $\hbar\omega$) defining the energetics of the Si − H bond. In fact, the two main vibrational modes of the Si − H bond are the stretching and bending mode [145] with the main parameters summarized in Table 1.1 [30]. However, as was previously shown, the experimental data is better fitted by the bending mode and therefore the values corresponding to this mode are employed. The formalism elaborated by Hess and co-authors and refined by Bravaix *et al.* with reasonably chosen simulation parameters allows for a perfect representation of the bond dissociation rate by the MP-mechanism (see [31] and the graph therein (Figure 1.10)).

Furthermore, the SP- and MP-mechanisms for defect creation are considered within Rauch's energy-driven paradigm, meaning that they are related to the regimes distinguished by Rauch *et al.* [17, 28]. The regime with low drain current and high carrier energies corresponds to the "hot-carrier" regime where the SP-mechanism plays the dominant role [31]. In this case the "lucky electron" model is valid and the device life-time is

$$1/\tau_\mathrm{SP} \sim \frac{I_\mathrm{d}}{W}\left(\frac{I_\mathrm{s}}{I_\mathrm{d}}\right)^m, \tag{1.16}$$

where I_s is the substrate current, W the device width and the factor m is the ratio between the powers in the impact ionization and interface state creation cross sections, i.e. $m \cong 11.0/4.0 \approx 2.7$.

Another limiting case corresponds to the high electron flux with low carrier energies. In this

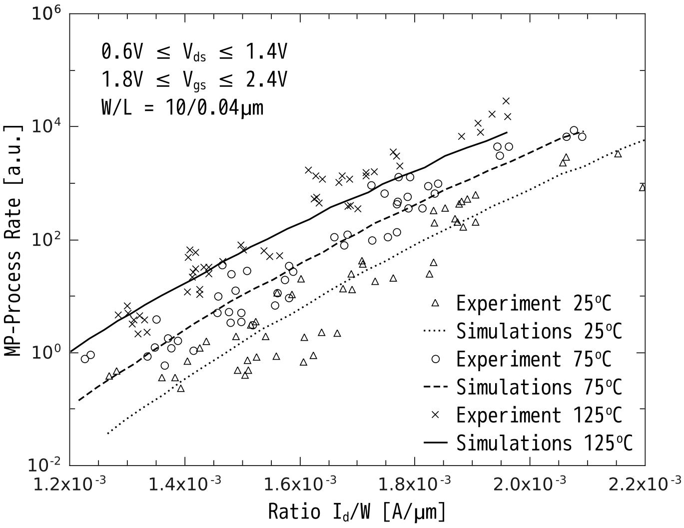


Figure 1.10: Experimental bond dissociation rate for the multiple-particle process vs. the theoretical one. The information about stress conditions is shown on the canvas. The data are borrowed from [31].

situation the MP-process dominates the bond dissociation and the device life-time is $1/R_{\mathrm{MP}}$ (1.15). According to the knee energy concept, $S_{\mathrm{MP}} \sim (V_{\mathrm{ds}} - \hbar\omega)^{1/2}$, and we have

$$\frac{1}{\tau_{\mathrm{MP}}} \sim \left[(V_{\mathrm{ds}} - \hbar\omega)^{1/2} \frac{I_{\mathrm{s}}}{W} \right]^{E_{\mathrm{B}}/\hbar\omega} \exp\left(-\frac{E_{\mathrm{emi}}}{\mathrm{k_B} T_L} \right) \approx \left(V_{\mathrm{ds}}^{1/2} \frac{I_{\mathrm{d}}}{W} \right)^{E_{\mathrm{B}}/\hbar\omega} . \tag{1.17}$$

The intermediate case with moderate drain current and moderate V_{ds} is governed by electron-electron scattering with the corresponding life-time [31]

$$\frac{1}{\tau_{\mathrm{EES}}} \sim \left(\frac{I_{\mathrm{d}}}{W} \right)^2 \left(\frac{I_{\mathrm{s}}}{I_{\mathrm{d}}} \right)^m . \tag{1.18}$$

This quadratic signature is due to impact ionization which generates electron-hole pairs which are still cold in terms of bond-breakage but being further accelerated by electron-electron scattering up to energies ensuring to trigger bond dissociation. Since under real device stress/operation conditions all modes are present, one writes the device life-time considering these competing mechanisms as

$$\frac{1}{\tau_{\mathrm{d}}} = \frac{K_{\mathrm{SP}}}{\tau_{\mathrm{SP}}} + \frac{K_{\mathrm{EES}}}{\tau_{\mathrm{EES}}} + \frac{K_{\mathrm{MP}}}{\tau_{\mathrm{MP}}}, \tag{1.19}$$

that is, different contributions are weighted with corresponding probabilities (K_{SP}, K_{EES}, K_{MP} which are fitting parameters) and summed. Figure 1.11 shows a fit of the model to experimental life-times.

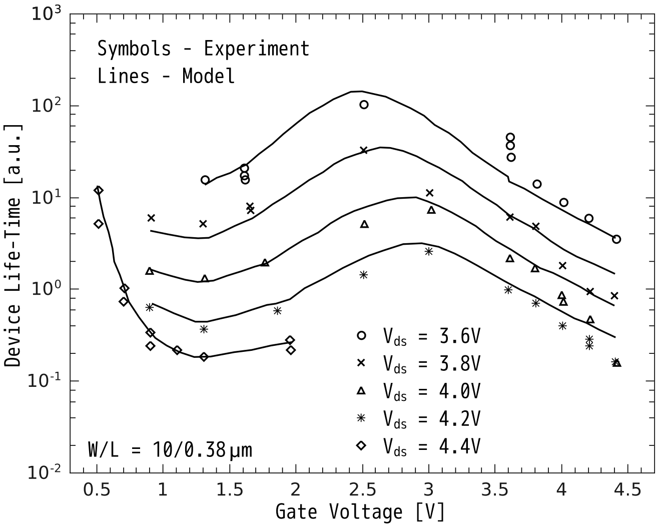


Figure 1.11: Comparison between the experimental device life-time and that calculated within the Bravaix framework (for devices fabricated in a 65nm node). The data are taken from [18].

1.5 HCD Model Based on the Carrier Distribution Function

This work presents a more detailed approach for hot-carrier degradation modeling and tries to more accurately capture the physical picture behind this phenomenon. This model incorporates the crucial features of the previous approaches for hot-carrier degradation modeling. In contrast to previous HCD models the aim of this work is to cover and link all levels related to this effect, starting from the microscopic mechanisms of defect generation and ending at the device level.

The main approaches to hot-carrier degradation modeling described in the previous sections were carefully analyzed and a comprehensive framework of a physics-based HCD model was established. As was demonstrated within the model by Hess, the degradation is controlled by the interplay between Single- and Multiple-Particle mechanisms of $\mathrm{Si-H}$ bond dissociation. This interplay is controlled by the way the carriers are distributed over energy, that is by the carrier energy distribution function. These considerations suggest that carrier transport and microscopic mechanisms of defect creation are two essential sub-tasks of the general problem. The Penzin model attempts to link the microscopic mechanisms of defect creation and the device level. While the energy-driven paradigm elaborated by Rauch and LaRosa is focused on the substitution of the DF by some simple approximations, the Bravaix model combines this paradigm with defect generation concepts.

In contrast to previously proposed models, the full hierarchical ladder connecting the microscopic origin of the phenomenon and device simulation level is arranged. Therefore, there are three main modules in the physics-based HCD model: the carrier transport module, the section responsible for the defect build-up, followed by the device simulation module (Figure 1.12). The carrier transport module allows for a thorough evaluation of the carrier energy distribution function for a particular device architecture. The DF represents populations of "hot" and "colder" carriers and thus controls the interplay between the SP- and MP-mechanisms and is then employed by the post-processor calculating defect density profiles. These profiles with information

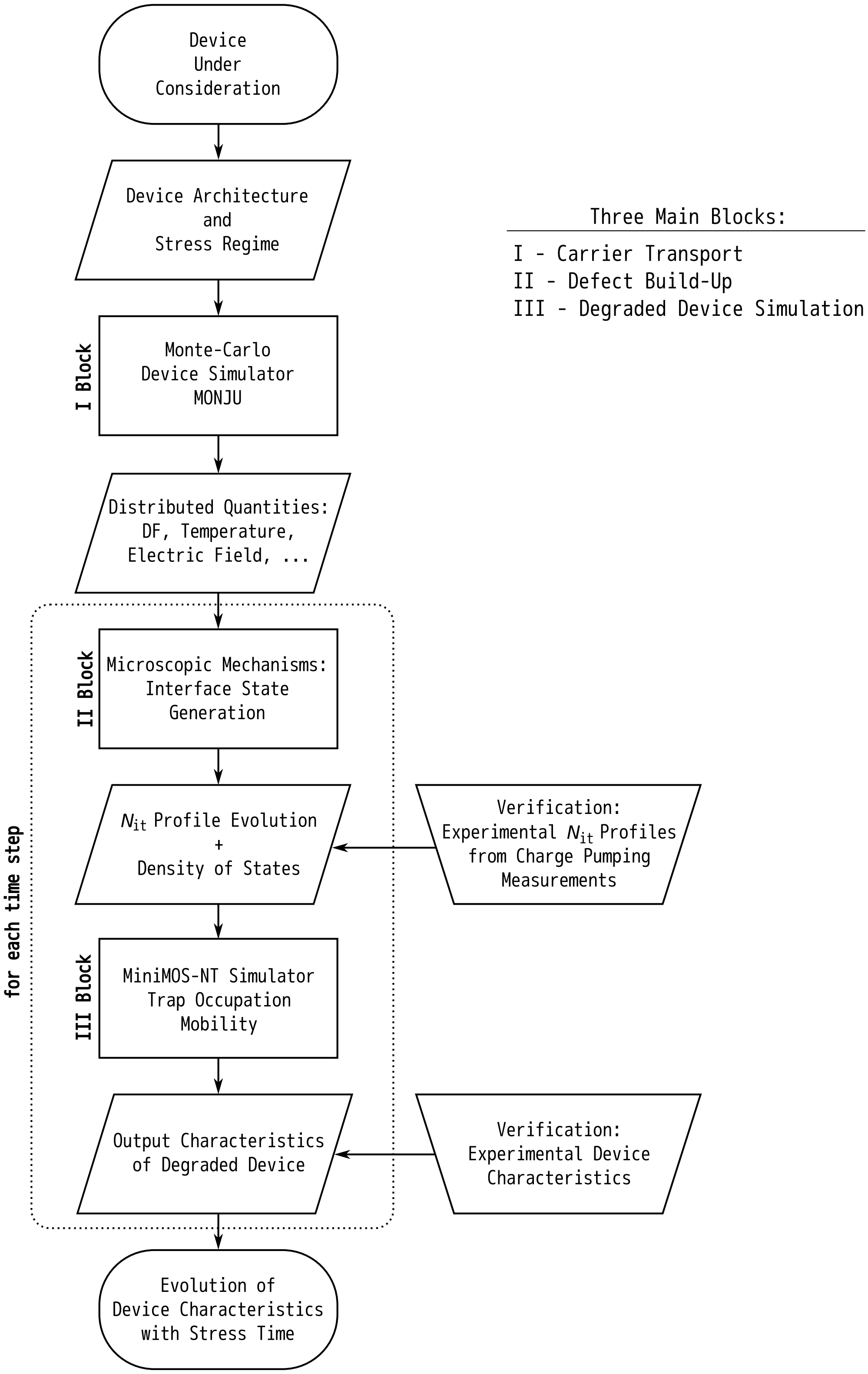


Figure 1.12: The flowchart of the proposed model for hot-carrier degradation depicting three main modules: carrier transport module, module for microscopic mechanisms for defect creation, and module for simulations of the degraded devices.

about the trap density-of-states (DOS) are used as input data for device simulations considering the distortion of device electrostatics and additional scattering events induced by charged traps. Furthermore, the device characteristics (such as output and transfer characteristics, transconductance, threshold voltage shift, etc.) of the degraded transistor are calculated. The feedback to calibrate the model is given by comparison with the experimental device characteristics. The principal breakthrough associated with this concept is the self-consistent consideration of both the transport and degradation aspects. Different fragments of the main framework are based on physical models presuming some assumptions and simplifications and these fragments act as sources of possible errors. Therefore, a thorough verification of all output information (used as the input for other modules) is to be performed. It should be emphasized that the double-check of the approach is being made. First the simulated N_{it} profile is compared vs. that extracted from charge-pumping measurements. Finally the simulated (employing these N_{it} profiles) set of device characteristics and the experimental one are collated.

2 Carrier Transport

The intensity of defect generation is controlled by the extent of how hot the carriers are. The measure of this may be the local value of the electric field, the carrier dynamic temperature, the extension of high-energy tails of the DF, etc. In the developed approach this carrier transport module provides the information regarding the carrier energy distribution function and is used for the simulation of the device characteristics. This information can only be obtained by a proper treatment of carrier transport, which is the essential sub-task of the HCD model. The carrier transport treatment is to be performed for transistors of different topologies covering the range from ultra-scaled short-channel devices to high-voltage transistors. In this Chapter the existing schemes for determining the distribution of hot carriers in semiconductor devices are described.

2.1 Boltzmann's Transport Equation

In the case of negligible quantum effects, the microscopic behavior of the system can be described within the Boltzmann transport equation. Otherwise, equations based on the Wigner-Boltzmann formula must be considered [153]. The BTE allows for the description of carriers motion in an inhomogeneous material with an arbitrary band structure to external and internal forces. The main components for mathematical models for semiconductor devices are derived from this semi-classical fundamental equation which reads [154]

$$\left\{\partial_t + \mathbf{v}(\mathbf{k}) \cdot \nabla_{\mathbf{r}} - \frac{\mathrm{q}}{\hbar}\,\mathbf{F}(\mathbf{r},t) \cdot \nabla_{\mathbf{k}}\right\} f(\mathbf{r},\mathbf{k},t) = S(\mathbf{r},\mathbf{k},t). \tag{2.1}$$

The solution of Boltzmanns equation is the time dependent carrier distribution function $f(\mathbf{r},\mathbf{k},t)$ in the six-dimensional phase space (three dimensions in real space $\mathbf{r}$ and three dimensions in momentum space $\mathbf{k}$). The right hand side of the BTE contains the rate of change of the distribution function due to collisions. The electron velocity $\mathbf{v}$ is determined from the band structure for the semiconductor under consideration. The electric field $\mathbf{F}(\mathbf{r},t)$ can be determined via Poissons equation [155]

$$\nabla_{\mathbf{r}} \cdot \mathbf{F} = \frac{\mathrm{q}}{\varepsilon}\left(p - n + N_{\mathrm{dd}}\right), \tag{2.2}$$

where N_{dd} is the differential doping density and ε is the permittivity of the material. The electron and hole concentrations, n and p, are the lowest order moments of the distribution function. If the distribution function is a known solution of (2.1), simultaneously with the Poisson equation, macroscopic quantities of interest such as the electron density, drift velocity and mean energy can be calculated according to

$$\begin{aligned} n(\mathbf{r},t) &= \frac{1}{(2\pi)^3}\int f(\mathbf{r},\mathbf{k},t)\mathrm{d}^3k, && (2.3)\\ \mathbf{v}(\mathbf{r},t) &= \frac{1}{(2\pi)^3}\frac{1}{n(\mathbf{r},t)}\int \mathbf{v}(\mathbf{k})f(\mathbf{r},\mathbf{k},t)\mathrm{d}^3k, && (2.4)\\ w(\mathbf{r},t) &= \frac{1}{(2\pi)^3}\frac{1}{n(\mathbf{r},t)}\int E(\mathbf{k})f(\mathbf{r},\mathbf{k},t)\mathrm{d}^3k. && (2.5) \end{aligned}$$

Due to its complicated nature, a general analytical solution of BTE can not be found. On the other hand, attempts to solve this equation numerically by discretization of the differential and integral operators is computationally very expensive. A widely used numerical method for solving the Boltzmann equation is the Monte Carlo approach. The MC method has been proven to produce reliable results but is computationally expensive. Moreover, if the distribution of high-energetic carriers is relevant, or if the carrier concentration is very low in specific regions of the device, Monte Carlo simulations produce a high variance in the results. Therefore, a common simplification is to investigate only some moments of the distribution function, such as the carrier concentration and the carrier temperature. In other words, the Boltzmann equation is simplified to the models, where the mentioned physical quantities are considered as the ensemble averages at the equilibrium state. For example, the carrier temperatures can be assumed to be in equilibrium with the lattice temperature. These approaches are usually limited in their predictive capability because they employ a number of fitting parameters used to match experimental data. Below, a short overview of the most widely used transport models is presented.

2.2 Hydrodynamic Transport Model

In the isotropic effective-mass approximation, the left-hand side of (2.1) can be used to deduce equations for electron density, drift-velocity and mean energy as defined in (2.3),(2.4), and (2.5) corresponding to equations of moments of zero, first and second-order in the wave vector. In order to close the hierarchy of equations a Maxwellian with electron temperature T_n is assumed for the distribution function

$$f(E) = A\,\exp\left(-\frac{E}{\mathrm{k_B}T_\mathrm{n}}\right). \tag{2.6}$$

Here, the prefactor A determines the carrier density, while the carrier temperature T_n determines the shape of the distribution function. Approximating the effect of scattering through the macroscopic relaxation time approximation (with relaxation time $\tau_\mathbf{v}$ for the drift velocity and τ_w for the mean energy) results in (e.g. [153])

$$\begin{aligned} \partial_t n + \nabla\cdot(n\mathbf{v}) &= 0, && (2.7)\\ \partial_t\mathbf{v} + \mathbf{v}\cdot\nabla\mathbf{v} + \frac{1}{n\,m^*}\nabla(n\mathrm{k_B}T_\mathrm{n}) &= -\frac{\mathrm{q}}{m^*}\,\mathbf{F} - \frac{\mathbf{v}}{\tau_\mathbf{v}}, && (2.8)\\ \partial_t w + \mathbf{v}\cdot\nabla w + \frac{1}{n}\nabla(n\mathrm{k_B}T_\mathrm{n}\mathbf{v} + \mathbf{Q}) &= -\mathrm{q}\mathbf{v}\cdot\mathbf{F} - \frac{w-w_0}{\tau_w}. && (2.9) \end{aligned}$$

Here, the mean energy is related to the electron T_n and the lattice temperature T_L by

$$\begin{aligned} w &= \frac{3}{2}\mathrm{k_B}T_\mathrm{n} + \frac{1}{2}m^*\mathbf{v}^2, && (2.10)\\ w_0 &= \frac{3}{2}\mathrm{k_B}T_\mathrm{L}, && (2.11)\end{aligned}$$

respectively. The drift energy $m^*\mathbf{v}^2/2$ is typically small compared to the thermal energy $3\mathrm{k_B}T_\mathrm{n}/2$. As a third-order moment the heat flux $\mathbf{Q}$ cancels for a Maxwellian, but is often assumed phenomenologically to be proportional to the gradient of the electron temperature. In the stationary regime the drift energy is negligible. After introducing the conduction current density $\mathbf{J} = -\mathrm{q}n\mathbf{v}$ and $u_\mathrm{Tn} = \mathrm{k_B}T_\mathrm{n}/\mathrm{q}$, one obtains

$$\begin{aligned} \nabla\cdot\mathbf{J} &= 0, && (2.12)\\ \mathbf{J} + \tau_\mathrm{v}\mathbf{J}\cdot\nabla\mathrm{v} &= \mathrm{q}\mu(n\mathbf{F} + u_\mathrm{Tn}\nabla n + n\nabla u_\mathrm{Tn}), && (2.13)\\ \nabla\cdot\mathbf{S} &= \mathbf{J}\cdot\mathbf{F} - \frac{3}{2}\mathrm{q}n\frac{u_\mathrm{Tn} - u_\mathrm{Tl}}{\tau_w}, && (2.14)\\ \mathbf{S} &= \mathbf{Q} - \frac{5}{2}u_\mathrm{Tn}\mathbf{J}, && (2.15)\end{aligned}$$

with the electron mobility $\mu = \mathrm{q}\tau_\mathbf{v}/m^*$. These system along with Poisson's equation (2.2) form the well known HD model.

2.3 Drift-Diffusion Transport Model

The DD model is the simplest current transport model which can be derived from the BTE by assuming that the electron temperature is equal to the lattice temperature. According to this approach the local temperature of the carrier gas can be estimated via the homogeneous energy balance equation

$$T_\mathrm{n} = T_\mathrm{L} + \frac{2}{3}\frac{\mathrm{q}}{\mathrm{k_B}}\tau_\mathrm{E}\mu F^2, \qquad (2.16)$$

where τ_E is the energy relaxation time. Consequently

$$f(E) = A\,\exp\left[-\frac{E}{3\mathrm{k_B}T_\mathrm{L}/2 + \mathrm{q}\tau_w\mu F^2}\right]. \qquad (2.17)$$

The substitution of (2.17) in (2.12) and (2.13) gives us

$$\begin{aligned} \nabla\cdot\mathbf{J} &= -\mathrm{q}G + \mathrm{q}\partial_t n, && (2.18)\\ \mathbf{J} &= \mathrm{q}\mu n\mathrm{F} + \mu\mathrm{k_B}T_\mathrm{L}\nabla n. && (2.19)\end{aligned}$$

where the net generation rate G. This approach for the BTE solution had been the mainstay for modeling transport in industrial applications for over thirty years. However, it should be noted that modeling of deep-submicron devices employing DD model is becoming more and more problematic. As long as the device channel length exceeds ~120nm, the simplified Drift-Diffusion and Hydrodynamic schemes for BTE solving are believed to be well-suited [156]. However, due to continuous device miniaturization these approaches fail for the description of ultra-scaled MOSFETs and thus a direct solution of BTE is pursued.

2.4 Monte Carlo Method

As mentioned above, one of the most popular computational methods to solve the BTE is the Monte-Carlo method, because it is flexible and allows for the incorporation of many features such as real band structure, various scattering mechanisms, and other physical models [154]. In the MC approach, carrier trajectories are traced at the microscopic level in a given device structure. As a consequence, carriers are subjected to the impact of the electric field, the device topology, and scattering mechanisms. Carrier flight and scattering are stochastically treated in accordance with probabilities describing the physical processes. A self-consistent MC solution (as DD and HD schemes) also considers the Poisson equation. At the same time, due to its stochastic nature, the method requires substantial computational resources and is very time-consuming. This disadvantage manifests itself while modeling processes determined by small populations, e.g. small currents or high-energy tails of the DF, and hence a trade-off between the computational efficiency and noise reduction must be considered. The latter problem becomes essential in the context of HCD modeling because, in many regimes, the bond-breakage process is triggered by particles characterized by high energies and low probabilities. In other worlds, due to the stochastic nature of the MC approach, the number of carriers to be simulated must be sufficiently high to obtain reliable statistics. Various statistical enhancement and iteration schemes employed in MC simulations result in additional varying degrees of simulation noise. Moreover, statistical enhancement aims at the reduction of the simulation time which is necessary for the desired device characteristics computation. Enhancement algorithms are quite useful when the system behavior is caused by rare events in the transport process, which is appropriate in the case for HC effects simulation. Most MC approaches employing statistical enhancement use population control techniques [154, 157]. They are based on the special splitting procedure of the particles entering a given phase space region of interest.

2.4.1 Multiple Refresh

The statistical enhancement method used in this work falls into the class of population control methods and conserves the total particle charge. It is based on a partitioned phase space, but instead of the particle weight the number of particles is fixed for each region. In general, the MC algorithm is not modified and is identical for all particles regardless of their statistical weight [154]. The particles can move from one region to another and thereby, the number of particles in a region can change during the MC simulation. To maintain the given particle number in a region, the simulation is halted at fixed time intervals. The number of particles is then examined and if this number deviates from the initial number in the region, the particle sub-ensemble of that region is replaced by an ensemble with a desired number of particles. This process is called refresh and because it is continually repeated – multiple refresh (MR) – to

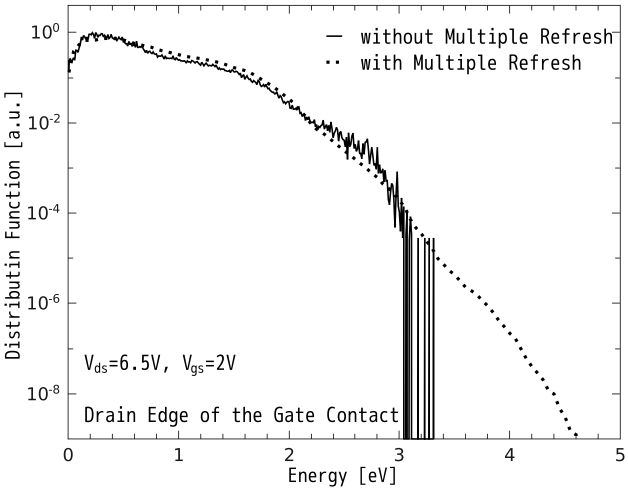


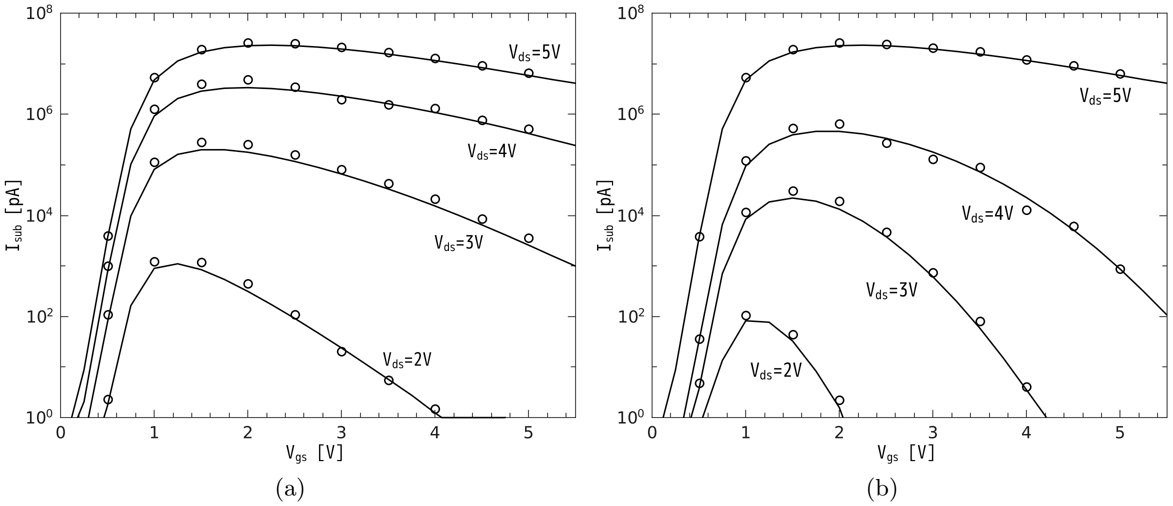
Figure 2.1: The electron distribution function for n-type MOSFET at $V_{\mathrm{ds}} = 6.5\mathrm{V}$ and $V_{\mathrm{gs}} = 2\mathrm{V}$ simulated with and without Multiple Refresh for the same computational time. Graph is taken for room temperature and corresponds to the drain edge of the gate contact.

maintain the given number of particles in all regions of the phase space during the simulation. The refresh does not produce any new information. The information contained in the sub-ensemble before the refresh is either identical after the refresh or it is reduced. In order to reduce the computational time only one region of interest may be refreshed. In Figure 2.1 the electron distribution function is shown for 5V n-type MOS transistor (for a more device details, see Section 3.1) at room temperature for $V_{\mathrm{ds}} = 6.5\mathrm{V}$ and $V_{\mathrm{gs}} = 2\mathrm{V}$ simulated with (dotted line) and without (straight line) MR. The depicted DF corresponds to the drain end of the gate contact. The simulation with and without MR were performed for the same computational time and the obtained DFs are identical within the statistical noise. In the case of the simulation performed without MR the statistical noise becomes too large for values of the distribution function below 10^{-4}. At the same time the simulation with MR demonstrates good results over the presented range of 10 orders of magnitude.

2.4.2 Non-Self-Consistent Simulation

In the case of a non-self-consistent (NSC) simulation the MC algorithm is performed for a given distribution of potential and carrier concentrations. This distribution can be obtained as the output of a previous self-consistent MC simulation or employing different simulation tools based on the DD or the HD model. As a consequence, the accuracy of the NSC simulation depends on the accuracy of the basic input information. In spite of a wrong reproduction of the low energy carrier distribution, the hot-carrier distribution is well simulated. As it was reported in [154] the NSC approach can be applied to a wide range of HC phenomena.

There exist several advantages to using a NSC method for efficient simulations of HC effects in comparison with a conventional MC simulation scheme [154]. However the main benefit of using NSC is a significant reduction in simulation time for reliable results in comparison with the self-consistent (SC) approach. The comparison of SC and NSC simulation results of a real 5V n-type MOSFET can be performed in order to confirm the validity of the NSC approach. The



(a)

(b)

Figure 2.2: Experimental (lines) and Non-Self-Consistently simulated (symbols) substrate current of n-type MOSFET with a gate length 0.5μm (a) and 2.0μm (b) for the series of the drain-source voltage $V_{\mathrm{ds}} = 2.0$, 3.0, 4.0, 5.0V.

substrate current due to impact-ionization is a key quantity for monitoring device degradation in nMOSFETs [5, 9]. Therefore, efficient simulation of hot-electron effects and especially of the substrate current is the most crucial criterion. The statistics were enhanced with the MR method in the case of the SC simulations as well as in the case of NSC. In Table 2.1 the resulting currents and corresponding simulation times are shown. Both methods demonstrate practically identical results for the substrate current. However, the simulation times differ by about three orders of magnitude. In Figure 2.2a and Figure 2.2b the measured and non-self-consistently calculated substrate current is shown for the 5V n-type MOSFETs with channel lengths of 0.5μm and 2.0μm, respectively. The simulation results clearly demonstrate a good agreement with the experimental data, i.e. the NSC simulation method is a reliable and efficient tool for hot-carrier simulations.

2.5 The Spherical Harmonics Expansion

An alternative approach for the calculation of the DF relies on an expansion of the DF by orthogonal polynomials [158, 159]. In particular, the spherical harmonics expansion (SHE) was

		Self-consistent		**Nonself-consistent**		**Experiment**
V_{ds} [V]	V_{gs} [V]	CPU Time [s]	I_{sub} [μA]	CPU Time [s]	I_{sub} [μA]	I_{sub} [μA]
2.0	2.0	53277	4.54×10^{-4}	445	4.04×10^{-4}	3.13×10^{-4}
4.0	2.0	55984	4.867	273	4.353	3.362
6.5	2.0	104267	129.1	285	128.1	128.6
6.5	2.5	154395	140.2	324	138.2	139.7

Table 2.1: Self-consistent and non-self-consistent simulation regimes - comparison of substrate current and computational time.

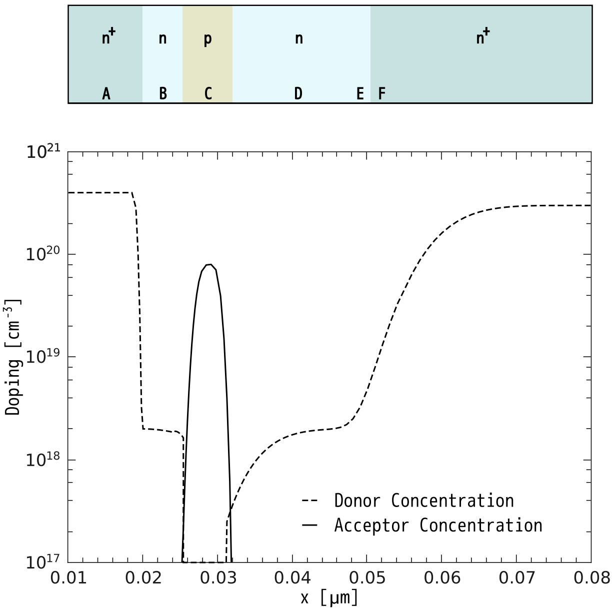


Figure 2.3: Schematic representation of the considered $n^+ \, p \, n^+$ structure (above) and 1D doping profile (below).

the first implementation that focused on the modeling of device characteristics, followed by a series of improvements in order to capture the full-band structure and magnetic field effects as well as various scattering mechanisms. The DF is expressed via a basis set of spherical functions and the discretization leads to a significant number of unknown variables. As a consequence, the problem of insufficient computer memory arises. In spite of the fact that two-dimensional implementations with lowest-order expansion were already developed in early 90s [161], the 2D realization with a higher expansion order has been achieved only recently [162]. The number of polynomials kept in the SHE implementation controls the accuracy of computational results. Simultaneously, the amount of computational memory drastically grows with increased expansion order, because of such peculiarities as the strong gradient of the DF and the large numerical range of values. To remedy this difficulty, the H-transform, the maximum entropy dissipation and the staggered grid schemes have been proposed [160]. Without these stabilization schemes, numerical instability occurs and simulations of 2D devices would be doubtful (see [159] and references therein).

The available version of the SHE solver is applicable only for 1D structure/device. Therefore, in this work the Monte-Carlo method is used for a BTE solution. The SHE approach is used only to control the results produced by the MC device simulator. For example, in Figure 2.4 the electron distribution functions calculated by the SHE and MC(NSC+MR) approaches for simple $n^+ \, p \, n^+$ structure (see Figure 2.3) are presented. The agreement between curves confirms the validity of the applied Monte-Carlo simulation schemes.

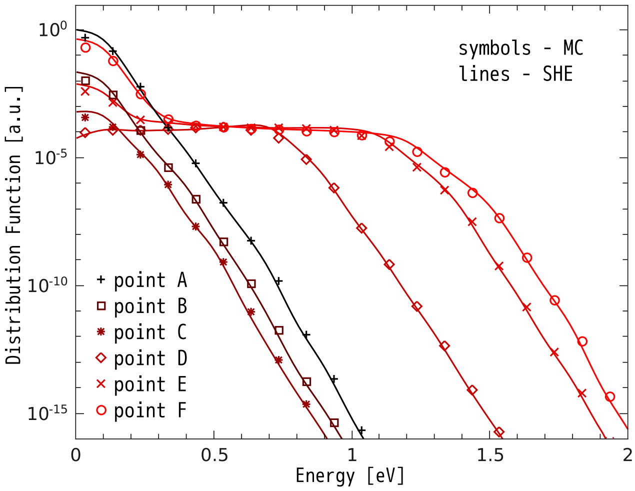


Figure 2.4: The electron distribution functions provided by the Spherical Harmonics Expansion and Monte-Carlo (Non-Self-Consistent and Multiple Refresh simulation regimes) approaches. Different letters corresponds to a series of considered position. A rather good agreement confirms the validity of the applied Monte-Carlo simulation schemes.

2.6 Employed Simulation Tools

As described above, theoretical models and simulation approaches have been continuously improved and implemented in simulators by various scientific groups. The improvements and developments of some groups already spans decades and continues at the present moment. A list of used simulation tools employed in this work is presented below.

- **MONJU** is a 2D full-band Monte-Carlo device simulator [154]. Such mechanisms as impact ionization, electron-phonon and scattering on ionized impurities are employed in MONJU. The simulator is capable of simulating real devices with real doping profiles. It is employed for the evaluation of the carrier DF, i.e. represents the transport module. MONJU is developed by the group of Prof. Jungemann [154].

- **MiniMOS-NT** is the 2D device and circuit simulator developed at our Institute [163]. The Drift-Diffusion and Hydrodynamics schemes for the Boltzmann transport equation solution are employed in MiniMOS-NT, which are valid for long-channel devices. MiniMOS-NT is much faster than any Monte-Carlo device simulator, thus, MONJU is used only for DF calculations.

- **SPRING** is a spherical harmonics expansion based solver of the Boltzmann transport equation [158, 162]. The program is developed by the group of Prof. Jungemann. The available version is applicable for simple 1D structures. It is used for verification of the simulation parameters for evaluation of the carrier DF by MONJU.

The employed simulation procedure to calculate the carrier DF consists of several steps, and is described by the flowchart shown in Figure 2.5. The device topology with a doping profile is obtained from the process simulator and is used as input data for the initial MiniMOS-NT simulation. During this simulation, potential and carrier density distributions are calculated for the chosen stress voltages. The resulting information is then used as input data for MONJU with

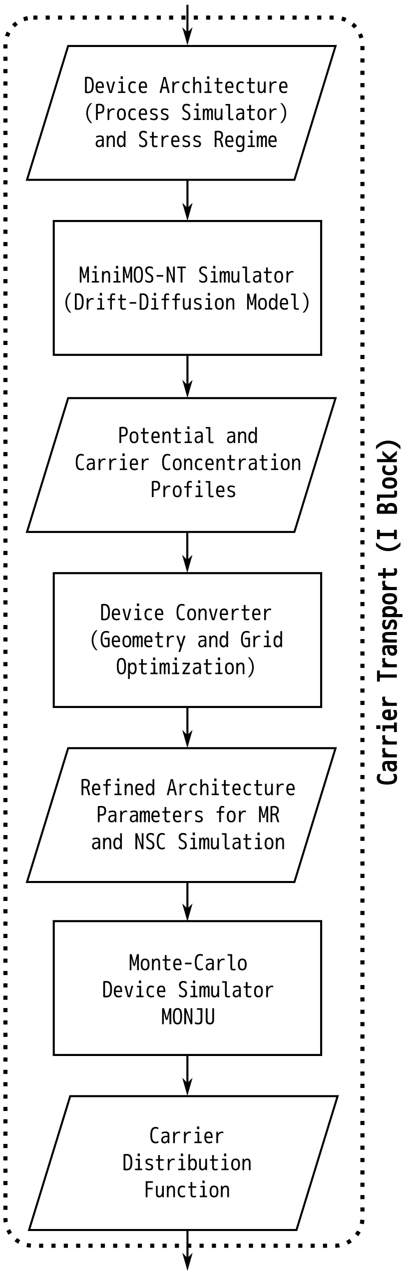
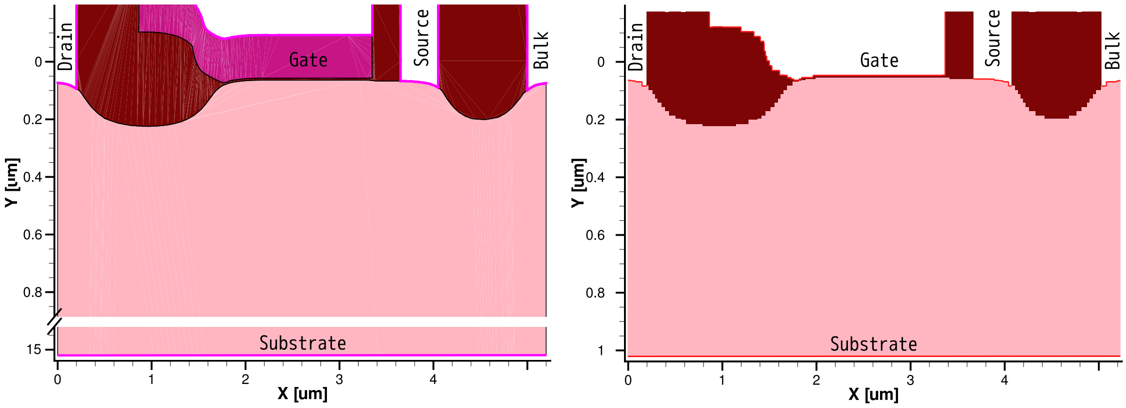


Figure 2.5: The simulation flowchart for the carrier transport block of developed HCD model.

a NSC/MR approach and for the optimization of the device topology. In some cases, the silicon bulk of the transistor may be as large as 15μm. It is obvious that the calculation of the physical quantities using the MC approach is a challenge for such a huge area. Moreover, according to the suggested scheme for HCD modeling, one is interested only in the carrier distribution function at the device interface. Therefore, the simulation efficiency can be increased by reducing the size of the silicon bulk, e.g. the device can be cut at the potential profile saturation point. Furthermore, in the present MC device simulators [154,157], because of computational efficiency, a rectangular grid is used. That is, due to the different types of the simulation grid which are used by MiniMOS-NT and MONJU (triangular and rectangular, respectively), data conversion is necessary. It should be mentioned that the realization of the MC method is possible in the case



(a) (b)

Figure 2.6: The cross section of the high voltage n-type LDMOS transistor (a) before and (b) after the conversion procedure.

of triangular grid type but leads to a significant increase of the computational time. An example of the converted n-type LDMOS transistor is demonstrated in Figure 2.6. At the final step of the transport block, the carrier DF is calculated by MONJU. Note, there is no requirement to perform the full procedure every time from the beginning. Several steps, such as the device geometry optimization might be omitted in the case of subsequent simulations.

3 Defect Creation

A comprehensive model for hot-carrier degradation should carefully capture the physical picture behind this detrimental effect. The developed model incorporates the main features of the previous approaches for hot-carrier degradation modeling. However, contrary to the previous HCD models the main aim of this work is to cover and link all levels related to this effect, starting from microscopic mechanisms of defect generation and ending at the device level. This means that essential peculiarities of this phenomenon (see Section 1.3) must be represented by the model. Among them are the interplay between SP- and MP-processes of $\mathrm{Si-H}$ bond dissociations [30, 32, 41, 99], the strong localization of the damage [2, 65], and the degradation saturation achieved at long stress times [29, 41]. Finally, the damage induced not only by the minority carriers (forming the channel) but also by the majority carriers is to be considered [27, 164].

3.1 Microscopic Model for Interface State Creation

The carrier transport module described in Section 2.6 allows for a thorough evaluation of the carrier energy distribution function for a particular device architecture. The DF represents populations of "hot" and "colder" carriers and thus controls the interplay between the SP- and MP-mechanisms (see Section 1.4). Under real hot-carrier stress conditions both SP- and MP-mechanisms are present and their interplay is described by the shape of the distribution function reflecting the topological features of a particular device as well as the bias conditions. The DF enters the corresponding rates of the bond depassivation processes via the carrier acceleration integral

$$\begin{aligned} I_{\mathrm{SP}} &= \int\limits_{E_{\mathrm{th,SP}}}^{\infty} f(E) g(E) \sigma_{\mathrm{SP}}(E) v(E) \mathrm{d}E, \\ I_{\mathrm{MP}} &= \int\limits_{E_{\mathrm{th,MP}}}^{\infty} f(E) g(E) \sigma_{\mathrm{MP}}(E) v(E) \mathrm{d}E. \end{aligned} \tag{3.1}$$

The integrand in (3.1) is obtained via the multiplication of the carrier flux with energies in the range $[E; E + \mathrm{d}E]$ by the probability to launch the reaction by a carrier at that energy. This flux is the product of the carrier distribution $f(E)$, density-of-states $g(E)$, and velocity $v(E)$. Therefore, the SP- and MP-related AIs have the same functional structure and differ only in their parameters. $E_{\mathrm{th,SP}}/E_{\mathrm{th,MP}}$ are (different) threshold energies of the SP/MP-processes and $\sigma_{\mathrm{SP}}/\sigma_{\mathrm{MP}}$ the Keldysh-like reaction cross sections [32, 99] – i.e. of the form

$$\sigma_{\mathrm{SP}}(E) = \sigma_{\mathrm{SP},0}(E - E_{\mathrm{th,SP}})^{p_{\mathrm{SP}}}, \tag{3.2}$$

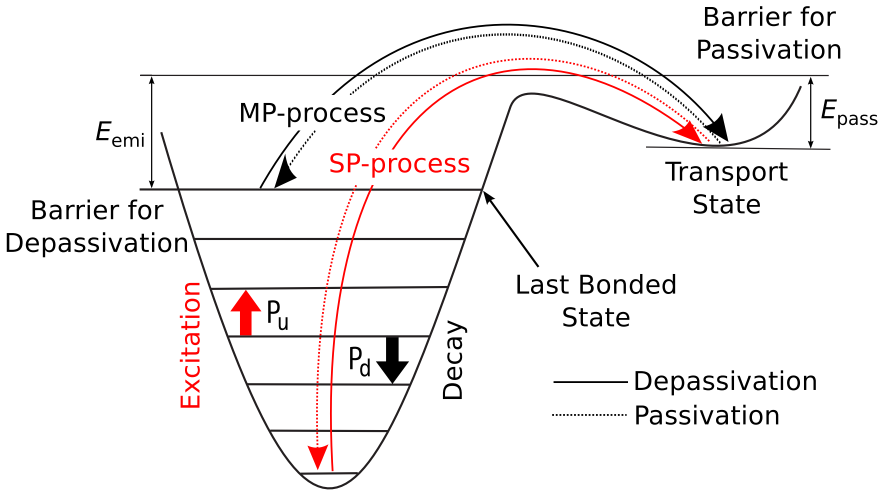


Figure 3.1: Schematic representation of the SP- and MP-processes.

and the cross section for the MP-process has the same structure. The single-particle bond-breakage case refers to the excitation of the bonding electron to an antibonding state while the multivibrational mode excitation is linked to the excitation/decay of the phonon modes [32,99].

To find the interface state generation rate corresponding to the SP-process one should multiply the acceleration integral by the attempt rate ν_{SP} (which is the fitting parameter in the model), i.e. $\lambda_{\mathrm{SP}} = \nu_{\mathrm{SP}} I_{\mathrm{SP}}$. Assuming the process follows first-order kinetics, the following dependence of the SP-related interface state density N_{SP} on time is obtained

$$N_{\mathrm{SP}}(t) = n_0 \left[1 - \exp(-\lambda_{\mathrm{SP}} t)\right], \tag{3.3}$$

where n_0 is the concentration of the virgin (passivated) $\mathrm{Si-H}$ bonds which can be broken. The situation with the MP-process is more complicated, because this process is linked to the excitation of the phonon modes by a cascade of the subsequent bombardments of the interface by carriers [30–32, 99]. Note that the $\mathrm{Si-H}$ bond can relax from an excited state to a lower one, while the balance with a reciprocal process also must be considered. Therefore, the bond is treated as a truncated harmonic oscillator (Figure 3.1) characterized by the ladder of bonded levels. The last level is designated as N_{l}. The $\mathrm{Si-H}$ bond-breakage process is described by the system gradually climbing the ladder of energetic states, a process which is eventually terminated when hydrogen leaves the last bonded level for the transport state (Figure 3.1). The reaction rate is defined by the height of the barrier E_{emi} separating the last level N_{l} and the transport state. Similarly, the passivation process is related to the hydrogen jumping into the opposite direction, determined by the barrier height E_{pass}. The corresponding rates $(P_{\mathrm{emi}}, P_{\mathrm{pass}})$ are assumed to obey the Arrhenius law

$$\begin{aligned} P_{\mathrm{emi}} &= \nu_{\mathrm{emi}} \exp\left(-\frac{E_{\mathrm{emi}}}{\mathrm{k_B} T_{\mathrm{L}}}\right), \\ P_{\mathrm{pass}} &= \nu_{\mathrm{pass}} \exp\left(-\frac{E_{\mathrm{pass}}}{\mathrm{k_B} T_{\mathrm{L}}}\right), \end{aligned} \tag{3.4}$$

where (ν_{emi}, ν_{pass}) are attempt frequencies for the processes (fitting parameters), while T_{L} is the lattice temperature. To obtain an expression for the phonon excitation and decay rates ($P_{\mathrm{u}}, P_{\mathrm{d}}$) (Figure 3.1) one can follow the formalism described in the papers by Hess *et al.* [32, 99] (for a more details, see Section 1.4)

$$\begin{aligned} P_{\mathrm{u}} &= w_{\mathrm{e}} \exp\left(-\frac{\hbar\omega}{\mathrm{k_B}T_{\mathrm{L}}}\right) + I_{\mathrm{MP}}, \\ P_{\mathrm{d}} &= w_{\mathrm{e}} + I_{\mathrm{MP}}, \end{aligned} \tag{3.5}$$

with w_{e} being the phonon frequency and $\hbar\omega$ the distance between the oscillator levels. The first two terms describe the population of the oscillator levels in the absence of the carrier acceleration while in the presence of the electron flux an additional term $I_{\mathrm{SP}}/I_{\mathrm{MP}}$ enters the expression. The kinetics of the MP-process is determined by a system of rate equations (see e.g. [30])

$$\frac{\mathrm{d}n_0}{\mathrm{d}t} = P_{\mathrm{d}} n_1 - P_{\mathrm{u}} n_0, \tag{3.6}$$

$$\frac{\mathrm{d}n_{\mathrm{i}}}{\mathrm{d}t} = P_{\mathrm{d}}(n_{\mathrm{i+1}} - n_{\mathrm{i}}) - P_{\mathrm{u}}(n_{\mathrm{i}} - n_{\mathrm{i-1}}), \tag{3.7}$$

$$\frac{\mathrm{d}n_{N_{\mathrm{l}}}}{\mathrm{d}t} = P_{\mathrm{u}} n_{N_{\mathrm{l}}-1} - P_{\mathrm{d}} n_{N_{\mathrm{l}}} - P_{\mathrm{emi}} n_{N_{\mathrm{l}}} + \widetilde{P}_{\mathrm{pass}} N_{\mathrm{MP}}^2. \tag{3.8}$$

As opposed to [30], in (3.8) all 4 terms are maintained (cf. with [30], Eq. 22): the first two are related to the oscillator excitation/deexcitation while the last two control the bond rupture and passivation process. It is assumed that the MP-related interface state concentration is equal to the density of the mobile hydrogen which is needed for the bond passivation. As a result, N_{MP} enters (3.8) in the second degree. To satisfy the dimensionality $\widetilde{P}_{\mathrm{pass}} = P_{\mathrm{pass}}/n_0$ is used in (3.8).

The model furthermore assumes that a steady-state between the individual energy levels is established rather quickly, much quicker than the hydrogen hopping step (depassivation/passivation). As a consequence, these reactions can be considered quasi-independently. Thus, first the last two terms in (3.8) are omitted and one considers the oscillator in equilibrium, i.e. one assumes that $\mathrm{d}n_{\mathrm{i}}/\mathrm{d}t = 0$ for each i and recurrently finds that $n_{\mathrm{i}}/n_0 = (P_{\mathrm{u}}/P_{\mathrm{d}})^{\mathrm{i}}$ (for the sake of simplicity it is also assumed that the oscillator is predominately in its ground state, i.e. $n_0 = \sum_{\mathrm{i}} n_{\mathrm{i}}$). Then it is possible to "switch on" hydrogen transitions and obtain the interface state evolution with time

$$N_{\mathrm{MP}} = n_0 \left(\frac{\lambda_{\mathrm{emi}}}{P_{\mathrm{pass}}} \left(\frac{P_{\mathrm{u}}}{P_{\mathrm{d}}}\right)^{N_{\mathrm{l}}} \left[1 - \exp(-\lambda_{\mathrm{MP}} t)\right] \right)^{1/2}. \tag{3.9}$$

It is worth emphasizing that in the case of weak stresses and/or short stress time, i.e. when $\lambda_{\mathrm{emi}} t \ll 1$, the expression (3.9) transforms to the root time dependence of N_{MP} as reported in [30].

The two main $\mathrm{Si-H}$ vibrational modes are the stretching and bending modes [30, 145], with their parameters listed in Table 1.1 [30]. It has been previously shown [63] that experimental data is better fitted by the bending mode and therefore for simulations one uses the parameters for this mode. The SP- and MP-processes lead to two types of traps which must have different distributions over energy in order to match with experimental data. Therefore, these mechanisms are considered independently, i.e. the total interface state density is calculated as the superposition of SP- and MP-related components weighted with corresponding probabilities

$$N_{\mathrm{it}} = p_{\mathrm{SP}} N_{\mathrm{SP}} + p_{\mathrm{MP}} N_{\mathrm{MP}}. \tag{3.10}$$

Only charged interface states contribute to the device performance degradation. Therefore, while modeling the transfer characteristic evolution during the hot-carrier stress one should consider effective charges stored in the interface states, not the total concentration N_{it}. These effective charges (Q_{SP} and Q_{MP}, the total Q_{it} is their sum) are defined as

$$Q_{\mathrm{SP/MP}} = \int\limits_{-\infty}^{\infty} g_{\mathrm{SP/MP}}(E) f_{\mathrm{oc}}(E, E_{\mathrm{F_n}}(x))\mathrm{d}E, \tag{3.11}$$

where $g_{\mathrm{SP}}(g_{\mathrm{MP}})$ are the DOS for the SP(MP)-related traps and f_{oc} is the carrier distribution function obtained for device operation conditions. The coordinate-dependent position of the quasi-Fermi level of electrons is designated as $E_{\mathrm{F_n}}$. Note that the functions g_{SP} and g_{MP} are coordinate dependent because of the normalization conditions, i.e.

$$\int\limits_{-\infty}^{\infty} g_{\mathrm{SP/MP}}(E, x)\mathrm{d}E = N_{\mathrm{SP/MP}}(x). \tag{3.12}$$

The lateral coordinate also enters the DF for operation conditions because the quasi-Fermi level for carriers captured by traps is also position dependent. The model is thus calibrated in order to represent the degradation of the linear drain current I_{dlin} over a wide range of stress and/or operation conditions by a proper determination of Q_{SP} and Q_{MP}.

The carrier mobility in the device channel reduced after irradiation because of radiation-induced charges at the interface produces Coulomb scattering centers, which affect the motion of carriers in the channel. Therefore, the effective mobility, $\mu(N_{\mathrm{it}})$, which is related to the build-up of interface trapped charge, N_{it}, can be defined through the equation [165, 166]

$$\mu(N_{\mathrm{it}}) = \frac{\mu_0}{1 + \alpha N_{\mathrm{it}}}, \tag{3.13}$$

where μ_0 is the mobility in absence of interface states, α is a proportionality coefficient.

For the evaluation of the model a high voltage 5V n-type MOSFET fabricated on a standard 0.35μm technology was used. The device cross-section/contour plot is schematically depicted in Figure 3.2. Since we are dealing with a long-channel device with a length of ~0.5μm and a relatively high operating voltage of 5V, it is possible to estimate that the SP-process will dominate hot-carrier degradation. Figure 3.3a demonstrates the evolution of the carrier distribution function along the channel. One can see that near the source and drain the DF behaves like a heated Maxwellian, but has deep high energy-tails at the drain end of the gate. The carrier acceleration integral plotted vs. the coordinate x (Figure 3.3b) features its peak near the area with the most extended high-energy tails of the DF. Such a behavior proves that the hot-carrier

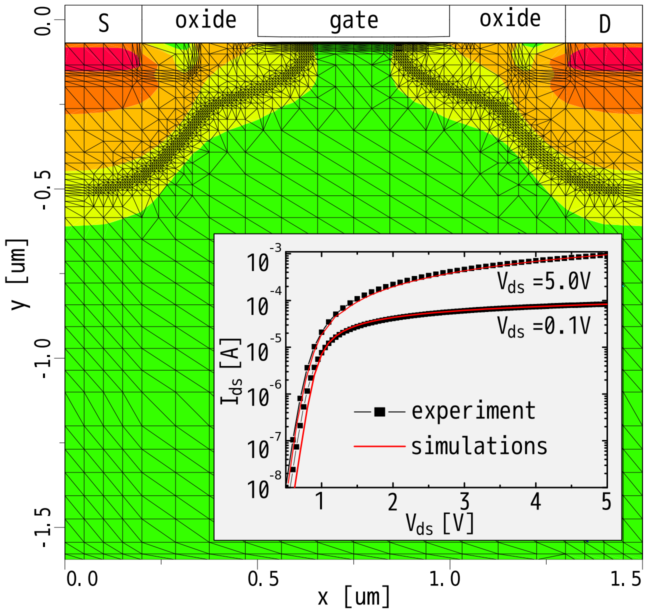
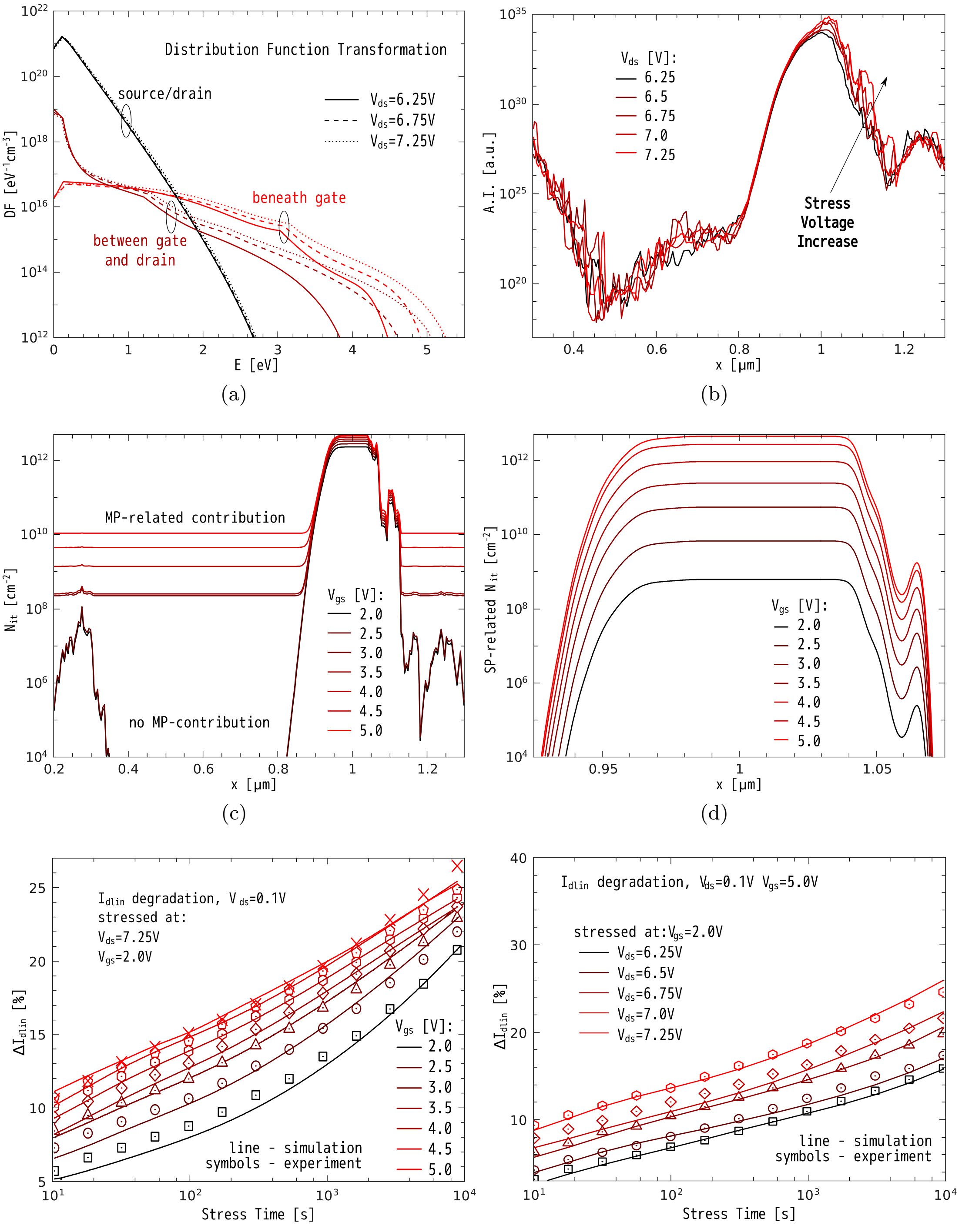


Figure 3.2: The topology of an n-MOSFET with the donor doping profile represented by the color map. The origin of the lateral coordinate corresponds to the beginning of the source. Inset: experimental I_{ds}-V_{ds} curves vs. simulated ones.

induced damage is controlled by the carrier AI which is defined by the shape of the DF. The family of the effective $N_{\mathrm{it}}(x)$ profiles calculated over various operating conditions at a fixed stress time t of 10s is shown in Figure 3.3c and Figure 3.3d. One can see that MP-induced defects come into play only for $V_{\mathrm{gs}} \geq 3.0$V. This circumstance means that the SP- and MP-related states are differently distributed over energy with the latter shifted to higher energies. This result agrees with the concept by Hess *et al.* where the double-power law dependence of the degradation was explained by introducing two time slopes for defects created by different processes [24, 40].

Figure 3.3d resolves the density of particles captured by the SP-traps in the region where the total trap concentration N_{it} is plotted for different V_{gs}. Since the main contribution to the total density Q_{it} is provided by the MP-process one may compare the behavior of densities related to the different types of traps. For the SP-process the distance between the curves saturates, meaning that interface states of this type are almost fully occupied. In contrast, for the MP-process the increase of charge density continues.

The model calibrated in the aforementioned manner allows for the I_{dlin} degradation to be represented at various V_{gs} (Figure 3.3e) and for different stress conditions, i.e. different V_{ds} (Figure 3.3f). There is no additional fitting parameter introduced into the model, which means that N_{it} effectively changes while switching from one stress condition to another. Finally, using this approach we are now able to represent the transfer characteristics of the degraded device at each time step, see Figure 3.4a.

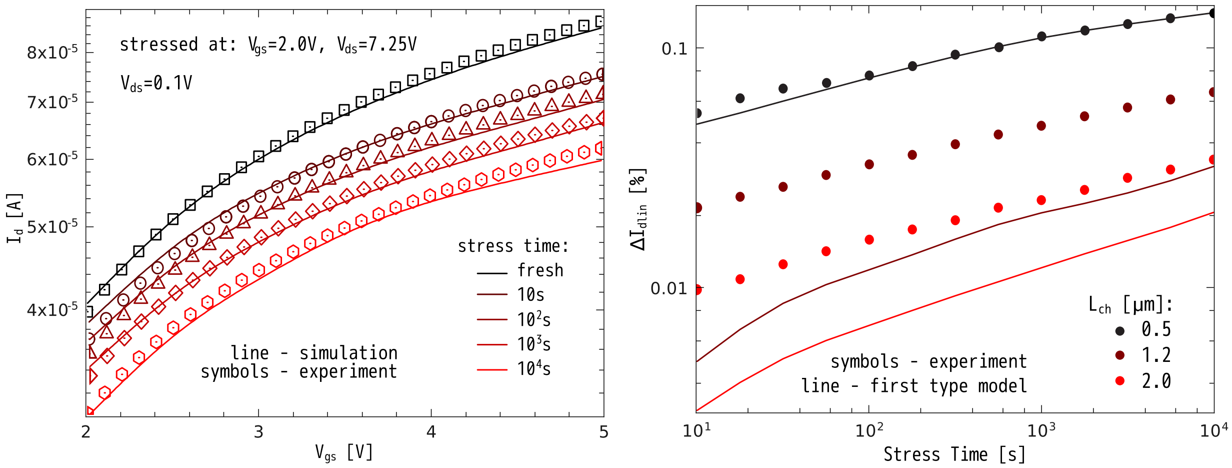


(a) (b)

(c) (d)

(e) (f)

Figure 3.3: Evolution of crucial characteristics of the degradation with the lateral coordinate: (a) carrier distribution function along the interface; (b) the carrier acceleration integral featuring a peak near the position of most prolonged high-energy tails of the distribution function; (c) the total interface charge density N_{it} and (d) stored on the SP-related traps N_{MP} in the region where the AI peaks. I_{dlin} degradation for (e) different operation V_{gs} and fixed stress conditions $V_{\mathrm{gs}} = 2.0\mathrm{V}$, $V_{\mathrm{ds}} = 7.25\mathrm{V}$ and for (f) different stress V_{ds} and fixed operation $V_{\mathrm{ds}} = 0.1\mathrm{V}$, $V_{\mathrm{gs}} = 5.0\mathrm{V}$.



(a) (b)

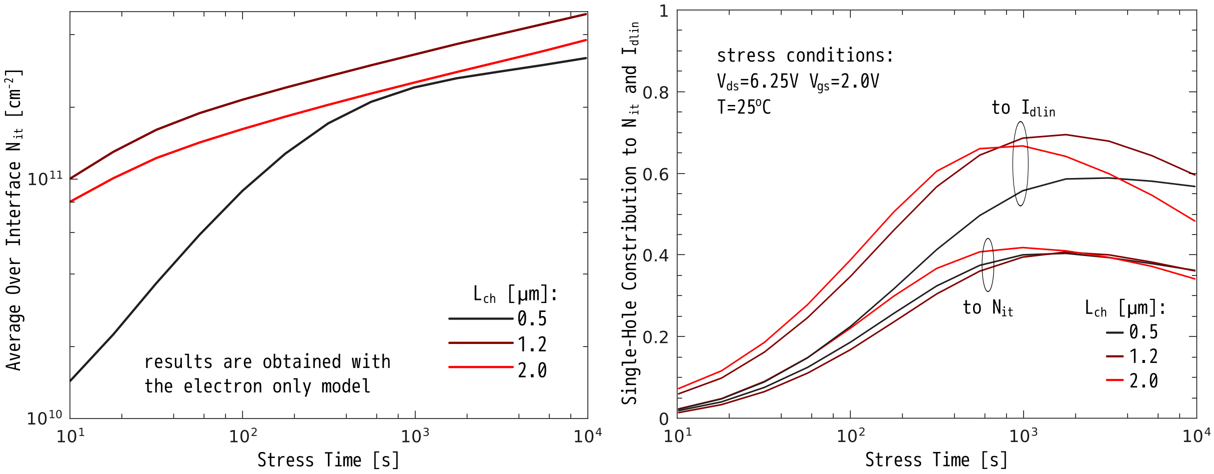
Figure 3.4: (a) The transformation of the transfer characteristics during stress: experiment (symbols) vs. theory (lines). (b) The degradation of the I_{dlin} current predicted by the electron-only HCD model for different channel lengths.

3.2 Secondary Generated Carriers as a Crucial Component for Modeling of HCD

The approach described in Section 3.1 is able to successfully represent the linear drain current degradation for a wide range of stress/operating conditions in the case of a 5V n-MOSFET with a channel length of $L_{\mathrm{ch}} = 0.5\mu\mathrm{m}$ (Figure 3.2). Unfortunately, the model completely fails to capture I_{dlin} degradation in MOSFETs with different channel lengths with the same set of fitting parameters. In fact, for devices with $L_{\mathrm{ch}} = 1.2$ and $2.0\mu\mathrm{m}$, the relative change of I_{dlin} (i.e. the ratio between its absolute change and the value typical for a “fresh” device) predicted by the model is substantially less than that observed experimentally, see Figure 3.4b.

The reason is that the concentration of interface states generated by electrons peaks outside the channel, thereby weakly affecting the device performance. Figure 3.5a, showing the average N_{it} (i.e. integrated over the interface and divided by the interface length), demonstrates that more severe degradation corresponds to a lower average concentration of interface states. Hence, longer devices are less sensitive to electron-induced N_{it}, suggesting that another mechanism leading to N_{it} created closer to the channel must be responsible for this discrepancy. One may envisage that the missing contribution to the damage is triggered by secondary generated (by impact ionization) holes which are accelerated by the electric field and thus create interface states shifted towards the source. Therefore, it is necessary to extend the presented version of the model (described in Section 3.1) in order to make it suitable to represent the degradation in transistors with different channel lengths within the same set of parameters.

As it is proposed in Section 3.1 the superposition of single- and multiple-carrier mechanisms of $\mathrm{Si-H}$ bond-breakage is considered. Both mechanisms are controlled by the carrier acceleration integral



(a) (b)

Figure 3.5: (a) The average (over the interface length) total degradation dose for different channel lengths. (b) The relative contribution provided by the single-hole component into the total I_{dlin} change and to the total N_{it}.

$$I_{\text{SP/MP,e/h}} = \int_{E_{\text{th}}}^{\infty} f(E)g(E)\sigma_{\text{SP/MP,e/h}}(E)v(E)\text{d}E. \tag{3.14}$$

Note that we use the acronyms SP-/MP- to distinguish between the two bond-breakage processes and ‘e/h’ to distinguish the carrier types. The secondary holes are generated by impact ionization caused by the injection of hot electrons. These holes are then accelerated by the electric field towards the source, thereby creating interface states shifted with respect to the electron-induced ones. Also, the N_{it} fraction induced by holes should be much less than their relative contribution to I_{dlin} change, Figure 3.5b.

For the SP-process first-order kinetics to describe the SP-induced contribution to the total N_{it} is assumed

$$N_{\text{SP}}(t) = n_0\left(1 - \exp[-(\nu_{\text{SP,e}}I_{\text{SP,e}} + \nu_{\text{SP,h}}I_{\text{SP,h}})t]\right), \tag{3.15}$$

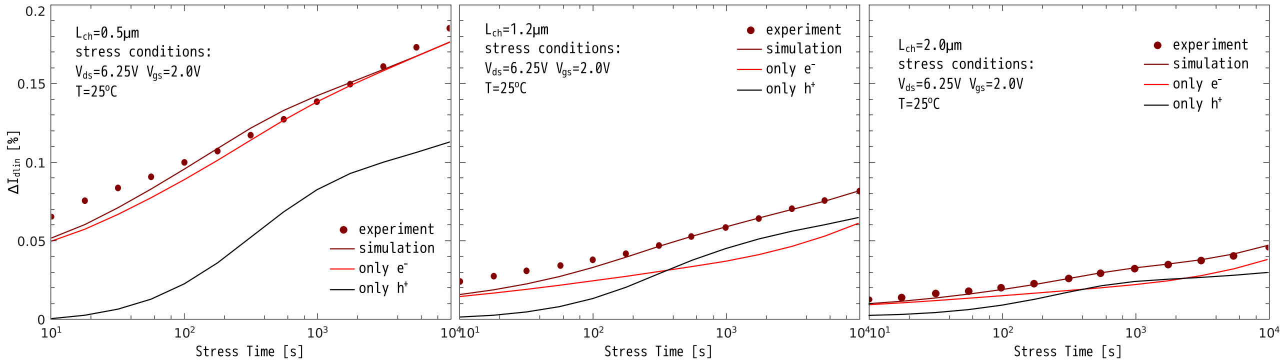
where $\nu_{\text{SP,e/h}}$ are prefactors related to electron/hole contributions.

The MP-induced portion of N_{it} is calculated analogously to (3.9) Following the procedure proposed in Section 3.1, one may obtain the new expressions for phonon excitation and decay rates with corresponding corrections for secondary carrier type

$$P_{\text{u}} = I_{\text{MP,e}} + I_{\text{MP,h}} + \omega_{\text{e}} \exp\left(-\frac{\hbar\omega}{\text{k}_{\text{B}}T_{\text{L}}}\right), \tag{3.16}$$

$$P_{\text{d}} = I_{\text{MP,e}} + I_{\text{MP,h}} + \omega_{\text{e}}. \tag{3.17}$$

The bond-breakage process is associated with the hydrogen hopping from bonded to transport state (with the rate λ_{emi}) with the backward reaction also taking place (P_{pass}). For the MP-process, as it was already shown in Section 3.1, even in the case of a high-voltage device the



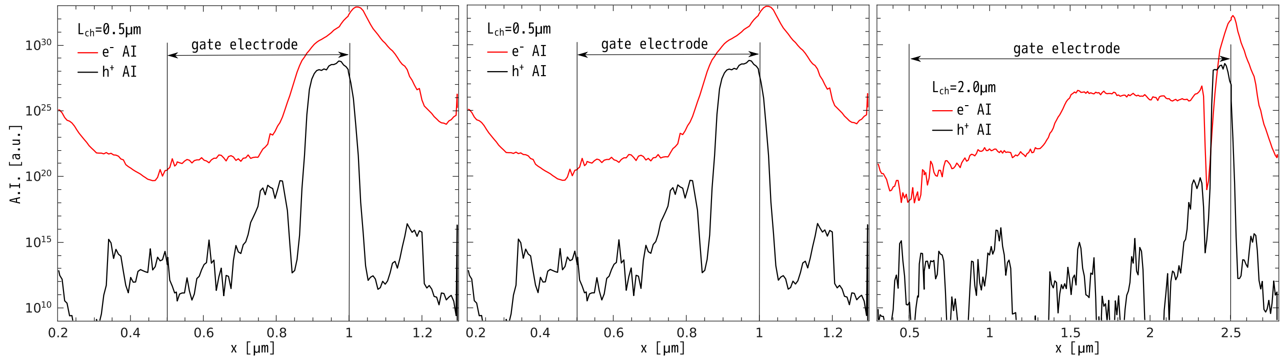
(a) (b) (c)

Figure 3.6: The relative I_{dlin} change vs. time: experiment, simulations and contributions of electrons and holes separately for channel lengths of (a) 0.5, (b) 1.2 and (c) 2.0μm.

MP-component makes a significant contribution to the total damage. However, in this case the carrier flux is still high enough, i.e. $I_{\mathrm{MP,e}} + I_{\mathrm{MP,h}} \gg \omega_{\mathrm{e}}$. As a result, the prefactor $P_{\mathrm{u}}/P_{\mathrm{d}}$ in (3.9) is unity and N_{MP} is homogeneously distributed along the lateral coordinate x as in Section 3.1.

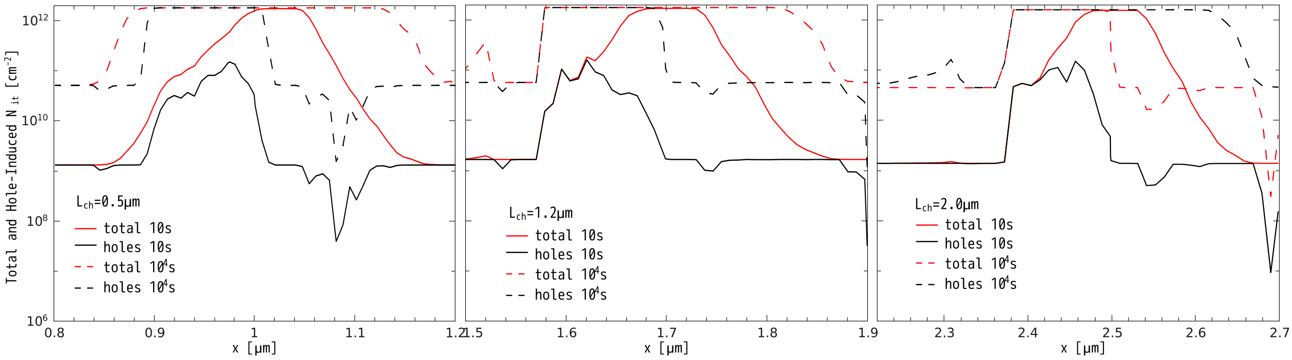
The verification of the extended model is performed using a series of three 5V n-MOSFETs (see Figure 3.2) with identical architecture differing only in channel lengths ($L_{\mathrm{ch}} = 0.5$, 1.2, and 2.0μm). Devices were fabricated on a standard 0.35μm technology and subjected to a hot-carrier stress at the gate voltage $V_{\mathrm{gs}} = 2.0$ and the drain voltage of $V_{\mathrm{ds}} = 6.25$V at 25°C. The model was calibrated in a manner which represents the I_{dlin} degradation for all devices with a single set of fitting parameters and reveals a good agreement between experimental and theoretical results, shown in Figure 3.6. AIs for electrons and holes plotted vs. x are depicted in Figure 3.7.

In Figure 3.8, which demonstrates the total N_{it} profile and the hole-related component of N_{it} calculated for 10s and 10^4s, one may explicitly see that the hole contribution is considerably shifted towards the source. The single-electron component generates traps situated outside the channel which explains why the hole-induced traps have stronger relative impact on I_{dlin}. Only the SP-component was considered while analyzing the relative contribution of holes. For both types of carriers the AI is already too high, thereby saturating the MP rate for this process. As a result, it is impossible to distinguish between multiple-electron and multiple-hole contributions. Additionally, N_{MP} is homogeneously distributed over x and speculations about the position-



(a) (b) (c)

Figure 3.7: The acceleration integrals for electrons and holes for the case of $L_{\mathrm{ch}} = 0.5$, 1.2 and 2.0 μm.



(a) (b) (c)

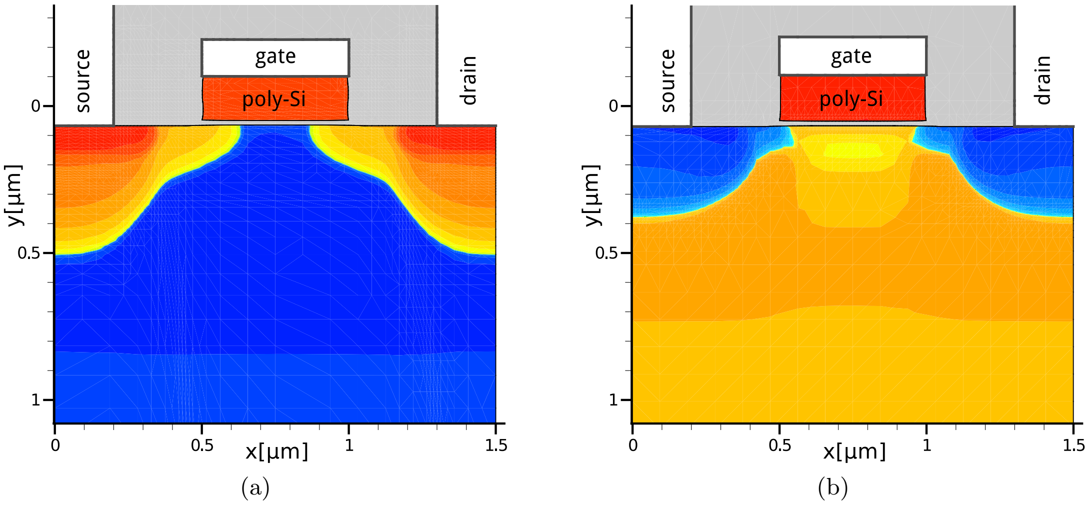
Figure 3.8: The total N_{it} profile and that induced only by holes for 10s and 10^4s and for three different channel lengths.

dependent impact of N_{it} on the device performance are possible only for the SP-component. The contribution of hot holes to the total concentration N_{it} (Figure 3.5b) is much less than the corresponding fraction of I_{dlin} change. This trend becomes more pronounced for longer devices. Furthermore, such a behavior is also supported by Figure 3.6 where the experimental I_{dlin} degradation is plotted against the theoretical one as well as the portion of I_{dlin} change induced by electrons and holes only. Note that in the case of $L_{\mathrm{ch}} = 0.5\mu$m the degradation may be represented employing only the electron-component, but this is not possible for longer devices.

To summarize, the extended model is verified by representing HCD in n-type MOS transistors with various channel lengths using a single number of fitting parameters. The model considers not only channel electrons but also secondary holes generated by impact ionization. It is shown that in spite of a less pronounced hole contribution to the total interface states density, the device behavior is more sensitive to the hole-induced trap generation as compared to the electron one. This is related to the fact that hole-induced N_{it} is situated closer to the channel, thereby affecting the drain current in a stronger fashion.

3.3 Analysis of Worst-Case HCD Conditions in the Case of n- and p-channel MOSFETs

Although the experimental worst-case conditions (WCC) was determined some time ago, a detailed numerical analysis of the matter is still missing. This is related to the lack of a comprehensive model for hot-carrier degradation. At the same time a predictive diagram reflecting how detrimental these given stress/operating conditions are for the device performance would be of great significance for reliability engineers. Furthermore, representations of such experimental diagrams by an HCD model should serve as a criterion of the model consistency. Therefore, the developed physics-based model for hot-carrier degradation is applied in order to represent the experimental color maps depicting the device damage as a function of $\{V_{\mathrm{ds}},V_{\mathrm{gs}}\}$ in the case of long-channel and high-voltage n- and p-MOSFETs.



(a) (b)

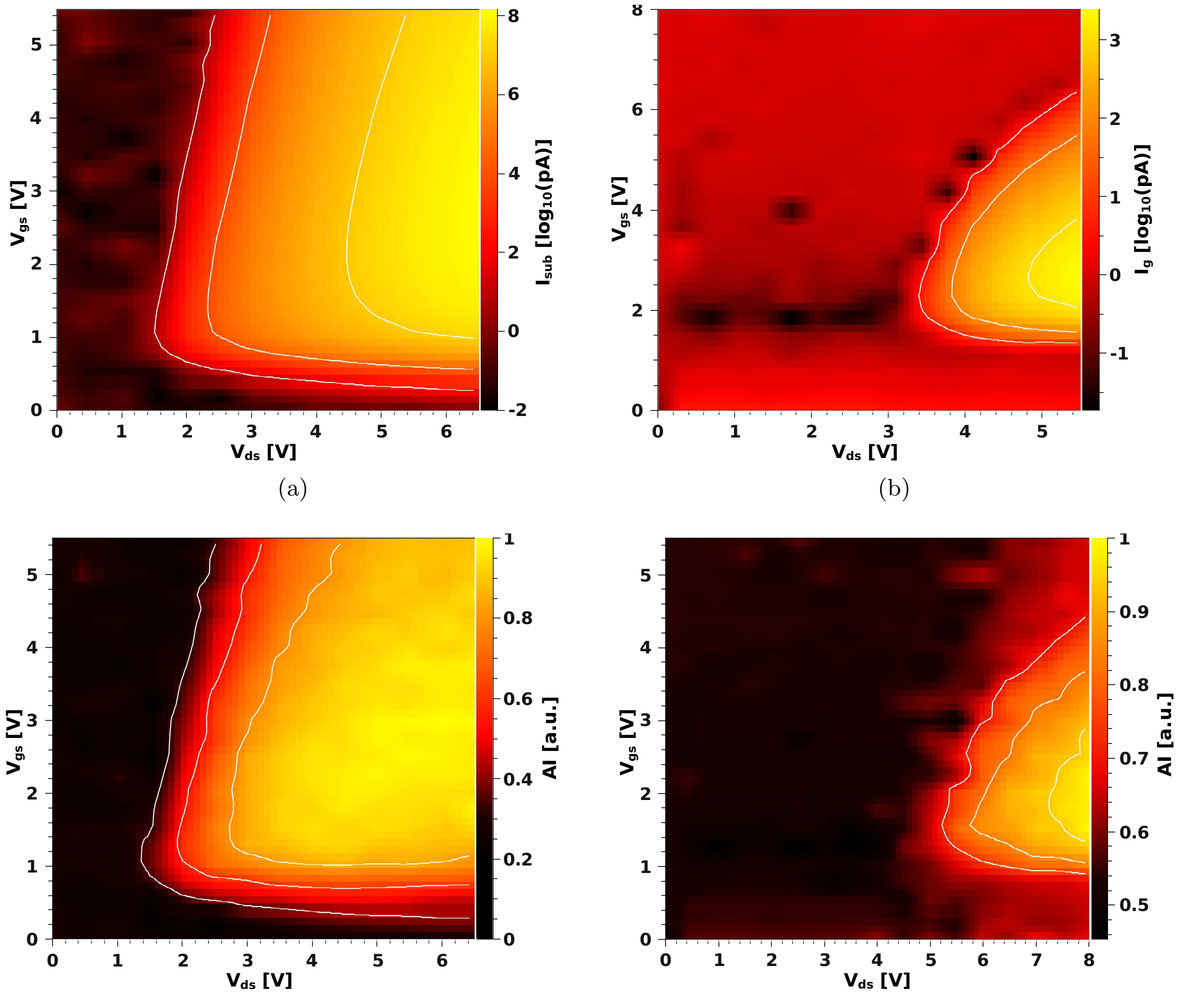
Figure 3.9: The topology of 5V (a) n-MOSFET and (b) p-MOSFET with the net doping profile highlighted.

3.3.1 WCC of Long-Channel Devices

The WCC of hot-carrier degradation in the case of a long-channel n-MOSFET are reached when the substrate current I_{sub} is at its maximum [29, 41, 70, 167]. It is usually assumed that this maximum corresponds to $V_{\mathrm{gs}} \cong (0.4-0.5)V_{\mathrm{ds}}$. As for long-channel p-MOSFET, WCC are observed at the maximum gate current I_{g} and an empirical link between the voltages is not established [50–52].

Long-channel n- and p-MOSFETs with a metallurgical gate contact length of 0.5μm (the device architectures and the net doping profiles are sketched in Figure 3.9) have been employed for monitoring the criteria reflecting the severity of HCD. All lateral coordinates x refer to an origin placed at the left edge of the source contact for both devices. Together with a relatively high operation voltage this channel length ensures that the single-electron mechanism of the $\mathrm{Si-H}$ bond dissociation is the dominating one [41]. In the case of n-MOSFETs, the substrate current as a function of varying $\{V_{\mathrm{ds}}, V_{\mathrm{gs}}\}$ was recorded and binned onto a color map (Figure 3.10a). For p-MOSFET $I_{\mathrm{g}} = I_{\mathrm{g}}(V_{\mathrm{ds}}, V_{\mathrm{gs}})$ was measured and depicted in Figure 3.10b.

Note that in a previous Section 3.2 the damage produced by channel holes is also considered. However, in Section 3.1 it is shown that in the case of the channel length of 0.5μm the linear drain current degradation is sufficiently described by the electron contribution only. Furthermore, as discussed in Section 3.1 the dose of damage provided by the MP-component is homogeneously distributed over the lateral coordinate. This is because the prefactor in the MP-induced interface state concentration defined by the acceleration integral is already saturated due to the high concentration of relatively "cold" carriers (represented by the low-energy fragment of the DF). As a result, the non-uniform nature of HCD is related to the single-carrier bond-breakage process with the corresponding contribution to the total interface state density localized near the drain end of the gate. Therefore, in order to find the stress conditions corresponding to the worst-case scenario, it is suggested to analyze the behavior of the acceleration integral, without considering the interaction between the SP- and MP-mechanisms. In literature, one may find other criteria



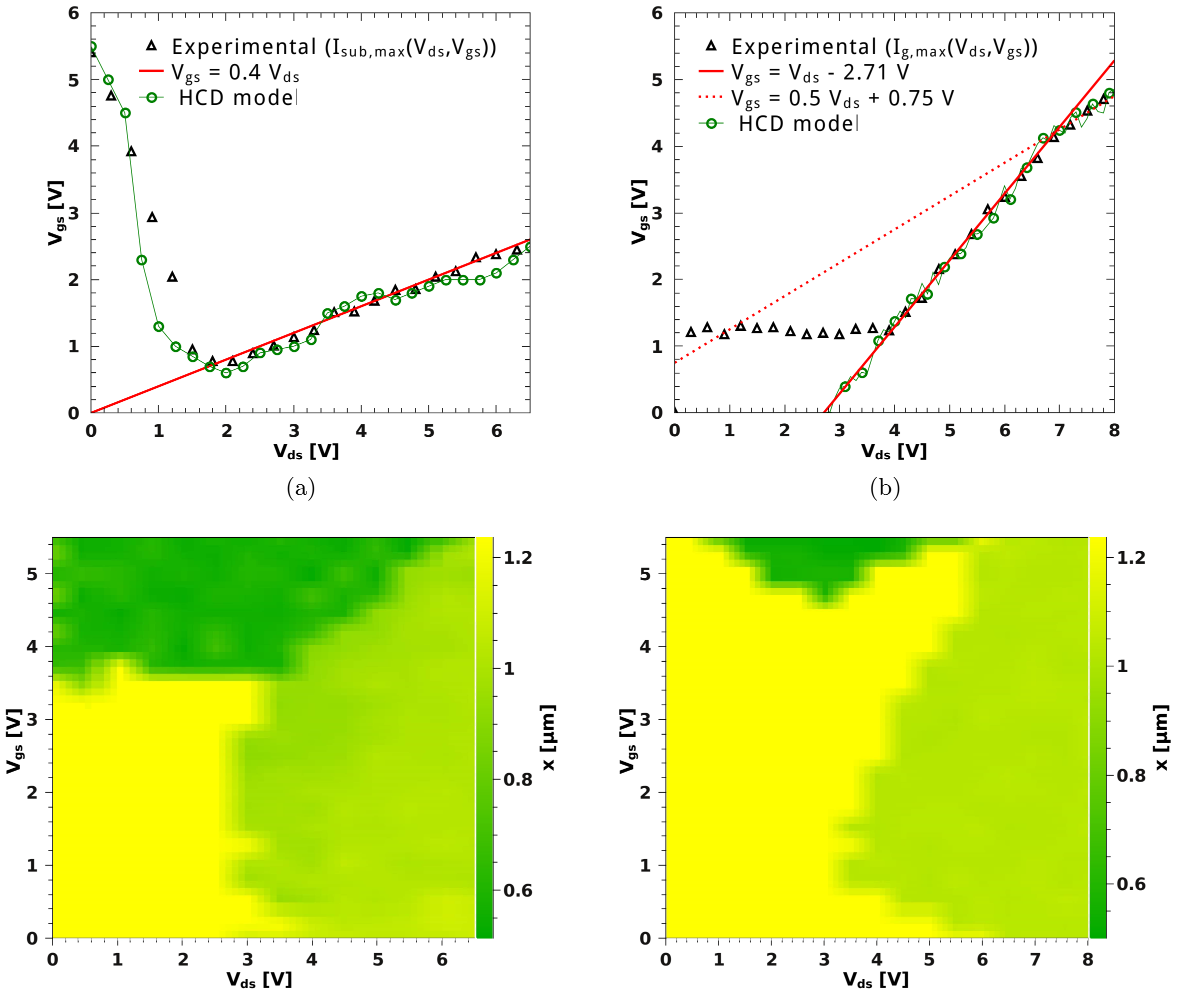
(a) (b)

(c) (d)

Figure 3.10: Experimental (a) I_{sub} and (b) I_{g} as a function of V_{gs} and V_{ds} for the n- and p-MOSFET, respectively. Maximum value of the acceleration integral as a function of V_{gs} and V_{ds} for (c) n- and (d) p-MOSFET.

how efficiently carriers interact with the bond, i.e. the maximum of the electric field, the carrier dynamic temperature, position where the DF demonstrates most extended high-energy tales, etc. However, as it will be additionally confirmed in the next Chapter 4, the N_{it} peak only corresponds to the maximum of the AI and is shifted with respect to maxima of other quantities.

In order to numerically calculate picture for the worst-case conditions, the following computational procedure is used. The carrier acceleration integral is used as a criterion to judge how severe the HCD is. In other words, the AI as a function of $\{V_{\text{ds}}, V_{\text{gs}}\}$ is calculated and the color maps (see Figure 3.10c and Figure 3.10d) is produced in the same fashion as Figure 3.10a and Figure 3.10b, respectively. The computational procedure presumes the following steps. Initially, V_{gs} and V_{ds} are varied in the range of 0..6V and 0..7V (for p-MOSFET the V_{ds} range is of 0..8V) with 0.2V step. For each pair $\{V_{\text{gs}}, V_{\text{ds}}\}$ the set of carrier DFs (i.e. at each point along the Si/SiO$_2$ interface) is calculated. This information is used to calculate the dependences of the carrier acceleration integral vs. the lateral coordinate. After processing these dependencies the



(a)

(b)

(c)

(d)

Figure 3.11: The interrelation between V_{gs} and V_{ds} corresponding to the WCC of (a) n- and (b) p-MOSFET. The position of the maximum acceleration integral as a function of V_{gs} and V_{ds} for (c) n- and (d) p-MOSFET.

maximum value of the acceleration integral I_{SP} is plotted as a function of $\{V_{\mathrm{ds}},V_{\mathrm{gs}}\}$. Therefore, the entire computational routine contains approximately 1000 different combinations of stress voltages while each pair $\{V_{\mathrm{gs}}, V_{\mathrm{ds}}\}$ includes time expensive Monte-Carlo simulations of the carrier DF.

The comparison between Figure 3.10a (Figure 3.10b) and Figure 3.10c (Figure 3.10d) demonstrates a good agreement between experimental and simulated color maps for both types of transistors. The dependence of the gate voltage on the drain bias corresponding to the worst-case conditions is also plotted, see Figure 3.11a and Figure 3.11b. This data is extracted from Figure 3.10a and Figure 3.10b where V_{gs} is presented as a function of V_{ds} which is found as the value guaranteeing the maximum of either experimental data (I_{sub} for n-MOSFET and I_{g} for p-MOSFET) or I_{SP} (simulations with HCD model) for a chosen value of V_{ds}. For a drain voltage above ~ 2.0V for the n-MOSFET this dependence follows the law $V_{\mathrm{gs}} = 0.4V_{\mathrm{ds}}$ typical for this device. Although for the p-MOSFET this dependence is more complicated (Figure 3.11b), the experimental and simulated pictures are in good agreement. Finally, the position where the maximum AI is observed is plotted vs. $\{V_{\mathrm{gs}}, V_{\mathrm{ds}}\}$, see Figure 3.11c and Figure 3.11d. Since the peak of the acceleration integral determines the location of the most severe degradation spot, this graph can be used to predict where the most degraded section of the device is located. Note that the color maps presented in Figure 3.11c and Figure 3.11d reflect the pattern shown in Figure 3.10c and Figure 3.10d, respectively. For instance, the orange area pronounced in Figure 3.10c (and bordered by the right contour curve) corresponds to the light green spot in Figure 3.11c.

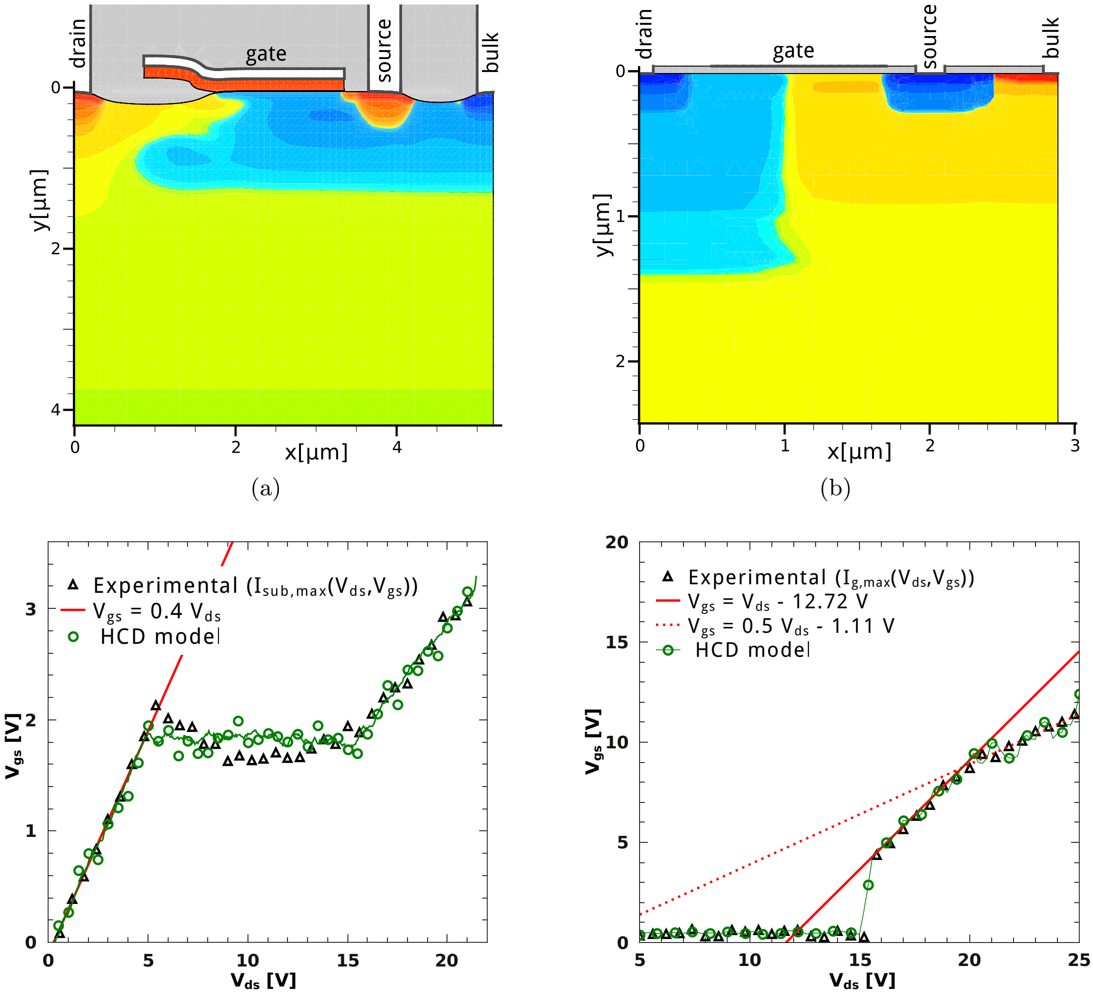
3.3.2 WCC of High-Voltage Devices

The developed approach for WCC analysis is applicable in the case of HV LDMOS transistors. A cross section of devices employed for the analysis of the model applicability for such a type of devices is shown in Figure 3.12. The first HV transistor is an n-type LDMOS integrated into an 180nm CMOS base process (Figure 3.12a). A 350nm thick shallow trench isolation (STI) oxide is used as a dielectric layer above the drift region of the high voltage device. The second device is a HV p-type LDMOS with a simplified geometry (flattened interface and absence of the STI, see Figure 3.12b). The operation voltage of these transistors is V_{ds} = 20V and V_{ds} = -20V, respectively. Following the same computational procedure as described above, the maximum value of the acceleration integral as a function of $\{V_{\mathrm{ds}}, V_{\mathrm{gs}}\}$ is calculated. The graphs obtained confirm the tendency that for the case of a n-type LDMOS the WCC are realized at the maximum I_{sub}. In contrast, for the p-type LDMOS, the severest HCD corresponds to the maximum gate current. This detailed picture agrees with the empirical concept concerning the HCD worst-case conditions documented in literature.

Finally, it is possible to conclude that the developed HCD model reproduces the WCC with a reasonable accuracy. The simulated diagrams representing the maximum AI as a function of gate and drain voltages agree well with the color maps $I_{\mathrm{sub}} = I_{\mathrm{sub}}(V_{\mathrm{ds}}, V_{\mathrm{gs}})$ for n-MOSFETs and $I_{\mathrm{g}} = I_{\mathrm{g}}(V_{\mathrm{ds}}, V_{\mathrm{gs}})$ for p-MOSFETs. Moreover, the suggested approach for HCD modeling predicts the position where the most degraded device section is localized (i.e. between the drain end of the gate and the drain in the case of long-channel devices).

3.4 Impact of the Carrier Distribution Function on HCD Modeling

As outlined in Section 1.5, the employed physical model for hot-carrier degradation includes three main sub-tasks: the carrier transport module, a module describing interface state generation and a module for the simulation of the degraded devices. The presented carrier transport module of the model is based on the solution of the BTE (Section 2.6) by means of a full-band MC device simulator MONJU [154]. However, Monte-Carlo simulations are time consuming and their substitution by faster simplified schemes – such as hydrodynamic (Section 2.2) and drift-diffusion (Section 2.3) models – for Boltzmann transport equation solving appears very attractive, see e.g. [168]. There is a considerable amount of papers (see e.g. [96,97,169]) devoted to the comparison between different transport schemes and showing that HD and DD models

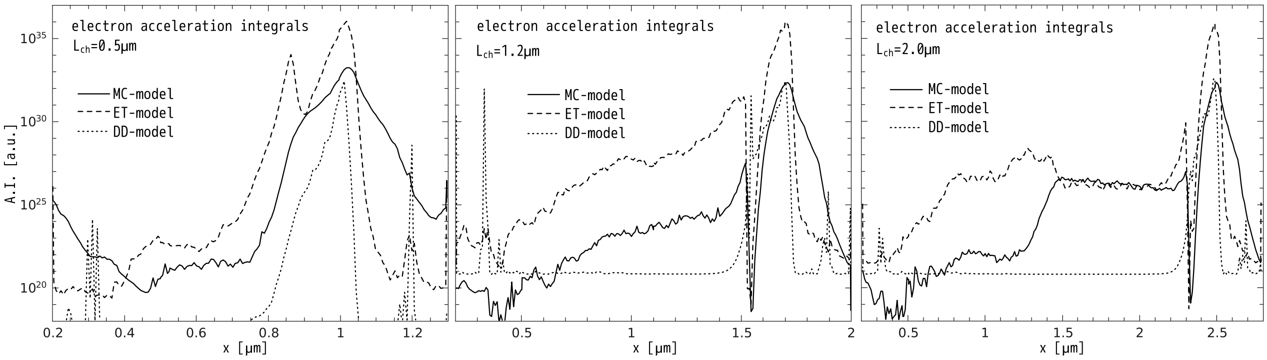


(a) (b) (c) (d)

Figure 3.12: The topology of HV (a) n- and (b) p-LDMOS with the net doping profile highlighted. The interrelation between V_{gs} and V_{ds} corresponding to the WCC of (c) n-MOSFET and (d) p-MOSFET.

fail to properly describe hot carriers in the case of a channel length less than 0.1μm. A detailed analysis of results obtained with MC, HD and DD approaches to the BTE solution was performed in [169] confirming that HD and DD schemes may be applicable for channel lengths longer than 0.1μm. The developed HCD model has been verified using MOSFETs with a channel length longer than 0.5μm, see Section 3.2. In these long-channel devices the SP-mechanism plays the dominant role. The SP-mechanism is triggered by the "hottest" carriers, i.e. is defined by the high-energy tail of the DF. As a consequence, the HCD model for this case will be the most sensitive to the chosen transport scheme. Therefore, it appears mandatory to investigate whether simplified treatments of carrier transport may be successfully employed in the model to substitute the time-consuming Monte-Carlo method. Thus, this Section aims to compare different (i.e. based on MC, HD and DD methods) realizations of the HCD model.

To verify these model versions, three 5V n-MOSFETs of the same architecture (Figure 3.2) were employed, differing only in channel lengths ($L_{\mathrm{ch}} = 0.5, 1.2$ and 2.0μm as in Section 3.2). Devices



(a) (b) (c)

Figure 3.13: The electron acceleration integral calculated with with Monte Carlo, Hydrodynamic, and Drift-Diffusion based versions of the HCD model for the case of (a) $L_{\rm ch} = 0.5\mu$m, (b) $L_{\rm ch} = 1.2\mu$m and (c) $L_{\rm ch} = 2.0\mu$m.

were stressed at a gate voltage $V_{\rm gs} = 2.0$ and a drain voltage of $V_{\rm ds} = 6.25$V at 25°C.

In the first approach, the DF is obtained by solving the BTE with the Monte-Carlo method. In the second version, only the average energy is taken from the MC solution in order to emulate the solution of an hydrodynamic transport model. The energy profile $\langle E\rangle(x)$ is then used to approximate the carrier DF as a function of position as it is normally done in energy transport based physical models described by

$$f(E,x) = A\ \exp\left(-\frac{E}{\langle E\rangle(x)}\right). \tag{3.18}$$

Finally, in the third variant, only the electric field from the Monte-Carlo solution is retained. Using this field profile one may obtain the average carrier energy as

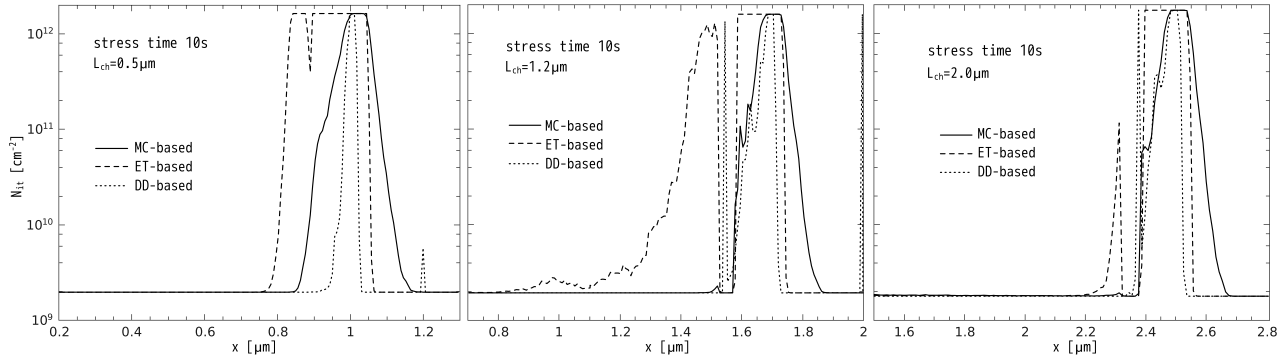
$$\langle E\rangle(x) = \frac{3\mathrm{k_B}T_\mathrm{L}}{2} + \mathrm{q}\tau_\mathrm{E}(x)\mu F(x)^2. \tag{3.19}$$

Note that in order to eliminate a possible origin of discrepancy related to different device simulators all the calculations were performed within MONJU.

3.4.1 Comparison of Different Transport Module Realizations

The electron AIs computed within the MC-, HD- and DD-based models are plotted in Figure 3.13 (note that all the findings are similar for the hole acceleration integral). The driving force of the degradation in the DD-based model is the electric field. However, the DF follows the electric field with a certain delay [93]. As a result, the maximum of the MC-based AI is shifted towards the drain with respect to the result of the drift-diffusion scheme. As demonstrated in Section 3.1 starting from source to drain first the maximum of the carrier average energy appears, followed by the electric field peak, and finally the maximum AI when calculated with the Monte-Carlo method. In Figure 3.13 the maxima of corresponding acceleration integrals are in this order.

The $N_{\rm it}$ profiles calculated employing different transport schemes also confirm this behavior,



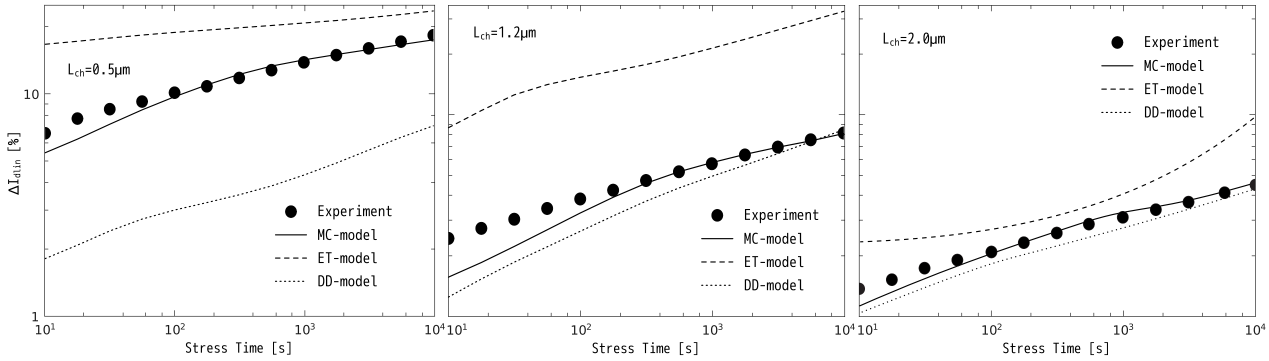
(a) (b) (c)

Figure 3.14: The simulated $N_{\text{it}}(x)$ profiles obtained with Monte Carlo, Hydrodynamic, and Drift-Diffusion based versions of the proposed HCD model for the case of (a) $L_{\text{ch}} = 0.5\mu$m, (b) $L_{\text{ch}} = 1.2\mu$m and (c) $L_{\text{ch}} = 2.0\mu$m.

see Figure 3.14. The interface state density evaluated with the HD-based model spuriously overestimates the damage as compared to DD and MC schemes. Such a trend was expected based on hot-carrier tunneling studies [170] where the tunneling process was also overestimated when DF was simulated employing the HD scheme. As a result, the linear drain current degradation predicted by the HD-based model is much stronger than those obtained employing the MC and DD approach (Figure 3.15). Finally, the DD-based model predicts ΔI_{dlin} close to the result obtained by the MC-based model for $L_{\text{ch}} = 1.2$ and 2.0μm but totally fails for $L_{\text{ch}} = 0.5\mu$m.

3.4.2 Analysis of WCC within Simplified Transport Schemes

For additional visualization of the obtained results in this Section, the worst-case conditions will be analyzed using simplified transport schemes. That is, the computational procedure presented in Section 3.3, which is based on the evaluation of the carrier distribution function along the Si/ SiO_2 interface is repeated for HD- and DD-based transport approaches. For these investigations only a long-channel 5V n-MOSFET (Figure 3.2) is employed. Moreover, the simplified treatment of carrier transport using a non-Maxwellian, analytical distribution function is also discussed.



(a) (b) (c)

Figure 3.15: The linear drain current degradation: experiment vs. simulations. The case of (a) $L_{\text{ch}} = 0.5\mu$m, (b) $L_{\text{ch}} = 1.2\mu$m and (c) $L_{\text{ch}} = 2.0\mu$m.

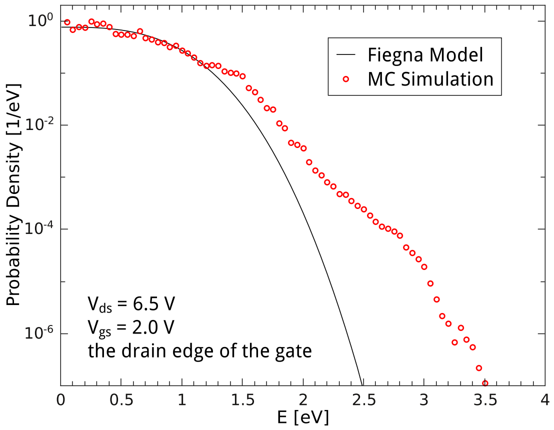


Figure 3.16: The carrier distribution function: the result of Monte-Carlo simulations and its fitting with the Fiegna model for a long-channel 5V n-type MOSFET.

The most popular approach for analytical modeling of non-Maxwellian distribution functions is the Fiegna model [96, 171]. This model is applicable for devices with a channel length longer than 0.5μm which is suitable for our case. Within this approach the following expression for the DF is proposed

$$f_{\mathrm{Fi}}(E,x) = C\ \exp\left(-\frac{\chi E^3}{F(x)^{1.5}}\right), \tag{3.20}$$

where C denotes the normalization constant and χ is a fitting parameter. This parameter may be adjusted in order to represent the real distribution function at a certain point. The choice of this point may substantially affect all results. The value of χ is found in a way which guarantees a minimal discrepancy between experimental and simulated charts $I_{\mathrm{sub}}(V_{\mathrm{ds}}, V_{\mathrm{gs}})$.

Explicitly, the value of χ used in this work is $1.62\times10^8\mathrm{eV}^{-3}\mathrm{V}^{1.5}\mathrm{cm}^{-1.5}$ (note that values reported in papers [171] and [96] are 1.3×10^8 and 0.8×10^8 $\mathrm{eV}^{-3}\mathrm{V}^{1.5}\mathrm{cm}^{-1.5}$, respectively). This parameter is found while fitting the distribution function at the drain end of the gate by the Fiegna expression, see Figure 3.16. The most significant discrepancy between these distribution functions is observed at higher energies. This circumstance suggests that the acceleration integrals calculated using these two different DFs should strongly diverge under conditions corresponding to the most extended high-energy tails of the distribution function.

The measured substrate current as a function of stress drain and gate voltages is depicted in Figure 3.10a. Figure 3.10c demonstrates the maximum value of the acceleration integral I_{SP} plotted vs. V_{ds} and V_{gs} and calculated within the HCD model which is based on the direct solution of the Boltzmann transport equation using the MC approach. Figure 3.17 demonstrates the same result as Figure 3.10c but the AI was calculated employing the fitting function $f_{\mathrm{Fi}}(E)$ and using HD/DD-based transport models. A comparison of these figures suggests several conclusions. In contrast to the Monte Carlo-based HCD model, which provides a very good agreement between the experiment and theory, the simplified versions which employs the Fiegna function instead of the real DF (as well as HD- and DD-based approaches) fail in the area of low V_{gs} with high V_{ds}. In practice this regime is characterized by a low I_{dlin} and thus degradation is weak (this trend is pronounced in Figure 3.10c). That is, if V_{ds} is fixed and V_{gs} varies, the substrate current first grows, reaches its maximum and then decreases. This tendency is also supported by

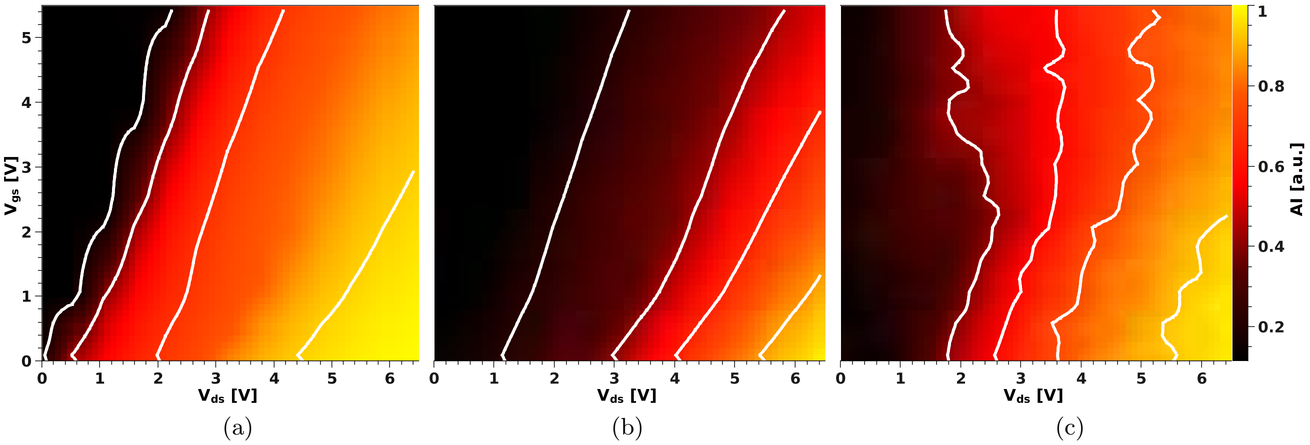


Figure 3.17: The maximum value of the acceleration integral as a function of $\{V_{\mathrm{ds}}, V_{\mathrm{gs}}\}$ calculated with the (a) Fiegna, (b) drift-diffusion and (c) hydrodynamic models.

Figure 3.11a. However, according to the predictions from Figure 3.17 the maximum degradation dose monotonously increases with V_{gs}, i.e. WCC are obtained at $V_{\mathrm{gs}} = 0\mathrm{V}$. Furthermore, in general, the color map generated employing the simplified models is not so obvious and it is difficult to judge relatively the WCC at all. In other words, one may finally conclude that simplified transport schemes cannot be used for proper HCD modeling.

4 The Charge-Pumping Technique

The developed physical model of hot-carrier induced degradation is based on a representation of the interface state profile of degraded devices. Therefore, a direct comparison of the simulation results with observed ones, i.e. with N_{it} profiles obtained experimentally is required for model verification. Moreover, it is very important to characterize the device interface as precisely as possible. The CP technique is one of the most useful technique for measuring the density of interface and oxide traps in MOS transistors. It has been proven as a reliable probe for Si/ SiO_2 interface states, such as the total number of traps, the energy and spatial distributions, and the variation of these quantities as a function of hot-carrier-injection stress. Additionally, only basic equipment is required and the set up is trivial. The technique has been a subject of many theoretical and simulation studies. These studies have resulted in the development of numerous variations of the basic technique in order to obtain more detailed information regarding the energy and spatial distribution of interface traps generated during hot-carrier stress experiments. The charge pumping method was originally proposed by Brugler and Jespers in 1969. 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 to the leakage current of the source and drain to substrate diodes. They showed that the current originates from a recombination of minority and majority carriers at traps at the Si/SiO_2 interface. Therefore, the method can be used to measure the interface trap density in MOSFETs for the evaluation of device degradation. The major breakthrough for the CP method was the thorough investigation and correct explanation of the method, applied directly to MOSFET structures by Groeseneken *et al.* in 1984 [172].

4.1 Basic Principles of Charge Pumping Measurements

The basic charge pumping experiment setup and working principles will be described in this Section. The experimental setup performance is shown in Figure 4.1. The MOSFET source and drain are tied together and slightly reverse biased. The time varying gate voltage is of sufficient amplitude for the surface under the gate to be driven into inversion and accumulation. The pulse train can be square, triangular, trapezoidal, sinusoidal, or tri-level. The charge pumping current is measured using an ammeter at the substrate, at the source/drain connection, or at the source and drain separately.

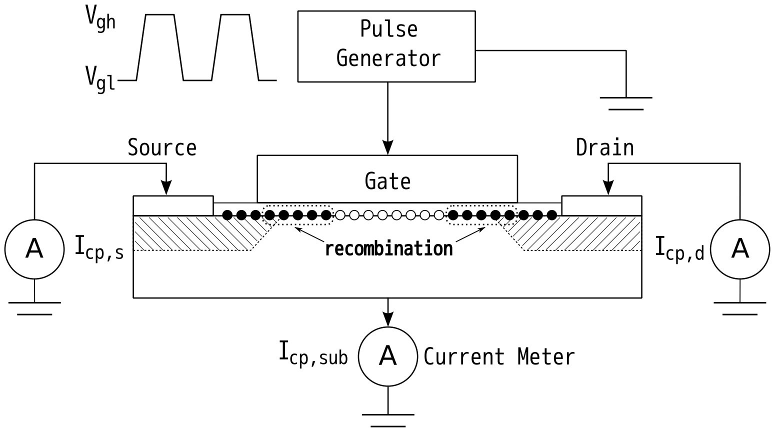
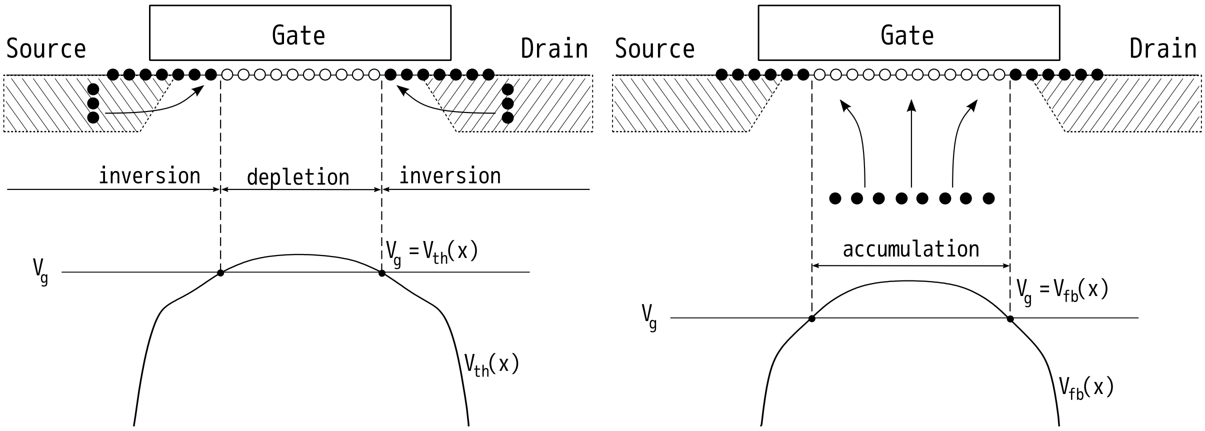


Figure 4.1: Basic experimental setup for charge-pumping measurements.

4.1.1 Local Threshold and Flatband Voltages

Foremost, it is necessary to introduce the notion of local threshold and local flatband voltages [173]. These quantities describe the dependencies of CP characteristics on the gate pulse amplitude V_{A}, the frequency f, the reverse bias on source and drain, etc. Moreover, the mentioned values are the key parameters for the determination of lateral defect distributions. The local threshold voltage $V_{\mathrm{th}}(x)$ can be defined as the gate voltage required to induce a certain density of minority carriers (electrons in the case of n-type MOSFET) and the local flatband voltage $V_{\mathrm{fb}}(x)$ is the gate voltage required to induce a desired density of majority carriers (holes) at each position x along the Si/SiO$_2$ interface. The values of these densities should be defined such that the interface states can capture holes and electrons during the inversion and accumulation periods of the gate pulse [173]. In other words, $V_{\mathrm{th}}(x)$ and $V_{\mathrm{fb}}(x)$ depend on the dynamics of the capture and emission process at the interface. The distribution of $V_{\mathrm{th}}(x)$ and $V_{\mathrm{fb}}(x)$ can be calculated with MiniMOS-NT using a widely-adopted routine [22,173–176]. Namely, as it was explaned above, the local threshold (flatband) voltage of a point at the interface of an n-type



(a) (b)

Figure 4.2: Definition of (a) local threshold voltage $V_{\mathrm{th}}(x)$ and (b) local flatband voltage $V_{\mathrm{fb}}(x)$ for nMOSFET. The positive direction of voltage points downwards.

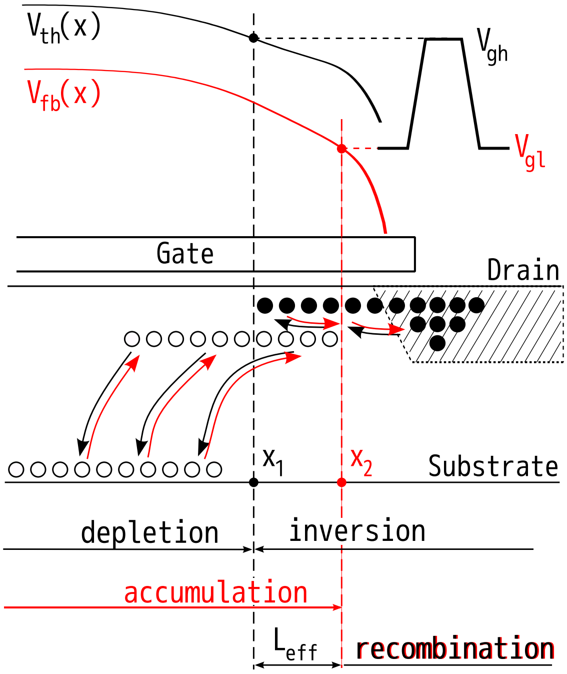


Figure 4.3: The response of the free carriers to the applied periodic gate pulse. At $V_\mathrm{g} = V_\mathrm{gh}$ electrons flood into x_1 and holes return to the substrate. At $V_\mathrm{g} = V_\mathrm{gl}$ electrons return to the drain and holes flood into x_2.

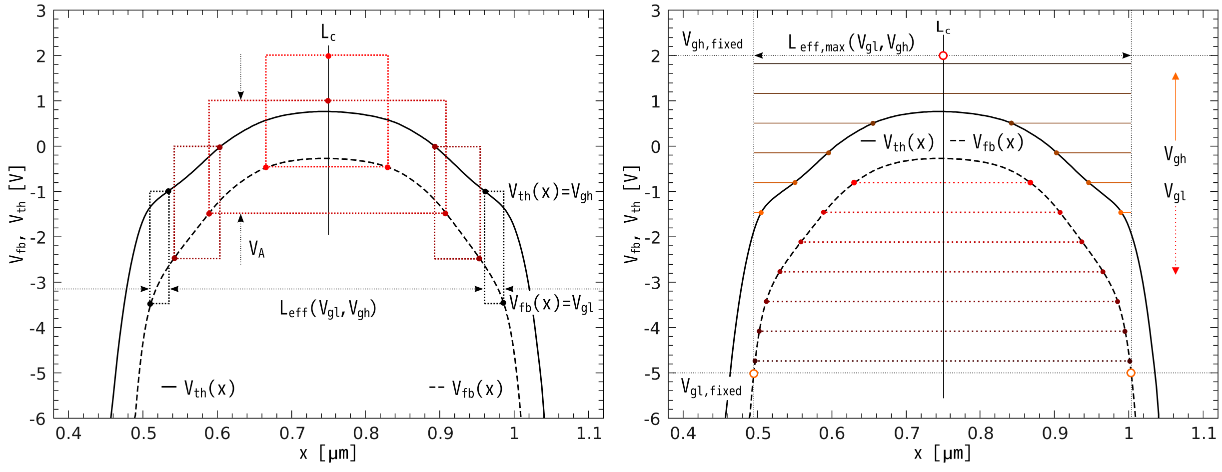
MOSFET is defined as the gate voltage required to accumulate the electron (hole) concentration $n_\mathrm{e/h}$ [173]

$$n_\mathrm{e/h} = \frac{1}{v_\mathrm{th}\sigma_\mathrm{e/h}\tau_\mathrm{e/h}}, \tag{4.1}$$

where

v_th	- thermal carrier velocity ($= 1 \times 10^7$ cm/s),
$\tau_\mathrm{e/h}$	- time constant for electron (hole) trapping ($= 1/(2f)$),
$\sigma_\mathrm{e/h}$	- capture cross section for electrons (holes).

Consequently, the dependence of the local threshold and flatband voltages on the coordinate along the interface x is determined by the oxide thickness and the amount of charges in the oxide and/or at the device interface. Additionally, these voltages significantly change with the position around the drain/source junctions as a result of variations in impurity concentration and two-dimensional effects in the potential. As shown in Figure 4.2a, an interface of an n-type MOSFET with a local threshold lower than V_g is flooded with electrons supplied from the source or drain. Similarly, as shown in Figure 4.2b, an interface whose local flatband voltage is higher than V_g is filled with holes from the substrate. Furthermore, it is necessary to consider the dynamic response of free carriers around the drain contact when periodic pulses of high level V_gh and low level V_gl are applied to the gate electrode. When V_g equals V_gh, the interface between the drain and x_1 is flooded with electrons from the drain (see, Figure 4.3). Simultaneously, the holes are rapidly swept into the substrate. As a consequence, interface states in this region are emptied of holes. However, when V_g falls to V_gl, free electrons go back to the drain and holes flood back from the substrate again and reach x_2. This tendency is reflected in Figure 4.3. Therefore, the interface traps up to x_2 are filled with holes.



(a)

(b)

Figure 4.4: Definition of the effective channel length using the notation of local threshold voltage $V_{\mathrm{th}}(x)$ and local flatband voltage $V_{\mathrm{fb}}(x)$ for n-type MOSFET in the case of (a) a constant and (b) a varying amplitude CP technique.

4.1.2 Effective Channel Length

It should be noted that the area flooded with holes and the area flooded with electrons overlap in the L_{eff} region, between x_2 and x_1

$$L_{\mathrm{eff}}(V_{\mathrm{gl}}, V_{\mathrm{gh}}) \equiv x_2(V_{\mathrm{gl}} = V_{\mathrm{fb}}(x)) - x_1(V_{\mathrm{gh}} = V_{\mathrm{th}}(x)). \tag{4.2}$$

Therefore, the interface states in this overlapping region L_{eff} are able to capture holes and electrons alternately. Through a repetition of the processes described above, recombination proceeds in L_{eff}. In contrast, outside of the L_{eff} region, the supply of either electrons or holes is insufficient for recombination to take place. Because the recombination process at position x begins almost abruptly when V_{gh} rises above $V_{\mathrm{th}}(x)$ or V_{gl} falls below $V_{\mathrm{fb}}(x)$ [173], the region in which recombination does proceed can be clearly distinguished from the region in which no recombination takes place in terms of the local threshold and flatband voltages. These findings lead to the conclusion that

$$I_{\mathrm{cp}}(V_{\mathrm{gl}}, V_{\mathrm{gh}}) = \mathrm{q}Wf \int_{L_{\mathrm{eff}}} N_{\mathrm{it}}(x)\mathrm{d}x. \tag{4.3}$$

This argument clarifies the geometrical characteristics which determine the value of the recombination currents I_{cp}. It should be noted that an overlapping region L_{eff} is also formed around the source as can be seen from an argument parallel to the above. In Figure 4.4 on increase/decrease of the effective channel length as a function of the applied gate pulse in the case of constant and varying amplitude CP technique is presented.

4.2 Interface State Density Profile of Unstressed Device

The density of interface traps along the Si/SiO$_2$ interface of an MOSFET is significantly influenced by the fabrication process and the device geometry. The requirement to consider a number of factors with a variety of physical properties results in the absence of an established model describing initial transistor defect profile at this time. Unfortunately, theoretical calculations of the defect concentration is a very complicated problem and no known solution exists. One of the most widely used experimental methods to obtain this information is the CP technique [173]. In papers devoted to extraction algorithms of the lateral interface state density $N_{\mathrm{it}}(x)$ profile from the CP current, different assumptions regarding the initial interface trap concentration (of a fresh device) have been used. Only a few papers verify the initial profile uniformity [164, 175, 177]. At the same time, most techniques for the $N_{\mathrm{it}}(x)$ extraction from CP data assume that the initial $N_{\mathrm{it}}(x)$ is homogeneous [176, 178]. This assumption is only a simplification, not supported by physical reasoning. The presence of a nonuniform pre-stressed N_{it} profile is usually ignored while modeling the evolution of $N_{\mathrm{it}}(x)$ during hot-carrier degradation. In fact, HCD-induced N_{it} may easily reach $10^{12}\mathrm{cm}^{-2}$, while the initial N_{it} has typical values of $\sim 10^{10}\mathrm{cm}^{-2}$. However, the hot-carrier stress is usually performed at voltages higher than operating ones. At operating conditions and/or short stress times, the interface state density is characterized by moderate values, comparable with pre-stressed N_{it}. Therefore, for a proper description of HCD at operating conditions the information concerning the initial interface state concentration is of significant importance. In this Section an attempt to combine relevant approaches for pre-stressed $N_{\mathrm{it}}(x)$ extraction from CP measurements [164, 173, 175, 177, 179] within a single framework (thereby implementing an exhaustive technique) is undertaken. The suggested scheme takes a number of factors affecting the CP current behavior into account.

For analysis of the pre-stressed N_{it} profile an n-type MOSFET fabricated on a standard 0.35μm CMOS process (device architecture and the net doping profile are sketched in Figure 3.2) is used. The constant base-level charge pumping technique [176, 178, 180] with a fixed base level $V_{\mathrm{gl}} = -5\mathrm{V}$ and V_{gh} increasing from -4V to 4V of the gate pulse has been employed to investigate the interface state density. A device channel width of $W = 10\mu\mathrm{m}$ was chosen to obtain a sufficient charge pumping signal. A standard experimental scheme with the gate of the transistor connected to a pulse generator and a small constant reverse bias applied to the source and drain was used. The substrate current (I_{cp}) was monitored. Only the charge pumping current of the unstressed device is considered in this Section.

4.2.1 Verification of the Interface State Density Distribution Uniformity

The uniformity of the initial interface trap distribution can be verified using the dependence of the maximum value of the charge pumping current ($I_{\mathrm{cp,max}}$) versus the nominal gate electrode length (L_{g}). A linear relationship between $I_{\mathrm{cp,max}}$ and L_{g} would suggest an absence of disparity between the interface state density near the source-drain region and the center of the channel. In this case the assumption that the pre-stressed N_{it} profiles are uniform along the channel is adequate [175]. Extrapolation to $I_{\mathrm{cp,max}} = 0$ reveals the difference (ΔL) between the nominal

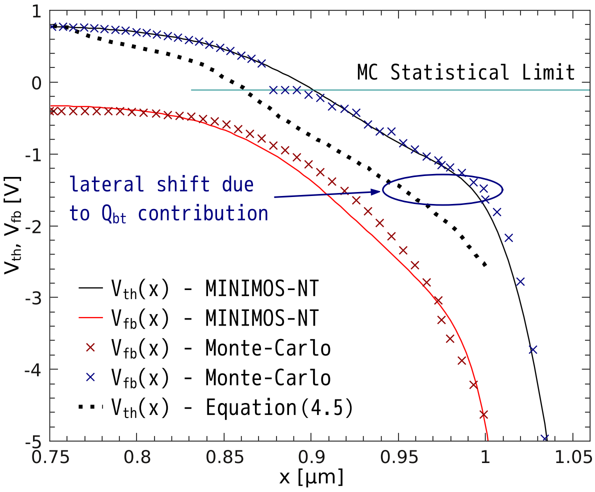


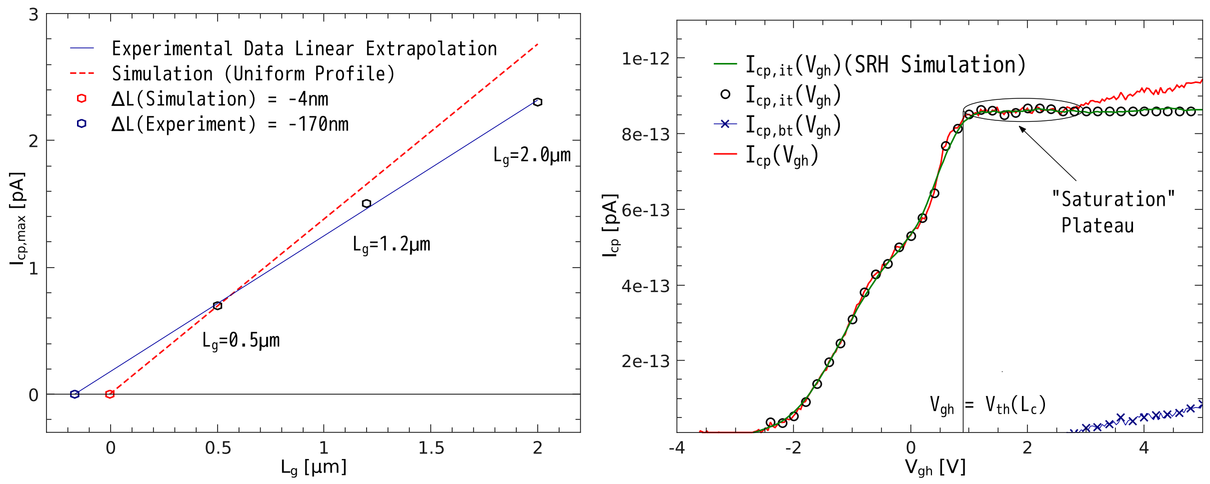
Figure 4.5: Local threshold $V_{\mathrm{th}}(x)$ and flatband $V_{\mathrm{fb}}(x)$ voltage profiles of a pre-stressed device calculated using different approaches.

(L_{g}) and effective (L_{eff}) channel lengths, which contributes to the CP current [173]

$$L_{\mathrm{eff}} = L_{\mathrm{gate}} - \Delta L \tag{4.4}$$

A limited number of works devoted to $N_{\mathrm{it}}(x)$ extraction techniques from the CP current, where this dependence has been verified, is available [164, 175, 177]. However, from the physical viewpoint, it is difficult to evaluate the legitimacy of the obtained findings in the last two papers. In fact, in [177] only two gate lengths are used with a limited L_{g} spread, while a suspiciously linear behavior is presented in [164]. This suggests that this issue demands a more detailed analysis.

To determine the effective length L_{eff}, a careful calculation of local flatband $V_{\mathrm{fb}}(x)$ and threshold $V_{\mathrm{th}}(x)$ voltage distributions as crucial quantities controlling the matter is performed. Due to the symmetry of the source and drain for the fresh device, results only for the drain half of the device are presented, confirmed in [178]. For an unstressed transistor $V_{\mathrm{th}}(x)$ and $V_{\mathrm{fb}}(x)$ can be obtained with the method described in Section 4.1 employing the device simulator MiniMOS-NT. The carrier distribution required for the local threshold and flatband voltage calculation was estimated to be $2 \times 10^{13}\mathrm{cm}^{-3}$ and $8 \times 10^{14}\mathrm{cm}^{-3}$ for the simulation. The validity of the obtained curves has been verified using the Monte-Carlo device simulator MONJU. Figure 4.5 shows a good agreement between the results. Note that in the low voltage region the agreement degrades due to the stochastic nature of the Monte-Carlo approach which can not provide enough precision in this area. As stated in [173, 175], the effective channel length for a constant base-level charge pumping technique is defined as $L_{\mathrm{eff}} = 2|l - L_{\mathrm{c}}|$ for $V_{\mathrm{gh}} \geqslant V_{\mathrm{th}}(L_{\mathrm{c}})$, where l is the solution of the equation $V_{\mathrm{fb}}(l) = V_{\mathrm{gl}}$ and L_{c} is the center point of the channel specified by the device geometry. The simulation yields an effective channel length of 0.504μm while V_{gl} was held at $-5V$ for the fresh device. Figure 4.6a depicts $I_{\mathrm{cp,max}}$ plotted vs. L_{g} and a simple linear extrapolation of CP data gives $L_{\mathrm{eff}} = 0.67\mu$m, i.e. the position is practically on the drain contact. Such an experimental value of the effective channel length is extremely dubious and supports the idea of a non-uniform $N_{\mathrm{it}}(x)$ profile. Let us emphasize that in the case of a homogeneous N_{it} distribution one may obtain a $V_{\mathrm{th}}(x)$ profile directly from the charge pumping current $I_{\mathrm{cp}}(V_{\mathrm{gh}})$



(a)

(b)

Figure 4.6: (a) Maximum value of the charge pumping current for different gate electrode lengths of the fresh 5V n-type MOSFET. (b) Charge pumping current of an undamaged device. Interface and border trap components are revealed.

for the undamaged device employing the method proposed in [175, 178]. According to this approach

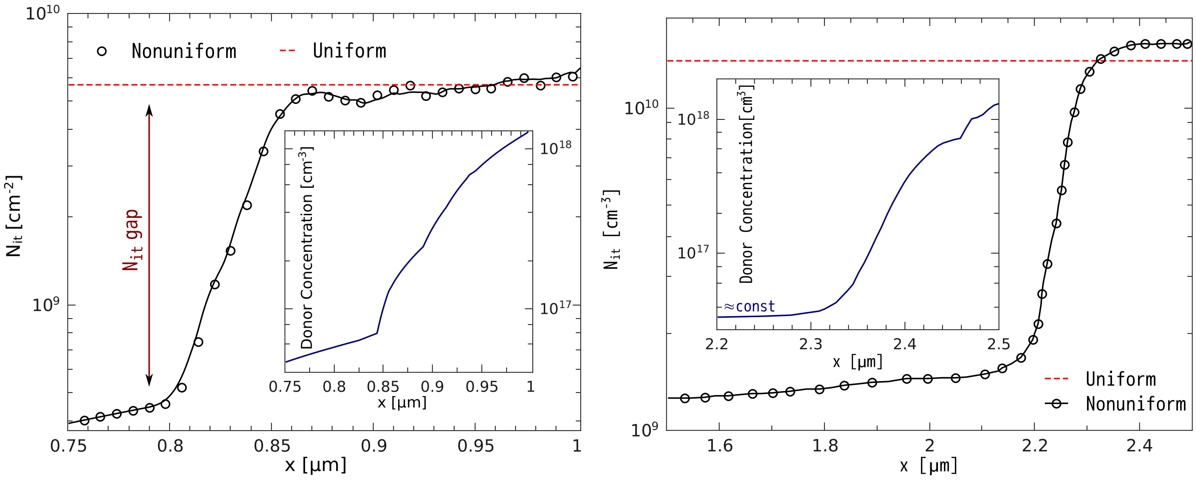
$$V_{\mathrm{th}}(x) = V_{\mathrm{gh}}; \quad x(V_{\mathrm{gh}}) = L_{\mathrm{c}} + \left(1 - \frac{I_{\mathrm{cp}}(V_{\mathrm{gh}})}{I_{\mathrm{cp,max}}}\right)\frac{L_{\mathrm{eff}}}{2}. \tag{4.5}$$

A comparison between this "experimental" and the simulated threshold voltage profiles $V_{\mathrm{th}}(x)$ can confirm the interface state density non-uniformity and is shown in Figure 4.5. However, such a wide discrepancy in the results can not be induced by the difference in initial N_{it} because the V_{th} shift corresponds to an interface trap density of $10^{12}\mathrm{cm}^{-2}$, i.e. for a heavily stressed device. This anomalous offset has a simple explanation: the method based on (4.5) does not work properly due to the border trap contribution to the CP current (see Section 1.3). As it was explained in detail in [181], a presence of trapped border charges Q_{bt} requires special attention to distinguish between contributions to I_{cp} induced by border ($I_{\mathrm{cp,bt}}(V_{\mathrm{gh}})$) and interface states ($I_{\mathrm{cp,it}}(V_{\mathrm{gh}})$)

$$I_{\mathrm{cp}}(V_{\mathrm{gh}}) = I_{\mathrm{cp,it}}(V_{\mathrm{gh}}) + I_{\mathrm{cp,bt}}(V_{\mathrm{gh}}). \tag{4.6}$$

The presence of $I_{\mathrm{cp,bt}}(V_{\mathrm{gh}})$ results in a spurious determination of $I_{\mathrm{cp,max}}$ and L_{eff} and, as a consequence, to an erroneous value for the uniform N_{it} profile. Thus, an additional separation procedure of CP current components must be performed. This separation of $I_{\mathrm{cp,bt}}(V_{\mathrm{gh}})$ from the total CP current relies on the following features of the employed charge pumping technique: at $V_{\mathrm{gh}} \geqslant V_{\mathrm{th}}(L_{\mathrm{c}})$ the interface state density can not increase I_{cp} because the active measurement area has already reached L_{eff}. That is, $I_{\mathrm{cp,it}}(V_{\mathrm{gh}} \geqslant V_{\mathrm{th}}(L_{\mathrm{c}})) = \mathrm{const}$. An explicitly revealed plateau (see Figure 4.6b) corresponds to the saturation of an interface trap induced dose of the charge pumping current and the absence of the Q_{bt} contribution. This means, for $V_{\mathrm{gh}} \geqslant V_{\mathrm{th}}(L_{\mathrm{c}})$ the border trap contribution to the CP signal is

$$I_{\mathrm{cp,bt}}(V_{\mathrm{gh}}) = I_{\mathrm{cp}}(V_{\mathrm{gh}}) - I_{\mathrm{cp}}(V_{\mathrm{th}}(L_{\mathrm{c}})). \tag{4.7}$$



(a) (b)

Figure 4.7: The lateral interface state density: comparison of uniform and nonuniform profiles for channel lengths of (a) 0.5 and (b) 2.0μm.

This finding is supported by the interface state dynamics based on the Shockley-Read-Hall theory [182] (see Figure 4.6b). The experimental data in Figure 4.6a was processed according to the described scheme.

Unfortunately, the method suggested above is not appropriate for the varying base-level technique because a continuous increase of the active measurement area occurs, when this method is implemented. Moreover, in the case of CP measurements with a lower gate pulse frequency (a few kHz [181]) it is not possible to expose the saturation plateau and a more complicated extraction scheme of border trap contributions (as proposed in [34]) must be employed.

4.2.2 Interface State Profile of an Unstressed Device

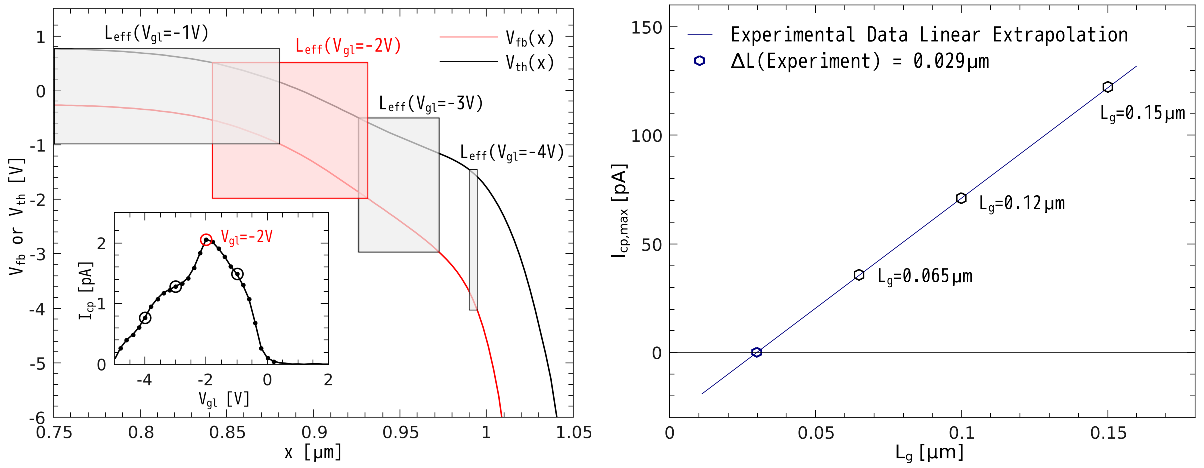
Applying this separation technique it is possible to use the technique for N_{it} extraction developed in [176]. The interface trap concentration of a pre-stressed device can be written as

$$N_{\mathrm{it}}(x) = \frac{1}{\mathrm{q}fW} \frac{\mathrm{d}I_{\mathrm{cp,it}}(V_{\mathrm{gh}})}{\mathrm{d}V_{\mathrm{gh}}} \frac{\mathrm{d}V_{\mathrm{th}}(x)}{\mathrm{d}x}, \tag{4.8}$$

while using the uniform approach [175] results in

$$N_{\mathrm{it,uni}} = \frac{1}{\mathrm{q}fW} \frac{I_{\mathrm{cp}}(V_{\mathrm{th}}(L_{\mathrm{c}}))}{L_{\mathrm{eff}}}. \tag{4.9}$$

Figure 4.7a depicts a comparison between the uniform and nonuniform interface state density profiles. It is believed that the gap in the center of the device can be linked to the peculiarities of the device fabrication process. The inset in Figure 4.7a represents the donor doping concentration with increased gradient at 0.85μm where $N_{\mathrm{it}}(x)$ reaches a uniform value, i.e. the transistor doping can affect the interface state density in this way. An identical tendency shown



(a) (b)

Figure 4.8: (a) The effective channel length determination in the case of constant amplitude charge-pumping current measurement. Inset: charge-pumping current vs. low-level of the gate pulse ($V_\mathrm{A} = 2V$). (b) Maximum value of the charge pumping current for different gate electrode lengths of the fresh CMOS device fabricated on 300mm Si wafers.

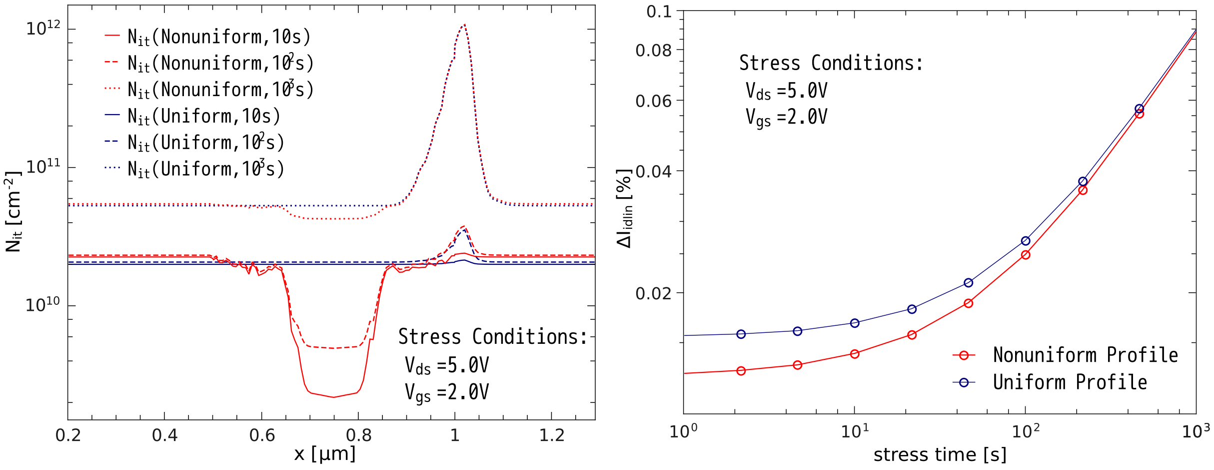
in Figure 4.7b, i.e the correlation between the N_it gap and the donor doping concentration profile, was revealed for the device of the same topology (produced using the same technology) but with a gate length of 2μm.

Moreover, the non-uniformity of the initial N_it profile can be easily visualized within a constant amplitude CP technique. In this case the gate pulse covers the device interface step by step. According to (4.3), the charge pumping current is proportional to the interface state density averaged over the effective measurement length. $I_\mathrm{cp,max}$ depicted in Figure 4.8a inset is reached at a base level of -2V. At the same time, the maximal effective length corresponds to -1V, i.e. in this area the N_it concentration is lower. This approach which is based on another measurement setup additionally confirms previous findings.

However, the demonstrated non-uniformity of the unstressed N_it profile is not a mandatory phenomenon. As an example, a CMOS device fabricated on 300mm (100) Si wafers using a metal-inserted poly-Si process (the high-k dielectric is mainly HfO$_2$) demonstrates in Figure 4.8b a linear dependence between the maximum value of the charge pumping current and gate electrode lengths. This ideal linear graph confirms a uniform initial N_it distribution for such a class of MOSFETs.

4.2.3 Impact of Pre-stressed Interface State Density Profile on HCD Modeling

The information on the initial interface state profile is significant in the context of hot-carrier degradation modeling. The impact of the initial N_it distribution on the degraded device characteristics within the developed physics-based model of HCD (see Section 3.1) is examined. The proposed approach considers the interplay between the single- and multiple-carrier mechanisms of Si – H bond-breakage. The bond is interpreted as a truncated harmonic oscillator with the corresponding ladder of eigenstates. All oscillator transitions are considered within a system



(a) (b)

Figure 4.9: (a) Interface state density profile evolution calculated with developed HCD model under assumption of a uniform initial profile and employing $N_{\mathrm{it}}(x)$ extracted according to the proposed scheme. (b) Comparison of the I_{dlin} degradation calculated with MiniMOS-NT using obtained $N_{\mathrm{it}}(x)$ profiles (a) as input degradation characteristics.

of rate equations. The solution of the system relies on the initial conditions, that is the initial concentration of the virgin $\mathrm{Si}-\mathrm{H}$ bonds. If some of these bonds are already pre-broken, the density of bonds available to be ruptured is reduced. Therefore, the pre-existing dangling bonds already impact the characteristics of the fresh device which differ from those typical for an ideal transistor (with pre-stressed $N_{\mathrm{it}} = 0$) of the same configuration. Due to this initial interface state density profile the solution of the rate equation system – i.e. $N_{\mathrm{it}}(x,t)$ – is modified.

Figure 4.9a (obtained for stress $V_{\mathrm{ds}} = 5.0\mathrm{V}$ and $V_{\mathrm{gs}} = 2.0\mathrm{V}$) illustrates this tendency, showing the evolution of the interface state profile with stress time, calculated by considering the uniform and the nonuniform initial N_{it} distributions. The lack of details obtained from the uniform N_{it} approach is apparent. It can be seen that the difference is more pronounced at short stress times because at longer times the total N_{it} is already rather high, thereby screening the impact of the pre-stressed N_{it} dose. This tendency is confirmed by Figure 4.9b which demonstrates that the linear drain current change (calculated using real and uniform $N_{\mathrm{it}}(x)$ profiles) is significantly modified for stress times shorter than 10^2s. This trend means that an initial interface trap density can affect the device behavior only in the case of short stress times and relatively weak stresses. The second demand corresponds only to the case of real device operating conditions when the applied voltages are relatively low, compared to accelerated hot-carrier stresses.

4.3 Charge-Pumping Extraction Techniques for the Hot-Carrier Induced Interface and Oxide Trap Spatial Distributions in MOSFETs

To reveal and understand the physical picture behind HCD, it is essential to obtain quantitative information regarding the spatial distribution of hot-carrier induced interface states and oxide trapped charges [108]. For this purpose, the charge-pumping technique proves to be a promising approach. However, among numerous variations of the basic CP technique [173], to obtain the N_{it} and N_{ot} lateral profiles the measurement method with varying amplitude (high/low-level) of the gate pulse is employed [66, 177–179]. In addition to the interface state generation the bulk oxide charge build-up is an important component often linked with the hot-carrier stress in literature, which is of special significant in high-voltage devices. Thus, in this Section a careful study of the most widely used extraction techniques of spatial trap distributions from CP data is presented, highlighting their limits of validity and linking them with the developed HCD model.

As in the previous Section 4.2, for these investigations a 5V n-type MOSFET fabricated on a standard 0.35μm CMOS process (Figure 3.2) is used. The lateral coordinate x refers to an origin placed at the left edge of the source contact. A device channel width of $W = 10\mu$m was chosen to obtain a sufficient charge pumping signal. As reported in Section 3.3, since we are dealing with a long-channel device (gate length $L_{\mathrm{g}} = 0.5\mu$m) the WCC of HCD are reached at $V_{\mathrm{gs}} = 0.4V_{\mathrm{ds}}$. Therefore, the device was stressed at $V_{\mathrm{ds}} = 6.5$V, $V_{\mathrm{gs}} = 2.6$V and a temperature of $T = 40^{\circ}$C for 10^5s.

An experimental scheme was employed whereby the gate of the transistor is connected to a pulse generator, while source, drain and substrate contacts are grounded. As it was suggested in [175,177], the CP current at the drain ($I_{\mathrm{cp,d}}$), source ($I_{\mathrm{cp,s}}$) and substrate ($I_{\mathrm{cp,sub}}$) are measured. Due to the presence of the damage dose provided by the multiple-carrier component of the $\mathrm{Si-H}$ bond dissociation process (explained in Section 3.1) it is important to perform separate current measurements. To be more precise, the assumption used in [178,179] that the damage is induced only near the drain edge of the gate contact is incorrect. As a consequence, the drain component of the CP current can not be defined as

$$I_{\mathrm{cp,d}} = I_{\mathrm{cp,sub}}(\mathrm{stress}) - \frac{I_{\mathrm{cp,sub}}(\mathrm{fresh})}{2}. \tag{4.10}$$

The gate of the MOSFET was pulsed by a trapezoidal waveform at a frequency of $f = 100$kHz. A $V_{\mathrm{gl}} = 5$V was used, while V_{gh} was increased from -4 to 3V (varying high-level technique) and V_{gh} was set at 4V, while V_{gl} was decreased from 3 to -5V (varying low-level technique) in 0.02V increments. Here, V_{gl} and V_{gh} are the low and the high levels of the gate pulse, respectively (see Figure 4.10). Such a small voltage step is required in order to obtain a sufficient spatial resolution.

From (4.3) the lateral distribution of interface traps can be calculated by [176]

$$N_{\mathrm{it}}(x) = -\frac{1}{\mathrm{q}Wf}\frac{\mathrm{d}I_{\mathrm{cp,d}}(V_{\mathrm{gh}})}{\mathrm{d}V_{\mathrm{gh}}}\frac{\mathrm{d}V_{\mathrm{th}}(x)}{\mathrm{d}x}, \tag{4.11}$$

in the case of varying high-level CP technique and

$$N_{\mathrm{it}}(x) = \frac{1}{\mathrm{q}Wf}\frac{\mathrm{d}I_{\mathrm{cp,d}}(V_{\mathrm{gl}})}{\mathrm{d}V_{\mathrm{gl}}}\frac{\mathrm{d}V_{\mathrm{fb}}(x)}{\mathrm{d}x}, \tag{4.12}$$

in the case of varying low-level CP technique. Here $V_{\mathrm{th}}(x)$ and $V_{\mathrm{fb}}(x)$ are the local threshold

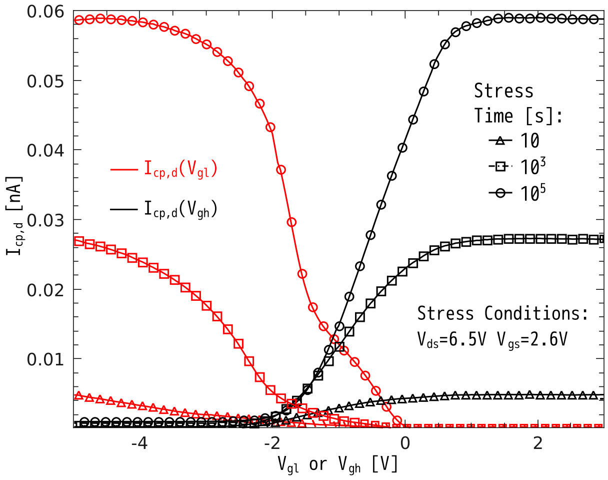


Figure 4.10: Time evolution of the measured charge-pumping current during hot-carrier stress using the varying high/low-level charge-pumping technique.

and flatband voltages described in Section 4.1. If the influence of the interplay between oxide and interface charges is small enough, $V_{\mathrm{th}}(x)$ and $V_{\mathrm{fb}}(x)$ keep the initial form corresponding to an undamaged device during hot-carrier stress. Otherwise necessary corrections of $V_{\mathrm{th}}(x)$ and $V_{\mathrm{fb}}(x)$ after each stress time step must be performed [66, 177, 178]

$$\begin{aligned} V_{\mathrm{th,s}}(x) &= V_{\mathrm{th}}(x) - \frac{\mathrm{q}\Delta N_{\mathrm{ot}}(x)}{C_{\mathrm{ox}}} + \frac{\mathrm{q}\Delta N_{\mathrm{it}}(x)}{2C_{\mathrm{ox}}} \\ V_{\mathrm{fb,s}}(x) &= V_{\mathrm{fb}}(x) - \frac{\mathrm{q}\Delta N_{\mathrm{ot}}(x)}{C_{\mathrm{ox}}} - \frac{\mathrm{q}\Delta N_{\mathrm{it}}(x)}{2C_{\mathrm{ox}}}, \end{aligned} \tag{4.13}$$

where $\Delta N_{\mathrm{it}}(x)$ and $\Delta N_{\mathrm{ot}}(x)$ are the changes between pre- and post-stress concentrations of interface traps and bulk oxide charges, respectively, C_{ox} is a parameter used to depict the oxide capacitance per unit area and will be described below. The coefficient 2 in the denominator of the last term in (4.13) is noteworthy. The interface states are assumed to obey a homogeneous distribution over energy in the band-gap [66, 183]. Also, traps situated above mid-gap assumed to be acceptor-like while those in the lower half of the band-gap are donor-like. In the case of an n-MOSFET with a positive bias applied to the gate, the net charge is negative. The reason behind this is the charging of the acceptor-like interface states. Not disposing at the real distribution of traps over energy it is suggested that the charge stored in these states is $Q_{\mathrm{it}} = -\mathrm{q}N_{\mathrm{it}}(\mathrm{q})/2$, thereby considering a uniform distribution.

In the case of the absence of a significant amount of oxide charges generated under hot-carrier stress it is possible to use a simplified method for N_{it} extraction and consider only the effect of N_{it} on the calculated V_{th}. Namely, only varying high-level or varying low-level CP technique can be used [174].

All approaches described below have been developed in order to obtain the spatial distribution of both types of defects ($N_{\mathrm{ot}}(x)$ and $N_{\mathrm{it}}(x)$) induced by HCD. There are four main concepts for the characterization of the interface states and oxide-trapped charges in MOSFETs: Lee's [178], Li's [179], Chim's [177] and Mahapatra's [66]. Furthermore, the key issues of these extraction procedures are emphasized.

The aforementioned extraction techniques assume that each value of the transistor's local threshold and flatband voltage corresponds to only one position of the coordinate x before and after stress. That is, the effect of the trapped charges on the $V_{\mathrm{th}}(x)$ and $V_{\mathrm{fb}}(x)$ distributions is one dimensional – the shift of the local threshold and flatband voltages at channel position x can be induced only by the charges at position x. This assumption can be additionally examined by using a modified charge-pumping technique where both the high level V_{gh} and the low level V_{gl} of the applied gate pulse are alternating during measurements [175,177]. However, an investigation of the modified CP experimental schemes for verification of this assumption is out of the scope of this work.

4.3.1 Local Oxide Capacitance

All considered methods employ the oxide capacitance C_{ox} as a key parameter of the extraction procedure. In the methods derived by Mahapatra [66] and by Li [179] C_{ox} is treated as a constant

$$C_{\mathrm{ox}} = \frac{\varepsilon_{\mathrm{ox}}}{t_{\mathrm{ox}}}, \tag{4.14}$$

where t_{ox} is the oxide thickness at the device center and $\varepsilon_{\mathrm{ox}}$ the dielectric permittivity.

This approach is applicable only for transistors with a fixed oxide thickness and leads to erroneous results when t_{ox} depends on interface coordinate x. Moreover, even in the case of uniform thickness distribution, such a scheme leads to spurious results. In fact, in such an approach the capacitor is considered ideal i.e. the electric field is assumed uniform. In work of Chim *et al.* [177] it was possible to find an extension of (4.14) when t_{ox} depends on x ($t_{\mathrm{ox}} = t_{\mathrm{ox}}(x)$). In practice, however, a substantial distortion occurs near the source/drain ends of the gate. The aforementioned simplification does not strongly affect the transistor characteristics. However, the electric field non-uniformity is of particular importance for the extraction of the interface state density profile after HC stress because $N_{\mathrm{it}}(x)$ peaks near the drain end of the gate where the capacitor non-ideality is most pronounced [41]. Additionally, Lee *et al.* [178] have already demonstrated that the coordinate dependence of the capacitance $C_{\mathrm{ox}}(x)$ due to the fringing effect is essential.

Conformal-Mapping Method

The conformal-mapping method is most helpful for the calculation of the fringing electric field in simple two-dimensional boundary problems (which is our case) by transforming the boundary to a soluble form [184]. A series of simulations based on the Lee *et al.* approach [178] (described below) allow us to draw the conclusion that, for a local oxide capacitance consideration, a simplified structure can be used. Namely, the gate contact can be interpreted geometrically as a ray instead of more complicated corner variants. The problem with the coordinate system is shown in Figure 4.11. Here, the origin of the considered system is placed at the drain end of the gate. The $z = x + iy$ plane is mapped into the $\omega = \varphi + i\psi$ plane with the functional relationship between z and ω described by

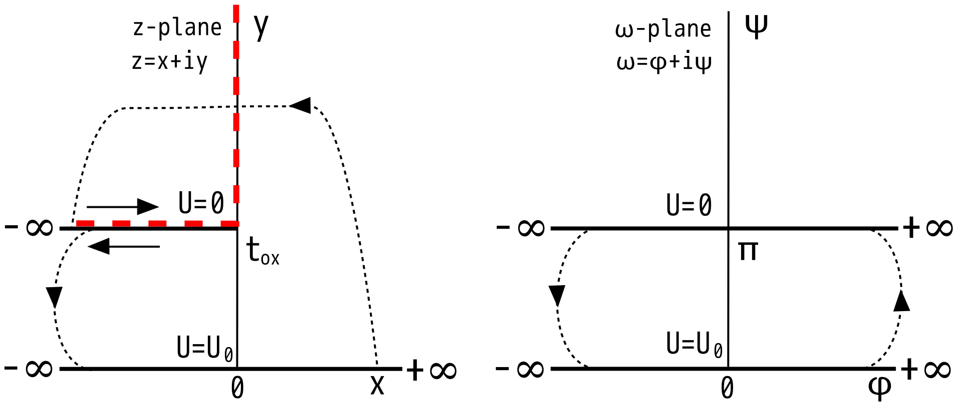


Figure 4.11: The conformal transformation used to solve the gate/drain fringing problem. The complicated case of the corner gate form (red dashed line) is reduced to a ray.

$$z = \frac{t_{\mathrm{ox}}}{\pi}\left[\omega + \exp(\omega)\right], \tag{4.15}$$

or in the coordinate notation

$$\left.\begin{aligned} x &= \frac{t_{\mathrm{ox}}}{\pi}\left[\varphi + \exp(\varphi)\cos\psi\right] \\ y &= \frac{t_{\mathrm{ox}}}{\pi}\left[\psi + \exp(\varphi)\sin\psi\right] \end{aligned}\right\}. \tag{4.16}$$

The suggested conformal transformation shown in Figure 4.11 reduces the pristine problem to the Laplace problem between two parallel infinitely long metallic plates at different potentials [184]. The analytical solution for the potential and the electric field distribution in the oxide near the gate end following from the conformal map and that obtained by MiniMOS-NT are presented in Figure 4.12. One can see a rather good agreement between the obtained results which means that the chosen geometrical system interpretation is correct.

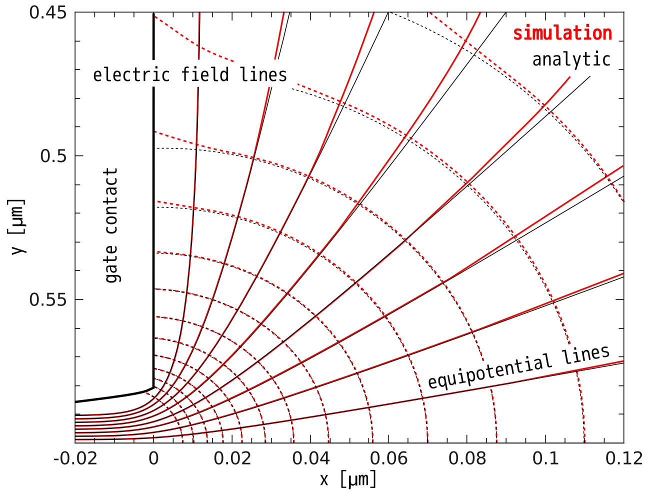


Figure 4.12: The comparison of equipotential and field-lines in the oxide near the gate corner as the analytical solution and simulated by means of device simulator MiniMOS-NT.

The local oxide capacitance can be defined as the ratio between the surface charge density σ_{surface} and the interface potential U_0, written as

$$C_{\text{ox}}(x) \equiv \frac{\sigma_{\text{surface}}(x)}{U_0}. \tag{4.17}$$

That is, for the determination of the local oxide capacitance it is first of all necessary to find the surface charge density σ_{surface}. According to [185]

$$\sigma_{\text{surface}}(x) \equiv \varepsilon_{\text{ox}} \left.\frac{\partial U(x,y)}{\partial y}\right|_{y=0}, \tag{4.18}$$

and the solution of the potential problem for the suggested conformal transformation is

$$U(\varphi,\psi) = \frac{\pi-\psi}{\pi}U_0, \ \text{ i.e } \ -\frac{\partial U}{\partial y} = \frac{U_0}{\pi}\frac{\partial\psi}{\partial y}. \tag{4.19}$$

After differentiation of (4.16) against y one can write

$$\frac{\partial\psi}{\partial y} = \frac{\pi}{t_{\text{ox}}}\frac{1+\exp(\varphi)\cos\psi}{1+2\exp(\varphi)\cos\psi+\exp(2\varphi)}. \tag{4.20}$$

Furthermore, when $y=0$ (i.e. at lower infinite plate) $\psi = 0$

$$\left.\frac{\partial\psi}{\partial y}\right|_{y=0} = \frac{\pi}{t_{\text{ox}}}\frac{1}{1+\exp(\varphi)}. \tag{4.21}$$

Substituting (4.21) in (4.17) gives us

$$C_{\text{ox}}(x) = \frac{\varepsilon_{\text{ox}}}{t_{\text{ox}}}\frac{1}{1+\mathrm{e}^{\varphi}}. \tag{4.22}$$

That is, the local oxide capacitance can be written as a parametric system

$$\left.\begin{aligned} x &= \frac{t_{\text{ox}}(x)}{\pi}[\varphi+\exp(\varphi)] \\ C_{\text{ox}}(x) &= \frac{\varepsilon_{\text{ox}}}{t_{\text{ox}}(x)[1+\exp(\varphi)]} \end{aligned}\right\}. \tag{4.23}$$

It should be mentioned that in (4.23) the parameter t_{ox} has been changed to $t_{\text{ox}}(x)$. As reported in [185], this substitution is legitimate until $\mathrm{d}t_{\text{ox}}(x)/\mathrm{d}x < 1$ which is correct for devices with a real topology of the gate contact. The value of $C_{\text{ox}}(x)$ in (4.23) has a maximum with $C_{\text{ox}}(= \varepsilon_{\text{ox}}/t_{\text{ox}})$ in the middle of the gate and decreases gradually toward a much lower value outside the gate edge due to the fringing effect. At $x \to -\infty$ (or $\varphi \to -\infty$) the obtained equation (4.23) asymptotically turns into an expression for the parallel-plate capacitance (4.14).

Estimation of Approach Importance Area

It is rather important to clarify at which distance from the drain edge of the gate contact the electric field becomes uniform. In other words, where the fringing effect is negligible and the parallel-plate capacitor approximation is applicable. For this purpose let us find a position on the x-axis where the electric field intensity differs by 1% from the uniform one, i.e $F_0 = U_0/t_{\mathrm{ox}}$ [184]. For any position of consideration in the Figure 4.11 system, the electric field intensity can be defined using the employed conformal transformation as $F = |\mathrm{d}\omega/\mathrm{d}z|$. Therefore, from (4.21) one may conclude

$$F = \frac{U_0}{t_{\mathrm{ox}}} \frac{1}{1+\exp(\omega)}, \tag{4.24}$$

or in terms of F_0

$$\frac{F}{F_0} = \frac{1}{\exp(\varphi)+1}. \tag{4.25}$$

Assuming $F/F_0 = 0.99$ we obtain $\exp(\varphi) = 0.0101$ and, consequently, $\varphi = -4.61$. A substitution of the obtained values in the expression for the x coordinate in (4.16) results in

$$x_0 = 4.61 \frac{t_{\mathrm{ox}}(x)}{\pi} = 1.47 t_{\mathrm{ox}}(x = L_{\mathrm{g}}). \tag{4.26}$$

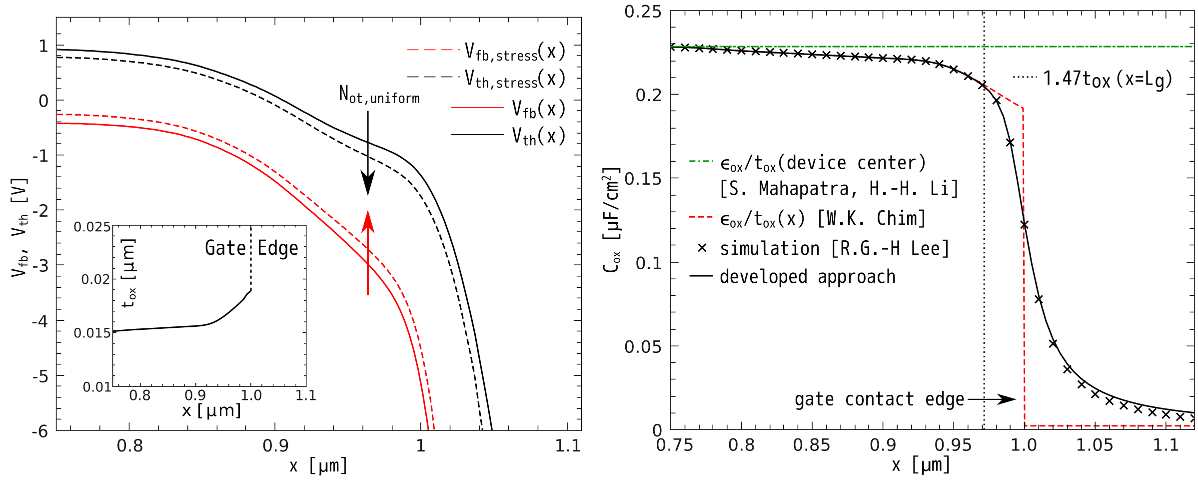
Here, $x = L_{\mathrm{g}}$ is the position of the drain end of the gate. Thus, one cane consider the local oxide capacitance as constant (i.e. use the parallel-plate capacitor approximation) already at a distance of 1.47 of the oxide thickness at the drain end of the gate contact. As a result, it is possible to operate with a simple expression (4.14) if the region of interest is outside of the mentioned area.

Lee's Approach for Modeling of Local Oxide Capacitance

For evaluating the local oxide capacitance using device modeling we employ the method suggested by Lee *et al.* [178]. To determine the $C_{\mathrm{ox}}(x)$ by means of simulations, a careful calculation of the local flatband $V_{\mathrm{fb}}(x)$ and threshold $V_{\mathrm{th}}(x)$ voltage distributions is performed. Due to the symmetry of the source and drain for the fresh device, one presents results only for the drain half of the device [178]. As mentioned in Section 4.1, for an unstressed transistor the $V_{\mathrm{th}}(x)$ and $V_{\mathrm{fb}}(x)$ profiles can be obtained by the device simulator MiniMOS-NT. For any point on the device interface the post-stress local threshold $V_{\mathrm{th}}(x)$ and local flatband $V_{\mathrm{fb}}(x)$ values are related to the pre-stressed ones by (4.13) which considers the change between pre- and post-stress concentrations of interface traps and bulk oxide charges. The presence of a uniform probe oxide charge $N_{\mathrm{ot,uniform}}$ leads to a local threshold voltage shift $\Delta V_{\mathrm{th,uniform}}(x) = V_{\mathrm{th,s(uniform)}}(x) - V_{\mathrm{th}}(x)$ [178]. Therefore, the local oxide capacitance can be found using

$$C_{\mathrm{ox}}(x) = -\frac{N_{\mathrm{ot,uniform}}}{\Delta V_{\mathrm{th,uniform}}(x)}. \tag{4.27}$$

A typical example of $\Delta V_{\mathrm{th,uniform}}(x)$ induced by a uniform oxide charge density of $5 \times 10^{11}\mathrm{cm}^{-2}$ is shown in Figure 4.13a.



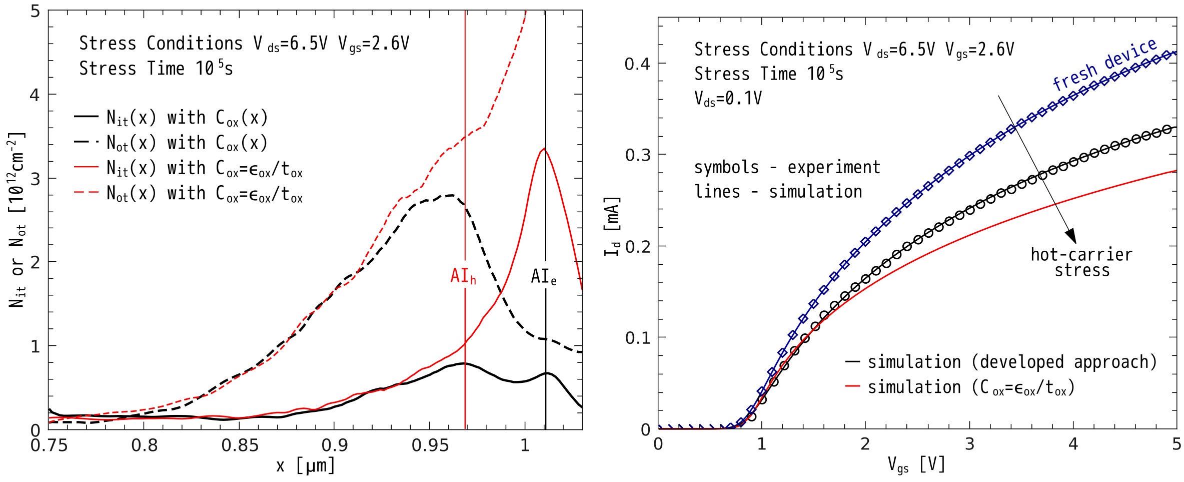
(a) (b)

Figure 4.13: (a) Local threshold and flatband voltage distributions with uniform oxide charge profiles. Inset: the dependence of the oxide thickness vs. the lateral coordinate. (b) The local oxide capacitance calculated using the approach of Lee *et al.*, compared with the newly developed analytical model. Here L_g is the position of the drain end of the gate.

4.3.2 Comparison of Interface State Profiles Extracted with Different Capacitance Approaches

A comparison between the simulations and the developed analytical approach for the local capacitance profile is presented in Figure 4.13b. It can be concluded that the obtained results are in good agreement. At the same time, the expression $C_{ox} = \varepsilon_{ox}/t_{ox}$ – even when corrected for $t_{ox} = t_{ox}(x)$ (Figure 4.13a, inset) – leads to substantially different results. The value of $C_{ox}(x)$ has its maximum with $C_{ox}(= \varepsilon_{ox}/t_{ox})$ in the middle of the gate and decreases gradually toward a much lower value outside of the gate edge due to the fringing effect. As expected, the most pronounced peculiarity is observed at the drain side of the gate where the abrupt change in the oxide thickness occurs, Figure 4.13a, inset. As predicted by analysis of the conformal map one can see that under the gate electrode the fringing effect can be neglected for distances longer than $1.47t_{ox}(x = 1\mu\text{m})$ from the gate edge (this is reflected in Figure 4.13b). Note that an abrupt reduction in $C_{ox}(x)$ is not a physical phenomenon and such an approach should not be used.

To extract the density of interface and oxide traps an analytical $C_{ox}(x)$ profile is incorporated into the scheme described in [177]. The results obtained under different assumptions regarding the $C_{ox}(x)$ distribution for a stress time of 10^5s are presented in Figure 4.14a. The extracted $N_{it}(x)$ and $N_{ot}(x)$ profiles were subjected to further validation as input parameters to simulate the transfer characteristics of the degraded device employing MiniMOS-NT. A comparison of the simulated and experimental curves once again confirms the applicability of the developed model for local oxide capacitance (see Figure 4.14b). The role of the fringing effect at large stress times is self-evident. Therefore, further in our investigations the developed model for the local oxide capacitance is used for all considered characterization methods.



(a) (b)

Figure 4.14: (a) The lateral distributions of $N_{\rm it}(x)$ and $N_{\rm ot}(x)$ calculated using different $C_{\rm ox}(x)$ approaches for 10^5s at $V_{\rm ds}$=6.5V and $V_{\rm gs}$=2.6V. Peaks of $N_{\rm it}(x)$ correspond to the maxima of electron and hole acceleration integrals. (b) A comparison of experimental and simulated transfer characteristics for fresh and stressed devices.

4.3.3 Extraction Techniques Description

The general features of the most commonly used extraction techniques are analyzed below.

Chim's Approach [177]

This extraction algorithm is based on the iterative correction scheme employing the equation system (4.13). The pre-stress CP edge is corrected for the additional charges associated with both $N_{\rm ot}$ and $N_{\rm it}$ generated due to the stress. For each position x_i, the maximum possible range where the post-stress local threshold voltage can lie is found. By comparing the varying pulse-high and varying pulse-low measured CP currents under consideration for a chosen $V_{\rm th}(x_i)$ value from this range, $N_{\rm it}$ is separated from $N_{\rm ot}$. Namely, for an assumed $V_{\rm th}(x_i)$ using (4.11) $N_{\rm it}(x_i)$ is calculated. Further, for a known value of the interface state density at x_i, the local flatband voltage $V_{\rm fb}(x_i)$ is found employing (4.12). The obtained results are $I_{\rm cp}(V_{\rm gh} = V_{\rm th}(x_i))$ and $I_{\rm cp}(V_{\rm gl} = V_{\rm fb}(x_i))$. Since the maximum value of the CP current $I_{\rm cp,max}$ is the same for both varying high/low-level measurements (the effective channel length is the same), the following condition must be satisfied (see, Figure 4.15)

$$I_{\rm cp,max} = I_{\rm cp}(V_{\rm gh} = V_{\rm th}(x_i)) + I_{\rm cp}(V_{\rm gl} = V_{\rm fb}(x_i)). \tag{4.28}$$

If (4.28) is not satisfied, the chosen $V_{\rm th}(x_i)$ value is incorrect and it is necessary to repeat this procedure from the start. Otherwise, a solution is found and we can go to the next coordinate x_{i+1}.

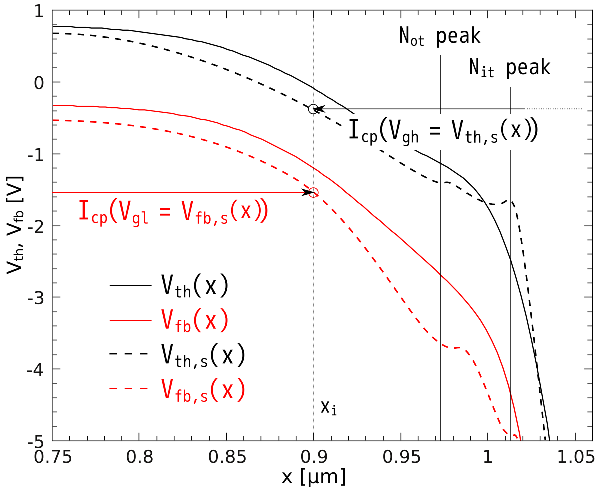


Figure 4.15: Typical local threshold and flatband voltage profiles before $\{V_{\mathrm{th}}(x), V_{\mathrm{fb}}(x)\}$ and after $\{V_{\mathrm{th,s}}(x), V_{\mathrm{fb,s}}(x)\}$ hot-carrier stress for n-type MOSFET.

Mahapatra's Approach [66]

Here, it is assumed that $N_{\mathrm{it}}(x)$ follows the lateral electric field distribution along the channel. The main idea is that the damage profile can be modeled best as Gaussian. However, a Gaussian distribution is inconvenient for analytical modeling because the integral results in the error-function. Therefore, the authors suggest to model the N_{it} profile by an analytically integrable function with a similar shape

$$N_{\mathrm{it}}(x) = \frac{N_{\mathrm{it,p}}}{\cosh^2\left(\alpha[x - x_{\mathrm{p}}]\right)}, \tag{4.29}$$

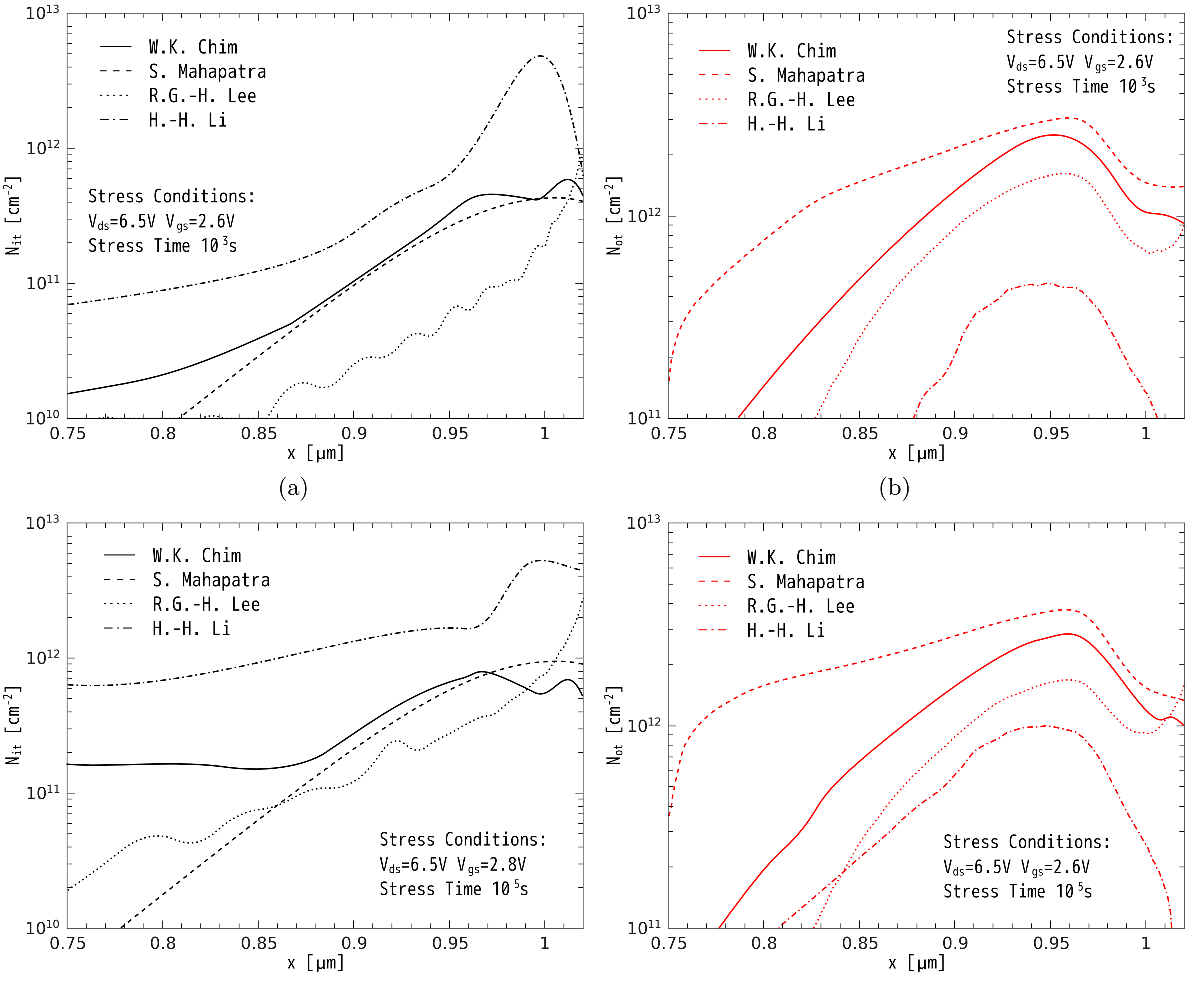
where $N_{\mathrm{it,p}}$ is the peak value of the damage, x_{p} is the position of the peak along the channel and α is a parameter responsible for the spatial spread of the damage. This number of parameters is defined by the fitting procedure of experimental CP data. An implementation of (4.29) in (4.13) significantly simplifies the extraction process.

Lee's Approach [178]

In this routine, it is supposed that the hot-carrier induced impact on $V_{\mathrm{th}}(x)$ for the damaged region near the drain side can be empirically represented as

$$V_{\mathrm{th,s}}(x) = \frac{V_{\mathrm{th}}(x_2) - V_{\mathrm{th}}(x_1)}{\ln(1 - (x_2 - x_1)/A)} \cdot \ln\left(1 - \frac{x - x_1}{A}\right) + V_{\mathrm{th}}(x_1), \tag{4.30}$$

while the undamaged region remains unchanged, i.e. $V_{\mathrm{th,s}}(x) = V_{\mathrm{th}}(x)$. Here $x_{1,2}$ are the boundaries of the damaged oxide region (defined by the evolution of the derivative of the charge-pumping curve with stress time) and A is the only unknown variable to be determined by the optimization technique. As in Mahapatra's approach, substituting of (4.30) in (4.13) results in a simplified defect profile characterization.



(a) (b) (c) (d)

Figure 4.16: The lateral distributions of (a),(c) $N_{\rm it}(x)$ and (b),(d) $N_{\rm ot}(x)$ calculated using different extraction approaches for 10^3s and 10^5s at $V_{\rm ds}$=6.5V and $V_{\rm gs}$=2.6V.

Li's Approach [179]

This technique stands out against the others because it requires only varying high-level measurements. At the same time, the suggested approach is original and worthy of attention. It is derived from the Shockley-Read-Hall theory which is used for the analytic charge-pumping current model. According to this model, $N_{\rm it}(x)$ is

$$N_{\rm it}(x) \simeq \frac{I_{\rm cp}(V_{\rm gl,2}) - I_{\rm cp}(V_{\rm gl,1})}{2{\rm k_B}T_{\rm L}WfB\int_0^\infty (P(V_{\rm gl,1},t) - P(V_{\rm gl,2},t)){\rm d}t}, \tag{4.31}$$

where $B \equiv \log(\sigma_{\rm n0}t_{\rm eme}(x)/\sigma_{\rm p0}t_{\rm emh}(x))$. Here $t_{\rm eme}(t_{\rm emh})$ is the non-steady-state electron (hole) emission time measured from two low gate voltages ($V_{\rm gl,1}$ and $V_{\rm gl,2}$). The fitting function P is obtained using device simulation and is equal to 1 at the center of the channel and rapidly decreases to 0 near the source/drain junction. The oxide-trapped charge density is derived as $N_{\rm ot}(x) = \Delta V_{\rm c}C_{\rm ox}$, where $V_{\rm c}$ is the difference of the voltage related to the half-maximum value of the $\Delta I_{\rm cp}/f - V_{\rm gh}$ curves.

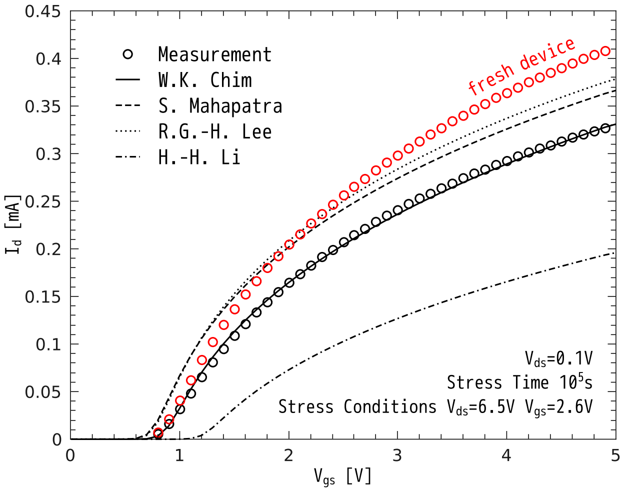


Figure 4.17: The comparison of the experimental and simulated transfer characteristics for fresh and stressed devices.

Extraction Methods Results Comparison

The profiles extracted for a stress time of 10^3s and 10^5s are depicted in Figure 4.16. The extracted $N_{\rm it}(x)$ and $N_{\rm ot}(x)$ dependences were subjected to further validation as input data of the device simulator MiniMOS-NT to model the transfer characteristics of the degraded device. A comparison of simulated and experimental curves checks the applicability of the considered techniques (see Figure 4.17).

One can see that the $I_{\rm d} - V_{\rm ds}$ dependence evaluated with Li's model spuriously overestimates the damage as compared to other extraction schemes. This fact can be explained by the necessity to recalculate the P function after every iteration with an erroneously localized $N_{\rm ot}(x)$ profile as an input parameter. As mentioned above, the spatial distribution of the oxide traps is defined using the change of half maximum of the charge-pumping curve. Such a variation is a very insensitive parameter and, as a consequence, the technique for calculating $N_{\rm it}$ and $N_{\rm ot}$ which is self-consistent, makes it questionable whether such additional efforts are justified [183].

At the same time, Mahapatra's approach predicts a transfer curve degradation close to the result obtained by Lee's method but not close enough to the experimental data. In literature one may find different criteria for evaluating the efficiency of interactions between the carriers and the bond, i.e. the maximum of the electric field, the carrier dynamic temperature, etc [29]. However, as proposed in Section 3.1, the $N_{\rm it}$ peak only corresponds to the maximum of the acceleration integral and is shifted away with respect to the maxima of other quantities. Consequently, for a proper HCD model it is not possible to use a simple analytical expression, such as (4.29) and (4.30). It is important to emphasize that the $N_{\rm it}(x)$ profiles extracted within Chim's technique features two peaks, see Figure 4.16a and Figure 4.16c. This finding is in good agreement with the results of the described HCD model, which shows that these peaks are related to the contributions induced by primary channel electrons and secondary generated holes (the latter is shifted towards the source) and correspond to the maxima of the electron and hole acceleration integrals (see Section 3.2). Moreover, in Section 3.1 one has already demonstrated that for long-channel devices the multiple-carrier process of $\mathrm{Si-H}$ bond-breakage leads to N_{it} homogeneously distributed over the interface, which is confirmed in Figure 4.16c (N_{it} plateau at the center of the gate).

4.4 Analysis of the Threshold Voltage Turn-Around Effect in n-MOSFETs Due to Hot-Carrier Stress

In this Section the turn-around effect of the threshold voltage shift during hot-carrier stress is investigated. Such a phenomenon is explained by the interplay between interface states and oxide traps, i.e. by the compensation of the rapidly created oxide charges by the more slowly created interface states. To prove this idea, a refined extraction scheme for the defect distribution from charge-pumping measurements has been employed. The obtained results are in good agreement with the findings of the proposed physics-based model of hot-carrier degradation (see Section 3.2). The approach considers not only damage produced by channel electrons but also by secondary generated channel holes. Although the contribution of the holes to the total defect creation is smaller compared to that of electrons, their impact on the threshold voltage shift is comparable with the electronic one. The reason behind this trend is that hole-induced traps are shifted towards the source, thereby more severely affecting the device behavior.

It is commonly accepted that processes related to trapping/detrapping in the oxide bulk play a crucial role in the bias temperature instability (BTI) [186, 187]. Since BTI and hot-carrier degradation are believed to have a similar microscopic origin, one may expect that bulk oxide traps with the density N_{ot} should also contribute to HCD. At the same time, N_{ot} dominates the recoverable component of BTI [138], suggesting that HCD should also recover considerably. However, in n-MOSFETs hot-carrier degradation demonstrates a rather small recovery over a wide range of temperatures [29, 41]. The situation is made even more complicated because of turn-around effects sometimes observed during hot-carrier stresses [109, 173]. These turn-around effects are related to the partial compensation of the charge stored in the oxide traps by interface state trapping characterized by the density N_{it} [29,173]. It is worth emphasized that a systematic study considering the evolution of the lateral defect profiles with stress time has not been carried out. The main scope of this part of the work is to analyze the turn-around of the V_{th} shift observed during HC stress. This strategy is supported by charge-pumping data and is linked to the developed HCD model. The model considers not only channel electrons but also secondary holes generated by impact ionization caused by the injection of hot electrons. These holes are then accelerated by the electric field towards the source, thereby creating interface states shifted with respect to the electron-induced ones. As a result, the N_{it} fraction induced by holes should be much less than their relative contribution to the linear drain current change (see Section 3.2).

For an investigation of the turn-around effect a 5V n-type MOSFET fabricated in a standard 0.35μm CMOS process (depicted in Figure 3.2) is employed. The device channel length is 0.5μm. The device was stressed at three combinations of drain and gate voltages: $\{V_{\mathrm{ds}}, V_{\mathrm{gs}}\}$=$\{6.5,2.6\}$, $\{6.75,2.0\}$ and $\{6.25,2.0\}$V and a temperature of T=40°C. Such a temperature regime was used because trapping/detrapping in the oxide bulk is triggered by non-elastic trap-assisted tunneling

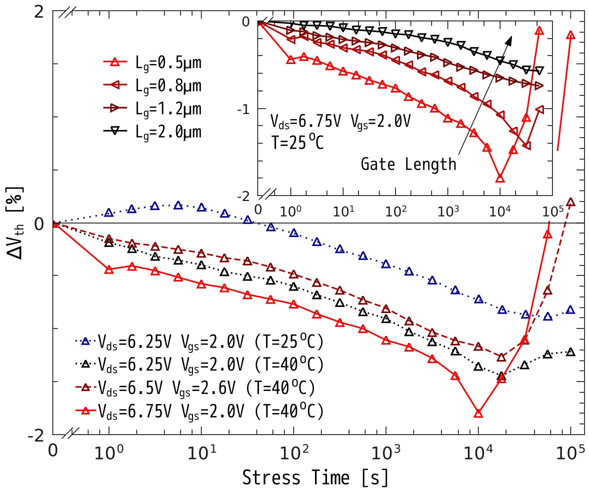
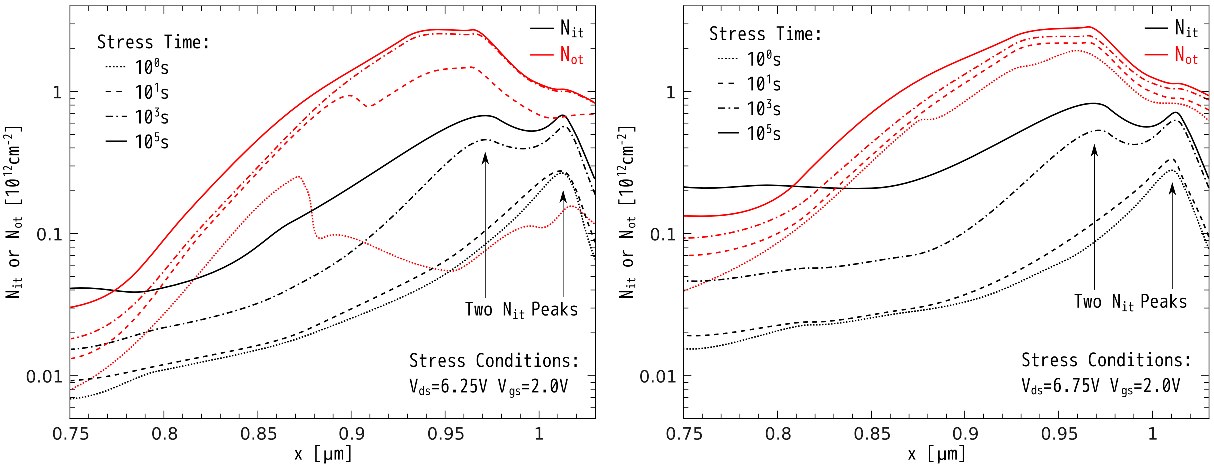


Figure 4.18: The shift of V_{th} as a function of stress time at various voltages. Inset: ΔV_{th} measured for devices of the same architecture but different only in channel lengths stressed at V_{ds}=6.75V and V_{gs}=2.0V.

which is accelerated at higher temperatures [188,189]. In the inset of Figure 4.18 the shift of the threshold voltage V_{th} vs. the stress time is depicted. One can see that the turn-around effect of the V_{th} shift [173] is less pronounced at room temperature. These turn-around effects are related to a partial compensation of the charge stored in the oxide traps by interface state trapping [29]. In this context, the chuck temperature was set to 40°C to accelerate the degradation processes.

Based on the analysis of Section 4.3, stress voltages close to the WCC were intentionally chosen in order to maximize the contribution of all degradation effects. The threshold voltage was monitored up to 10^5s. Figure 4.18 demonstrates that V_{th} first decreases due to hole trapping in the oxide bulk while after 10^3s V_{th} increases due to the trapping of electrons by interface traps. To check these speculations, the CP technique with varying amplitude of the gate pulse [66,177] has been employed to investigate the $N_{\text{it}}(x)$ and $N_{\text{ot}}(x)$ distributions. A standard experimental scheme is used, where the gate of the transistor is connected to a pulse generator and a small constant reverse bias is applied to the source and drain. The measured current is the substrate current and a gate pulse frequency is 25kHz. A measurement setup is used with $V_{\text{gl}} = 5\text{V}$ and V_{gh} is increased from -4 to 2V (varying high level technique) and $V_{\text{gh}} = 3\text{V}$ with decreasing V_{gl} from 2 to -5V (varying base level technique) in 0.02V increments.

Furthermore, to extract the evolution of $N_{\text{it}}(x)$ and $N_{\text{ot}}(x)$ profiles with the stress time, the analytical $C_{\text{ox}}(x)$ distribution (4.23) incorporated into the scheme of Chim's *et al.* extraction technique [177], based on system (4.13) is employed. The extracted $N_{\text{it}}(x)$ profile features two peaks, see Figure 4.19. This finding is in good agreement with the results of the developed HCD model which shows that these peaks are related to the contributions induced by primary channel electrons ans secondary generated holes (note that the latter one is shifted towards the source). In Section 3.1 and Section 3.2 it is demonstrated that in long-channel devices the multiple-carrier process of $\text{Si}-\text{H}$ bond-breakage leads to N_{it} homogeneously distributed over x (see Figure 4.19). At the same time, this is the component which gives the main contribution to ΔV_{th} because the device is most sensitive to N_{it} created closer to the channel middle (and in this device area the SP-related contribution is negligible). Hence, notwithstanding the fact that



(a) (b)

Figure 4.19: The evolution of $N_{\mathrm{it}}(x)$ and $N_{\mathrm{ot}}(x)$ profiles with stress time for (a) V_{ds}=6.25V and (b) V_{ds}=6.75V, the gate-source stress voltage was set to 2V.

the maximum value of N_{ot} is higher than the initial N_{it} peak value (Figure 4.19), N_{it} starts to prevail over N_{ot} at longer times ($\geqslant 10^4$s in this particular case). This is because at these times only the peak N_{ot} value is higher than the N_{it} maximum while within the device center (which defines the V_{th} behavior) N_{it} is higher. The strong localization of the N_{it} SP-component near the drain edge of the gate and, as a consequence, the secondary influence on ΔV_{th} can be confirmed by the saturation of V_{th} and a turn-around effect. This trend is depicted in Figure 4.18, inset for four MOSFETs of the same architecture with different gate lengths.

One may conclude that the hole contribution is considerably shifted towards the source. The single-electron component generates traps situated outside of the channel which explains why the hole-induced traps have a stronger relative impact on the threshold voltage turn-around effect. The contribution of channel holes to the total defect concentration is much less than the corresponding fraction of V_{th} change. This trend becomes more pronounced for longer devices (Figure 4.18, inset), see also Section 3.2. Another tendency confirming these considerations is that the ΔV_{th} turn-around effect is less pronounced at room temperature (Figure 4.18 for the stress regime with $V_{\mathrm{ds}} = 6.25V$ and $V_{\mathrm{gs}} = 2.0V$). This is because trapping/detrapping in the oxide bulk is triggered by non-elastic trap-assisted tunneling which is accelerated at higher temperatures [188].

In Figure 4.20 the $N_{\mathrm{ot}}(x)$ profiles are resolved for short stress times. Initially, two peaks are differently positioned in space while after ~5s they unite into a single common hump. It is worth emphasizing that the position of the $N_{\mathrm{it}}(x)$ and $N_{\mathrm{ot}}(x)$ peaks does not change with time (Figure 4.19). This is in good agreement with the model predictions where the position of the carrier acceleration integral maximum coincides only with the $N_{\mathrm{it}}(x)$ peak position.

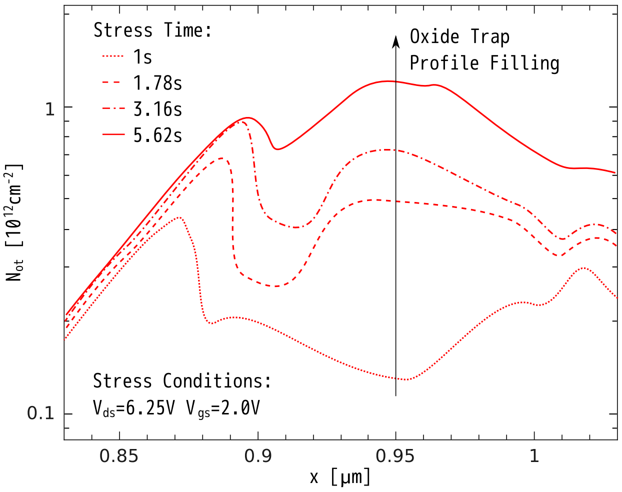


Figure 4.20: Filling of the oxide traps revealed for small stress times at the stress regime with V_{ds}=6.5V and V_{gs}=2.6V.

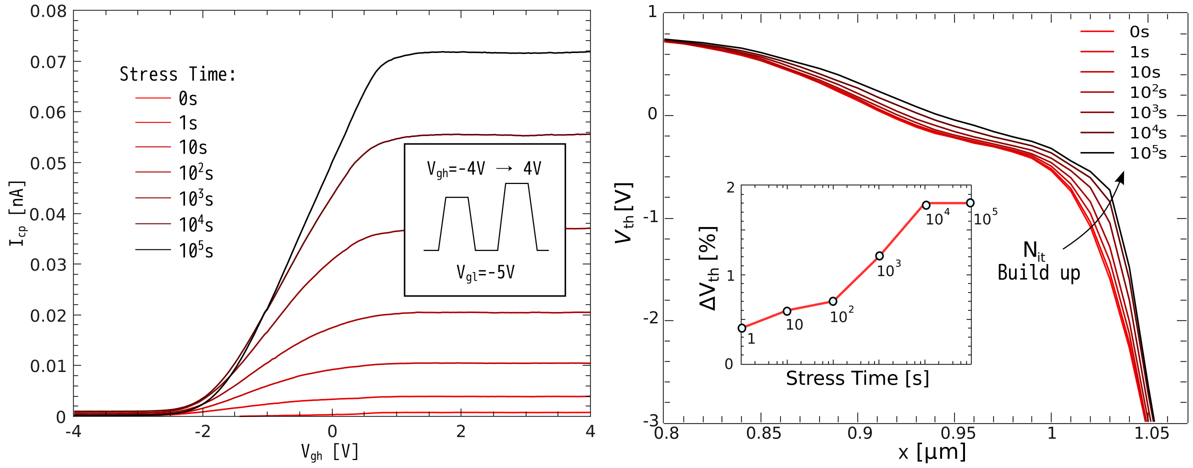
4.5 HCD Caused Interface State Profile – Simulations vs. Experiment

In this Section a direct comparison of simulated and experimental interface state density profiles is undertaken. For this purpose a 5V n-type MOSFET presented in Figure 3.2 is used as an experimental base. The devices have been stressed at a gate voltage V_{gs} of 2.0V and source-drain voltage V_{ds} of 6.75V. The ambient temperature was $T = 23^{\mathrm{o}}\mathrm{C}$ and the stress time $t = 10^5$s. The constant-base-level CP technique [176, 178, 179, 190, 191] with $V_{\mathrm{gl}} = -5\mathrm{V}$ and increased V_{gh} from -4V to 4V in 0.004V increments is used. Here V_{gl} and V_{gh} are the base and the high level of the gate pulse, respectively (see Figure 4.21a inset). As stated above, such a small step in gate pulse amplitude is set to obtain a detailed spatial resolution. Below the approach suggested in [174, 176, 190] is employed for the extraction of the N_{it} profile from the CP current.

The $V_{\mathrm{th}}(x)$ distribution keeps its initial form corresponding to an undamaged device during HC stress only if the influence of the interplay between oxide and interface charges is small enough [174–176]. As explained in Section 4.3 in the case of the absence of a significant amount of oxide charges generated under hot-carrier stress one can use a simplified method for interface state density extraction and consider only the effect of $N_{\mathrm{it}}(x)$ on the calculated $V_{\mathrm{th}}(x)$. To be more precise, in the case of the technique with varying high level of the gate pulse, the extraction procedure can be reduced to a self-consistent solution of the system

$$\left.\begin{aligned} N_{\mathrm{it}}(x) &= \frac{1}{\mathrm{q}fW}\frac{\mathrm{d}I_{\mathrm{cp,d}}(V_{\mathrm{gh}})}{\mathrm{d}V_{\mathrm{gh}}}\frac{\mathrm{d}V_{\mathrm{th,s}}(x)}{\mathrm{d}x} \\ V_{\mathrm{th,s}}(x) &= V_{\mathrm{th}}(x) + \frac{\mathrm{q}\mathit{\Delta} N_{\mathrm{it}}(x)}{2C_{\mathrm{ox}}(x)} \end{aligned}\right\} . \tag{4.32}$$

In [22, 179, 190, 191] the impact of the bulk oxide trap N_{ot} and interface trap density N_{it} on the CP current has been analyzed. A change in $N_{\mathrm{ot}}(x)$ without changing $N_{\mathrm{it}}(x)$ results only in a local shift of I_{cp} along the V_{gh} axis. On the other hand, an increase of $N_{\mathrm{it}}(x)$ without changing



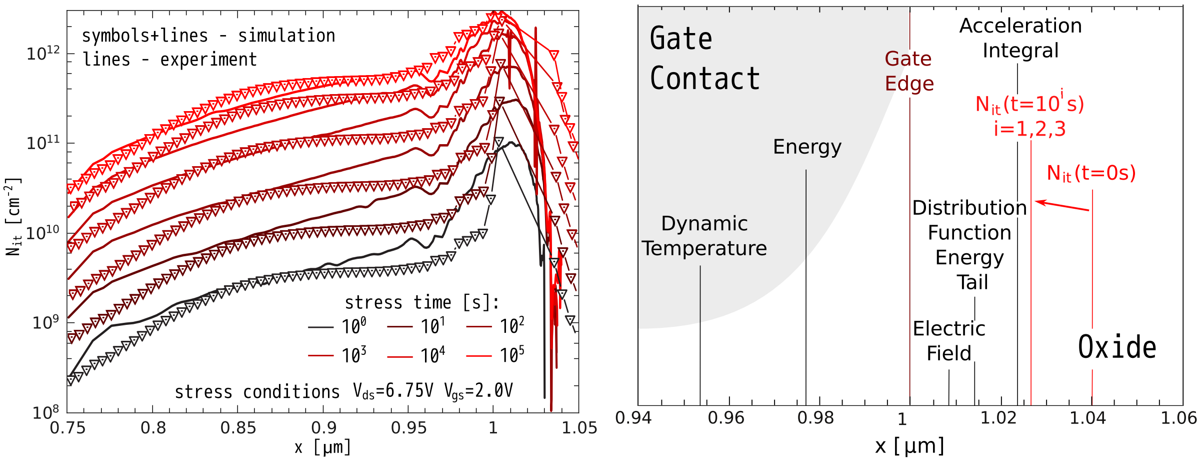
(a) (b)

Figure 4.21: (a) Time evolution of the measured $I_{\mathrm{cp}} - V_{\mathrm{gh}}$ relationship during hot-carrier stress using the constant-base-level charge-pumping technique. Inset shows the gate pulse train. (b) Local threshold $V_{\mathrm{th}}(x)$ voltage distributions along the channel before and after the stress. The shift of the threshold voltage ΔV_{th} measured by the maximum transconductance method is shown in inset.

$N_{\mathrm{ot}}(x)$ will only cause a local increase of $\mathrm{d}I_{\mathrm{cp}}/\mathrm{d}V_{\mathrm{gh}}$. Thus, based on the shape evolution of the I_{cp} and $\mathrm{d}I_{\mathrm{cp}}/\mathrm{d}V_{\mathrm{gh}}$ curves for different HC stress times it is believed that no significant amount of oxide charges are generated under the chosen stress conditions. This fact is confirmed by a measurement of the threshold voltage shift ΔV_{th} by the maximum transconductance method (see Figure 4.21b inset). The increase of ΔV_{th} demonstrates the dominance of N_{it} generation during the entire measurement time (see Section 4.4). Simply speaking, the aforementioned facts allow us to conclude that system (4.32) is valid and applicable for the established stress regime. A comparison of the extracted and the simulated data with HCD model interface state density profiles for several stress times is demonstrated in Figure 4.22a.

As mentioned in Chapter 1, it is usually assumed that the most severe hot-carrier degradation is localized of the electric field (or the driving force) peaks. This is because in this area carriers gain their maximum energy and the Si – H bond dissociation process is most intensive. However, the maxima of such physical quantities as the electric field, electron energy or dynamic temperature are not necessarily observed at the same position. Moreover, since both SP- and MP-processes are controlled by the AIs, just the maxima of these integrals correspond to the N_{it} peak. The AI maximum is controlled by the shape of the distribution function or – more precisely – the depth of the high energy tails. Therefore, the device area where the most extended tails of the DF are observed corresponds to the highest degradation dose. To conclude, it is rather important to compare coordinates for the most intensive trap creation, determined according to different criteria and the maxima of N_{it} (as well as N_{it} peak for the undamaged device). This information is provided in Figure 4.22b.

One essential feature should be emphasized. As demonstrated in Section 4.2, the “virgin” device is characterized by the pre-existing interface state density and the initial N_{it} has a peak at a certain position which differs from the position of the maximal AI(s) and is described by features of the fabrication process. This results in a shift of the interface state density profile maximum



(a) (b)

Figure 4.22: (a) The interface state density vs. coordinate x for several stress times: experiment vs. theory. (b) The position of the maxima of main physical quantities: the electric field, average carrier energy, acceleration integral, position of the most extended tails of the, etc.

for the first few seconds of the degradation (Figure 4.22b). After 10s of hot-carrier stress the peak of the experimental N_{it} profile practically coincides with the peak of I_{SP} (peak of the theoretical N_{it}), with the difference in positions less than $\sim$ 4nm. Such a system behavior depicted in Figure 4.22b allowing for the conclusion that the description of the hot-carrier degradation is untenable in the framework of the most principal physical quantities. Figure 4.22b shows that the peak positions are distributed in space and the extension of this distribution is $\sim$ 100nm which is already rather large (especially compared to the distance between N_{it} and AI maxima). Moreover, these results once again confirm that the maximum of the interface state profile should be described in terms of the AI rather than linking it with the position of the deepest high-energy tail of the distribution function [23,40]. Both SP- and MP-mechanisms of $\mathrm{Si}-\mathrm{H}$ bond-breakage are described by the carrier AI controlled by the DF. The structure of the integrand herein contains a superposition of rapidly decaying high-energy tails of the distribution function and quickly increasing capture cross sections, carrier velocity and the density-of-states. Therefore, the maximum of this integral is shifted with respect to the position where the DF demonstrates the deepest high-energy tails.

After 10^3s it is possible to observe a weak second shift ($\sim$17nm) of the N_{it} peak to the drain contact side. This could be explained by the necessity of a self-consistent simulation of the DF (or still existence of N_{ot} build-up influence). Figure 4.23 demonstrates a strong localization of the degradation portion associated with the single-particle process. The interface state density corresponding to the SP-mechanism abruptly changes practically from 0 to a significant value (compared to Figure 4.24a). In contrast, the MP-related acceleration integral does not change so drastically with the coordinate. Three pronounced humps in the profile mimic the energy tails of the DF. The first (centered at $x \approx 0.25\mu$m) and the third (positioned at $\approx 1.25\mu$m) hump are associated with the p-n junctions of the source/channel and channel/drain sections of the MOSFET while the second one is observed near the drain end of the gate, i.e. just near the areas where the DF demonstrates deepest high-energy tails. This tendency is reflected on the SP- and MP-induced interface states profiles plotted in Figure 4.24a and Figure 4.24b, respectively.

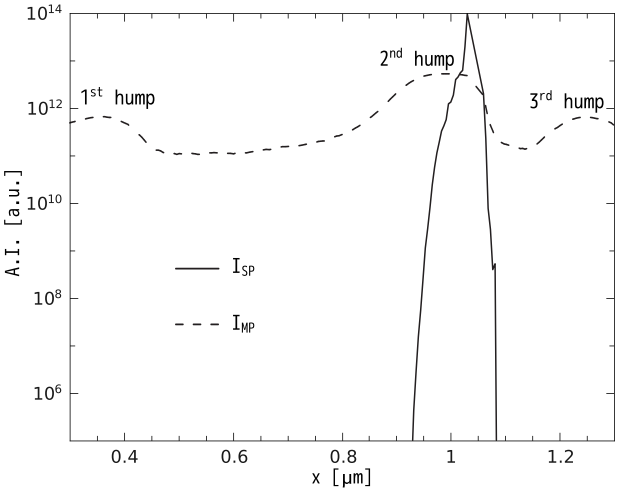
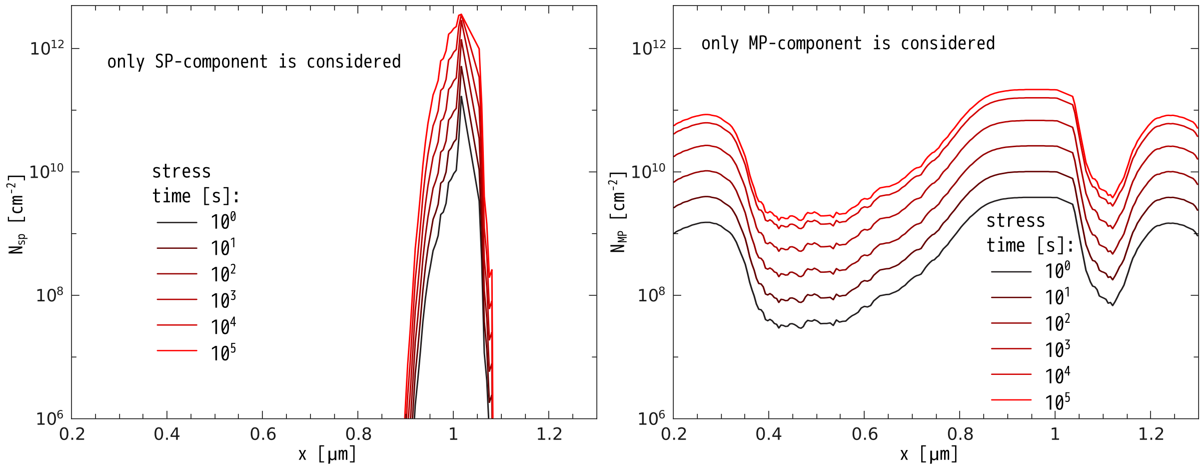


Figure 4.23: The dependence of the acceleration integral for the SP- (solid line) and MP- (dash line) processes. Integrals are given in arbitrary units.

Note that the AI for the MP-process controls the interface state density via the prefactor $(P_\mathrm{u}/P_\mathrm{d})$ in expression (3.9) and thus its variation is not as large as in the case of the SP-component. Figure 4.24b shows that N_MP represents the behavior of the corresponding AI. This means that electrons in the channel cannot reach high energies in this range and are "cold" from the perspective of the SP-process but still have sufficient energy to contribute to the MP-mechanism. In conclusion, the MP-component is present at all values of the x coordinate, in contrast to the SP-component which contributes only in a rather narrow region. It should be noted that only N_MP is responsible for the flat plateau situated left of the maximum of the total concentration N_it (this plateau is schematically shown in Figure 4.22a).

It is worth commenting on the choice of parameters entering the acceleration integral. For the SP-process parameters similar to those reported in previous papers are employed (e.g. [30–32, 99]): the threshold energy for the bond-breakage process is set equal to $E_\mathrm{th,SP} = 1.5\mathrm{eV}$ and the exponent in the Keldysh-like reaction cross section $p_\mathrm{it,SP} = 11$. The parameters characterizing the energetics of the truncated harmonic oscillator were similar to those listed in Table 1.1, however, experimental data is fitted with our simulations, taking $p_\mathrm{it,MP} \sim 0.1$, i.e. the process is practically energy-independent. Actually, $p_\mathrm{it,SP}$ and $p_\mathrm{it,MP}$ describe different physical processes, i.e. the excitation of the bonding electron to the antibonding state in the former case and electron-phonon interaction in the latter one. Generally speaking, these processes are described by different values of reaction cross sections. The value of $E_\mathrm{th,MP}$ equal to 1.5eV (used in the model) should also be verified, because it is also inherited from the SP-process. Although the pair of parameters $p_\mathrm{it,MP} \sim 0.1$, $E_\mathrm{th,MP} = 1.5\mathrm{eV}$ satisfactorily described the matter, this issue must be clarified. A weak energy dependence of the cross section for the MP-process may also be related to the parameter dispersion, when the large standard deviation dictates that an effectively smaller value of the parameter produces the main contribution.

Due to the employment of the Monte-Carlo method, calculations of the distribution function with a good spatial resolution and accurate high-energy tails is a time-consuming challenge. Therefore, in spite of the good agreement between the simulated and extracted N_it profiles,



(a) (b)

Figure 4.24: Both (a) SP- and (b) MP-related components of the total interface state density plotted vs. the lateral coordinate.

there is room for improvement at the expense of increased computational costs. Although good agreement between the experiment and simulation has been achieved, the theoretical N_{it} profile may be further refined. The stochastic nature of the Monte-Carlo method is responsible for some computational noise while evaluating the distribution function. Therefore, one may envisage a niche for improvement related to perfect calculations of the DF. Alternatively, a refinement may be achieved in the extraction technique of the N_{it} profile from CP data. In (4.32) $C_{\mathrm{ox}}(x)$ is considered as a constant with stress time, which could work only as a first order approach. The presence of the oxide and/or interface charges transforms C_{ox} into a function of the interface coordinate and degradation time. Already at the present stage one can emphasize the necessity of a self-consistent procedure for the extraction of the N_{it} profile from CP measurements – after each time step of the HC stress influence of the N_{it} and N_{ot} on the threshold and/or flatband voltages should be taken into account. Otherwise a discrepancy between the extracted and the experimental values could amount up to one order of magnitude.

The main advantage of the developed model is that one does not have to recalibrate it when switching from one device architecture and/or stress conditions to another. In other words, the set of parameters describing the $\mathrm{Si-H}$ dissociation kinetics is fitted once and does not depend on process conditions. As for the transport module, the Boltzmann transport equation is to be solved each time the device topology and stress conditions are changed in order to obtain the new set of DFs corresponding to the current situation. As a result, the model is suitable to predict the device life-time not only for accelerated stress conditions but also under normal operating conditions and thus is useful for development and reliability engineers. Another advantage is that the model provides information about the N_{it} profiles, and in in particular captures the strong localization of HCD. Two additional important peculiarities of hot-carrier degradation captured by the model are the saturation of the damage as well as the representation of the worst-case conditions.

5 Analytic Modeling Approach for HCD

In this Chapter an analytical approach for hot-carrier degradation modeling based on the rigorous physics-based TCAD model is developed. The model employs an analytical approximation of the carrier acceleration integral by a fitting formula. The essential features of hot-carrier degradation such as the interplay between single- and multiple-particle components of $\mathrm{Si-H}$ bond dissociation, mobility degradation during interface state build-up, as well as the saturation of degradation at long stress times are presented. As a result, the change of the linear drain current can be represented by the analytical expression over a wide range of stress conditions. The analytical model can be used to study the impact of device geometric parameters on hot-carrier degradation.

5.1 An Analytical Approach for Physical Modeling of HC Induced Degradation

One of the manifestations of HCD is the linear drain current I_{dlin} degradation. The I_{dlin} change over time is a consequence of a large number of microscopic contributions. These contributions are related to the dangling bonds produced due to the $\mathrm{Si-H}$ bond dissociation. A dangling bond can capture a carrier, thereby producing a charged defect, which perturbs the electrostatics of a device and degrades the mobility. The driving force of the bond-breakage process is the energy deposited by carriers and thus the distribution of the energy describes the intensity of the bond dissociation mechanism. At the same time, the carrier ensemble evolves as it moves through the device. Physically this means that the carrier distribution function (and the acceleration integral) changes with the coordinate along the interface. Since the acceleration integral defines the characteristic time for trap creation, the interface states are distributed spatially and over the characteristic time. This time, in turn, describes the activation exponent and thus the dynamics of the defect generation process.

The analogous situation is typical for negative bias temperature instability where the dispersion of time constants (describing defect activation) stems from the distribution of the activation energy [136,192,193]. The mathematical formalism for this case was elaborated using the Fermi-Dirac function derivative as the probability density for the activation energy. This shape of the distribution leads to a power law dependence of the degrading parameter, see [40,194,195]. In the case of HCD the situation is more complicated because the activation of the $\mathrm{Si-H}$ bond dissociation process is triggered by the AI which is coordinate dependent, see Section 3.1. The bonds located at the coordinate corresponding to a smaller value of the AI will be broken later than those located near the drain end of the gate, i.e. near the typical AI peak. Therefore, one

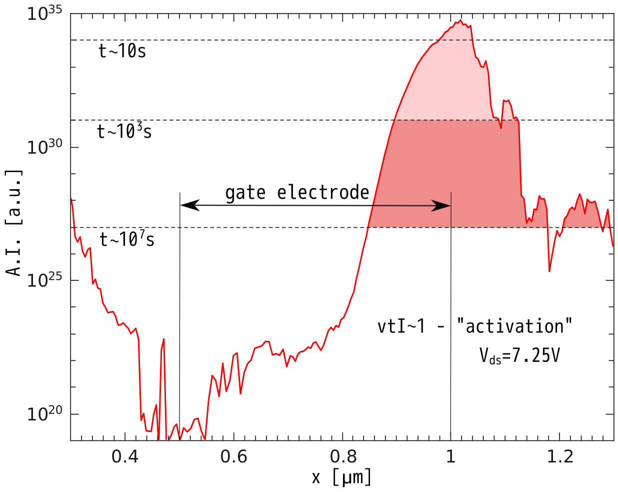


Figure 5.1: An example of a TCAD result for the AI with characteristic activation times being sketched.

should consider the matter in terms of the distribution of the acceleration integral $I(x)$ with the lateral coordinate x.

Based on the detailed model for HCD the acceleration integral is fitted with an analytical formula. Such an approach allows for the interface state density N_{it} and, hence, the degradation of the linear drain current, to be expressed analytically. One of the main features of this concept is that it is based on the physical principles behind HCD rather than on empirical fitting of experimental data. In particular, this means that the interplay between the single- and multiple-carrier mechanisms of the defect creation is considered by the model. This is very important because HCD tests are performed at accelerated conditions where carriers are rather hot. At real operation conditions, however, the voltages are lower, resulting in lower carrier energies. As a consequence, the role of the multiple-particle process is accelerated. Therefore, under real operating conditions the relative contribution of the latter mechanism is expected to be increased and the transition from stress to operating conditions may only be captured by a physics-based HCD model.

For these investigations a HV 5V n-MOSFET (sketched in Figure 3.2) designed on a standard 0.35μm technology is employed. To ensure such a high operating voltage a thicker gate oxide (with an oxide thickness of 15nm), a longer channel thickness and special measures concerning the source/drain doping were undertaken. The devices have been subjected to hot-carrier stress for 10^4s at room temperature, using the gate voltage $V_{\mathrm{gs}} = 2.0$V and 5 different drain voltages V_{ds} = 6.25, 6.5, 6.75, 7.0, 7.25V. To calculate the carrier distribution function the proposed TCAD model for hot-carrier degradation, as described in Section 3.1, is employed. A typical shape of the $I(x)$ profile (calculated for $V_{\mathrm{ds}} = 7.25$V) is represented in Figure 5.1. This information will be further used as an additional verification data. Moreover, the information on $I(x)$ is crucial for the derivation of the analytical model because it will be based on fitting of this profile by an analytical expression. As an additional verification data input the dependence of the AI on the coordinate along the interface x, calculated with the TCAD model described in Section 3.1, is employed.

The linear drain current degradation due to N_{it} build-up is [30]

$$\frac{\Delta I_{\mathrm{dlin}}}{I_{\mathrm{dlin0}}} = \frac{\Delta V_{\mathrm{th}}}{V_{\mathrm{gs}} - V_{\mathrm{th}}} + \frac{\Delta \mu}{\mu_0}, \tag{5.1}$$

where I_{dlin0} and μ_0 are the linear drain current and mobility of the fresh device, respectively, and $\Delta\mu$ (ΔV_{th}) is the change in mobility (threshold voltage). The mobility of the degraded device is calculated via (3.13). Relying on the experimental observation that, in the considered device ΔV_{th} is weak, ΔI_{dlin} and N_{it} are linked as

$$\Delta I_{\mathrm{dlin}} = \frac{\alpha N_{\mathrm{it}}}{1 + \alpha N_{\mathrm{it}}}. \tag{5.2}$$

As in the TCAD version two competing mechanisms for $\mathrm{Si-H}$ bond-breakage are considered, i.e. the single-particle and multiple-particle processes [30]. The interface states generated by the SP-process are characterized by the density $N_{\mathrm{it,SP}}$ (see (3.3)) which is found by assuming first-order kinetics for the bond-breakage reaction. As shown in Section 3.1, the MP-related contribution is not coordinate-dependent because the acceleration integral reaches high values. Therefore, the density $N_{\mathrm{it,MP}}$ is not considered as a spatially distributed quantity.

Since our aim is to develop an analytical model which does not use any information from the TCAD version (e.g. lateral profiles such as $N_{\mathrm{it}}(x)$ and $I(x)$), for the SP-process the average (over the interface) an effective trap concentration is introduced

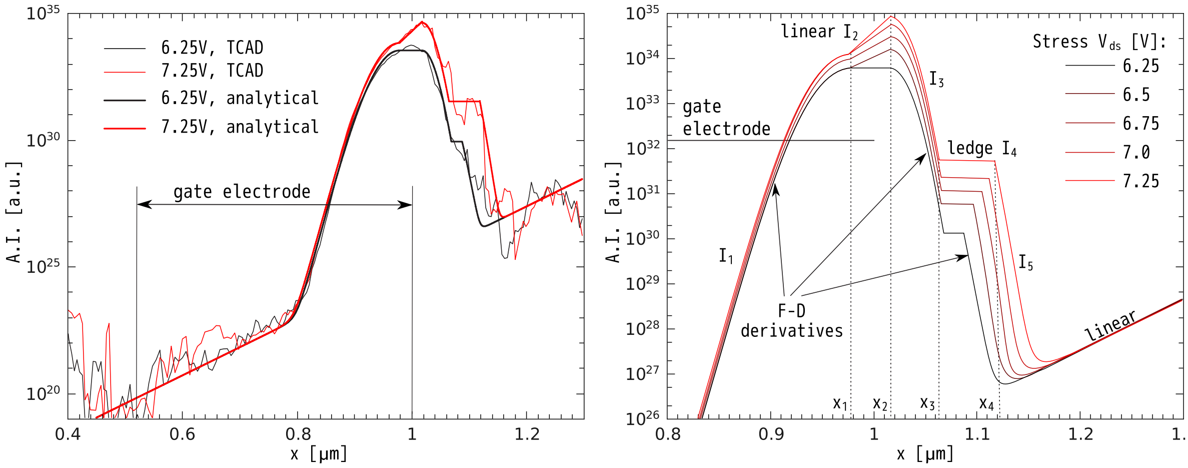
$$\overline{N}_{\mathrm{SP}} = N_0 \left\{ 1 - \frac{1}{x_{\max} - x_{\min}} \int\limits_{x_{\min}}^{x_{\max}} \exp\left[-t\nu I(x)\right] \mathrm{d}x \right\}, \tag{5.3}$$

with $x_{\max}$ and $x_{\min}$ as the boundaries of the $\mathrm{Si/SiO_2}$ interface. The integration over x captures all values of the AI. The probability that a $\mathrm{Si-H}$ bond is broken within a certain time period is given by (5.3). The coordinate-dependent acceleration integral $I(x)$ defines the characteristic time of defect creation for each particular position along the interface.

Figure 5.1 shows the AI profile $I(x)$ for $V_{\mathrm{ds}} = 7.25\mathrm{V}$ which can be decomposed into two parts, i.e. the linear dependence (on a log-scale) on the coordinate I_{linear} and the peak $I_{\mathrm{peak}}(x)$, see Figure 5.1 and Figure 5.2a

$$\begin{aligned} I(x) &= I_{\mathrm{linear}}(x) + I_{\mathrm{peak}}(x) \\ I_{\mathrm{linear}} &= A_1 \exp(\alpha[x - x_{\min}]). \end{aligned} \tag{5.4}$$

The decaying line observed in the range 0.3...0.5μm can be omitted because only bonds where the condition $\nu t I \sim 1$ is satisfied have a significant probability to be broken; i.e. this fragment corresponds to the stress time $t > 10^7$s and the dissociation process is not triggered within the given time period. The slopes of the AI peak are well described by Fermi-Dirac derivatives on a log-scale. The peak of the AI is represented by piecewise functions, see Figure 5.1 and



(a) (b)

Figure 5.2: (a) A representation of the AI with an analytical expression for various V_{ds}. (b) The profile of the acceleration integral: comparison between the AI calculated with the TCAD model and the analytical expression.

Figure 5.2b

$$
\begin{aligned}
& I_{\mathrm{peak}}(x) = I_1(x) + I_2(x) + I_3(x) + I_4(x) + I_5(x) \\
& F(x, \langle x \rangle, A, \sigma) = A \exp\left(c \, \frac{\exp\left([x - \langle x \rangle]/\sigma\right)}{\left[1 + \exp\left([x - \langle x \rangle]/\sigma\right)\right]^2} \right) \\
& I_1 = F(x, x_1, A_2, \sigma_1) && x \leq x_1 \\
& I_2 = A_2 \exp\left(c \left[1/4 + \beta(x - x_1)\right]\right) && x \in (x_1; x_2] \\
& I_3 = F(x, x_2, A_3, \sigma_2) && x \in (x_2; x_3] \\
& I_4 = I_3(x_3) && x \in (x_3; x_4] \\
& I_5 = F(x, x_5, A_4, \sigma_2) && x \geq x_4 \\
& x_5 = x_2 + x_4 - x_3
\end{aligned}
\tag{5.5}
$$

It can be observed (Figure 5.2a and Figure 5.2b) that the slope of the left front (i.e. I_1) and the right front (I_3 and I_5 with the same σ_2) do not vary with V_{ds}. The position of the maxima of I_1 and I_3 (x_1 and x_2, respectively) remains the same with the increasing V_{ds} and the only varying parameters are the slope of I_2 (parameter β) and the extension of the ledge I_4 (i.e. $x_4 - x_3$); the position of the I_5 maximum is linked to the end of I_4, see (5.5). The values of those parameters which vary with the stress V_{ds} (i.e. fitting parameters of the model) are shown in Figure 5.3. Their variation with V_{ds} can be represented by a linear dependence. Note that scattering in these parameters is related to the noise in the TCAD data, which is a consequence of the MC method employed for the solution of the Boltzmann transport equation. This circumstance allows for the calculate of the AI for those stress conditions for which the set of DFs has not been calculated. For this purpose values of must be interpolated/extrapolated $\{A_2, A_3, A_4, \beta, x_4\}$

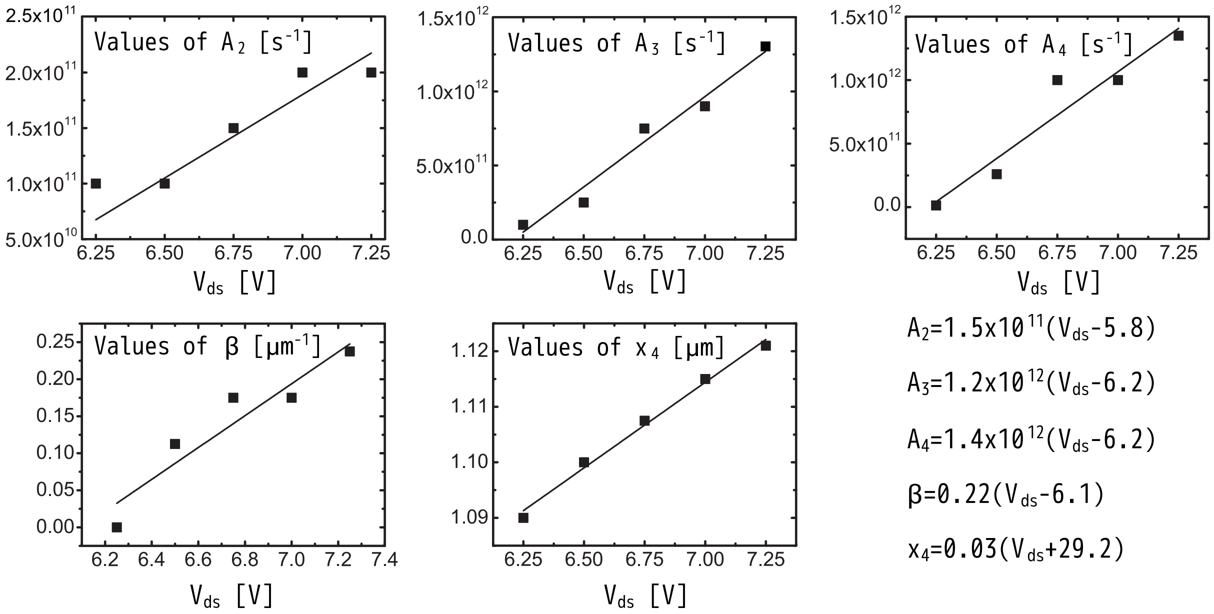


Figure 5.3: The parameters A_2, A_3, A_4, β and x_4 vs. V_{ds} as well as their fitting by the linear dependence on V_{ds}.

using the findings from Figure 5.3.

The full solution of (5.3) is written as

$$\overline{N}_{\mathrm{SP}} = 1 - (x_{\max} - x_{\min})^{-1} \sum_{\mathrm{i}} J_{\mathrm{i}}, \tag{5.6}$$

where J_{i} is the contribution produced by the term I_{i}, i.e.

$$J_{\mathrm{i}} = \int \exp\left[-\nu t I_{\mathrm{i}}(x)\right] \mathrm{d}x. \tag{5.7}$$

The terms J_{i} are expressed via the exponential integrals of the first kind $\mathrm{Ei}_1(x)$ which is defined as [197]

$$\mathrm{Ei}_1(x) = \int_x^{\infty} \frac{\exp -t}{t} \mathrm{d}t, \qquad |\arg(x) < \pi|. \tag{5.8}$$

The contributions J_{i} have the following structure (terms J_3 and J_5 are similar to J_1 and thus are omitted here)

$$\begin{aligned}
J_1 &= \alpha^{-1}\mathrm{Ei}_1\left(\nu t A_1 \exp\left[-\alpha(x - x_{\min})\right]\right)\Big|_{x_{\min}}^{\tilde{x}} \\
\tilde{x} &= x_1 + \sigma_1 \ln\frac{1 - 2\tilde{s} - \sqrt{1 - 4\tilde{s}}}{2\tilde{s}} \\
\tilde{s} &= c^{-1}\ln(\ln 2/\nu t A_1) \\
J_2 &= -\,(c\alpha)^{-1}\mathrm{Ei}_1\left(\nu t A_2 \exp\left(\alpha\left[1/4 + \beta(x - x_1)\right]\right)\right)\Big|_{x_1}^{x_2} \\
J_4 &= -\,\alpha^{-1}\exp\left[-\nu t I_3(x_1)\right]\mathrm{Ei}_1\left(\nu t A_3 \exp\left[-\alpha(x - x_0)\right]\right)\Big|_{x_3}^{x_4}.
\end{aligned} \tag{5.9}$$

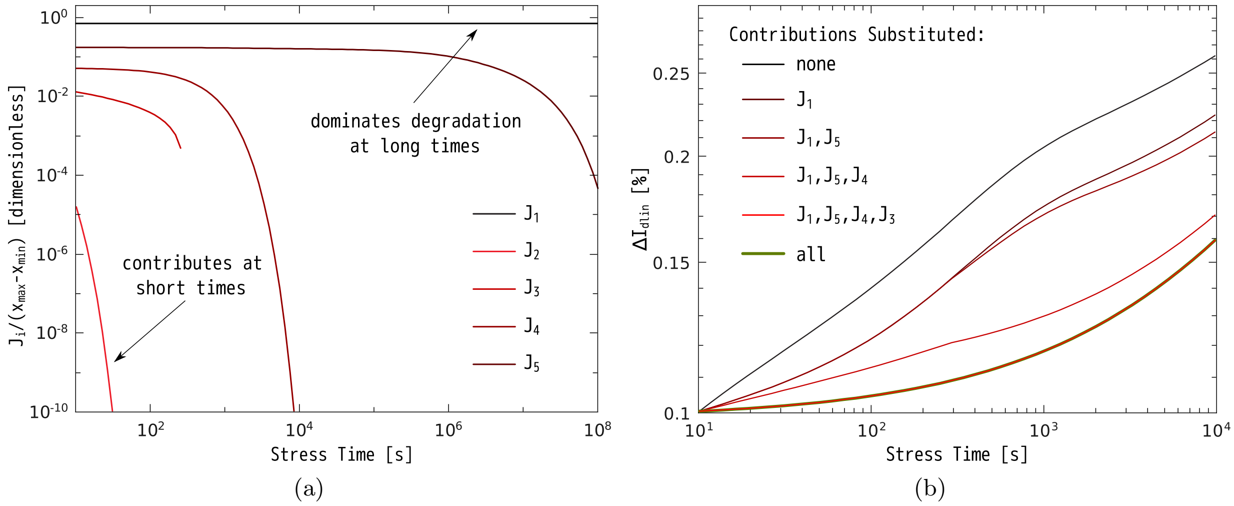


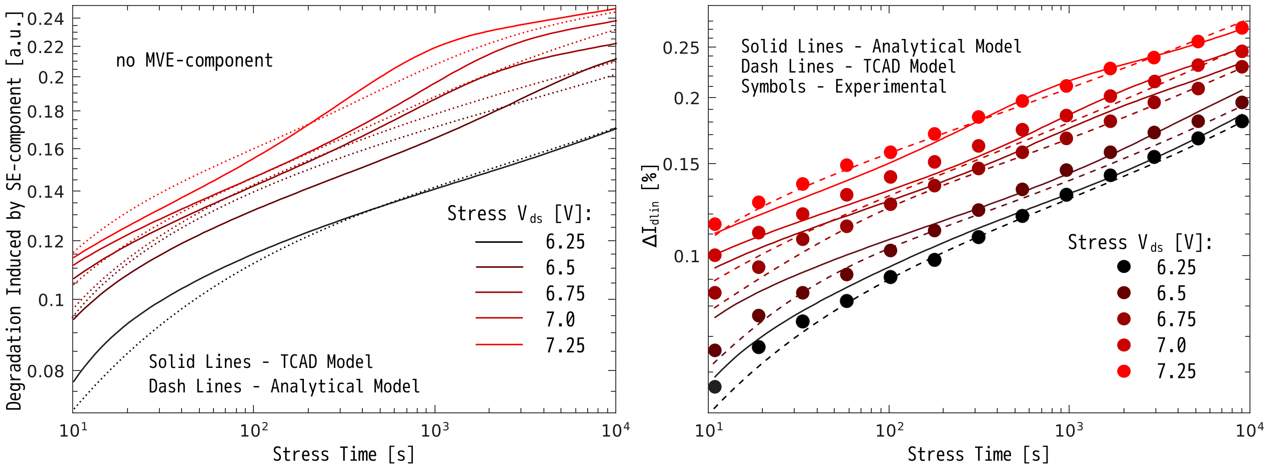
Figure 5.4: (a) Different contributions to the degradation J_i divided by the gate length for V_{ds} =7.25V. (b) I_{dlin} degradation calculated by substituting one (or some) contribution(s) J_{i} by the constant value (V_{ds} =7.25V).

Figure 5.4a demonstrates the time dependences of the integrals J_{i} calculated for V_{ds} = 7.25V. The terms J_{i} reflect the fraction of "virgin" bonds and a decay in J_{i} means that the dissociation process was triggered during this time period. It is evident that the "tip" of the AI peak (the term I_2 of the AI and the J_2 in ΔI_{dlin}) affects I_{dlin} rather weakly already at t = 10s. In fact, I_2 varies approximately in the range of $10^{34} - 10^{35}$a.u. (Figure 5.2a) and the criterion $\nu t I_1 \sim 1$ leads to bond-breakage events predominately triggered within the time period of 0.2s to 2s. In contrast, I_1 varies by more than 11 orders of magnitude and thus J_1 contributes to ΔI_{dlin} over several decades and defines the degradation at long times. These considerations are summarized in Figure 5.4b where the time-dependent values of some J_{i} terms are substituted with constants. For instance, the term J_1 does not substantially change during several decades (Figure 5.4a) and thus may be represented by the value $J_1(t = 0)$. Figure 5.4b shows that during the time period of 0..10^4s terms $\{J_2, J_3, J_5\}$ are not significant and may be omitted.

Since in the experiment it is impossible to separate the SP- and MP-components, one employs the TCAD model to show the I_{dlin} degradation induced solely by the SP-mechanism. This step is important because it allows for the verification of the adequacy of the acceleration integral profile representation with a fitting formula. Figure 5.5a shows good agreement between the results calculated by TCAD and analytical models. The major discrepancy is observed at short times and is related to the noise seen in the AI peak and the related stochastic nature of the Monte-Carlo method.

Being sure that the SP-component is reasonably represented with the analytical expression the SP- and MP-contributions were combined. ΔI_{dlin} considering also the MP-component is compared with experimental data and plotted in Figure 5.5b, which demonstrates a very good agreement between simulation and experiment.

This model based on the analytical representation of the AI does not employ time consuming MC simulations required to calculate the set of carrier DFs. One may profit from such an



(a) (b)

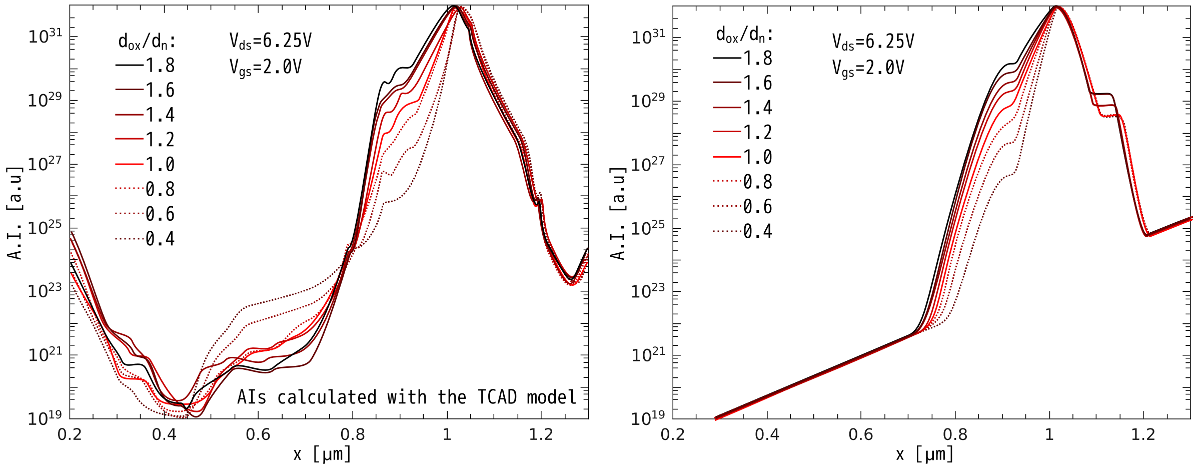
Figure 5.5: (a) Comparison between the ΔI_{dlin} portion induced by the SP-component obtained within the TCAD model and the analytical approach. (b) I_{dlin} degradation: experiments vs. simulations. Very good agreement between the experimental data and the results of both TCAD and analytical model is obtained.

advantage while considering the impact of statistical variations in device geometry (e.g. in the oxide thickness, gate length, etc) on HCD. If one should recalculate the entire set of carrier DFs with the MC method at each value of the fluctuating parameter, the simulation process will be inadmissibly time expensive. In contrast, within the developed model, the DFs are calculated only once for several reference values of the fluctuating quantity and the fitting parameters of the model (controlling the AI) are defined by interpolation using dependences such as those in Figure 5.3. Afterwards, one may use the analytical form of the AI and operate with simple expressions, which can be further weighted considering the distribution of the fluctuating quantity and statistically processed.

5.2 Impact of Gate Oxide Thickness Variations on HCD

The shape of the coordinate profile $N_{\mathrm{it}}(x)$ is defined by the device architecture. As such, the topological features such as the doping profiles, oxide thickness, gate extension, etc. can strongly affect the device life-time under hot-carrier stress. At the same time, these architectural peculiarities vary from device to device even in the case when these devices are fabricated within the same technological cycle [1]. As a result, it is expected that such fluctuations can impact HCD. In this Section of the work, the developed analytical approach to hot-carrier degradation modeling described in the previous Section is employed in order to analyze the impact of the oxide thickness (t_{ox}) variation on hot-carrier degradation.

A “brut-force” approach to analyze the impact of device architecture variations on the behaviour of the transistor degraded during hot-carrier stress may be performed in the following manner. At the initial step a set of topologically identical devices differing only in the value of the fluctuating parameter is generated. Then, for each particular device (with the unique value of the varying parameters) the entire described procedure is applied. The results are then



(a)

(b)

Figure 5.6: The AI profiles calculated with (a) the calibrated TCAD model and with (b) the analytical model for different oxide thicknesses (except t_{ox}, the device topology is identical).

binned into histograms and statistically processed, weighting the statistical distribution of the fluctuating quantity.

Due to the stochastic nature of the MC approach, such a computational procedure, however, would lead to an enormous huge computational burden. And even more dramatic, if we vary this geometrical parameter with a reasonably small step, we should pay especial attention to the calculation accuracy. Otherwise, no prominent difference between results calculated for fairly similar (but still not identical) devices will be obtained. To summarize, all these circumstance make this “brut-force” strategy practically unrealistic.

Therefore, an analytical approach is used, which is calibrated using the results obtained with the TCAD version of the HCD model. We vary an architectural parameter (in this work the oxide thickness) and for some reference values we launch the entire routine of the TCAD model to obtain the exact dependences $\Delta I_{\mathrm{dlin}}(t)$. Then, we interpolate parameters in the analytical formula for the acceleration integral in order to cover the entire range of the fluctuating parameter.

A 5V n-type MOSFET fabricated on a standard 0.35μm process with the nominal oxide thickness $t_{\mathrm{n}} = 14.83$nm is used. That is, the model was calibrated to describe the linear drain current change under hot-carrier stress for the device with oxide thickness t_{n}. One varied t_{ox} in a wide range from $0.4t_{\mathrm{n}}$ to $1.8t_{\mathrm{n}}$. Afterwards, for the reference values of t_{ox} ($t_{\mathrm{ox}}/t_{\mathrm{n}}$ was 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, and 1.8) the complete simulation scheme (including the Monte-Carlo step) is performed and the AI profiles are calculated, see Figure 5.6a (stress voltages were $V_{\mathrm{ds}} = 6.25$V and $V_{\mathrm{gs}} = 2.0$V). Figure 5.6b demonstrates the series of AIs calculated for the varying t_{ox} using the analytical formula. A direct comparison between acceleration integrals for particular thicknesses of $1.8t_{\mathrm{n}}$ and $1.0t_{\mathrm{n}}$ is presented in Figure 5.7. From Figure 5.6 it can be concluded that the slope of the right front of AI (represented by fragments I_3, I_5, see Figure 5.2b) remains the same (slight deviations in the shape of I_3, I_5 and the ledge I_4 boundaries are related to the statistical noise within the Monte-Carlo approach). Therefore, the only varying parameters are the height of I_1 (parameter A_1) and the slope of I_2 (parameter β). Their dependences are shown

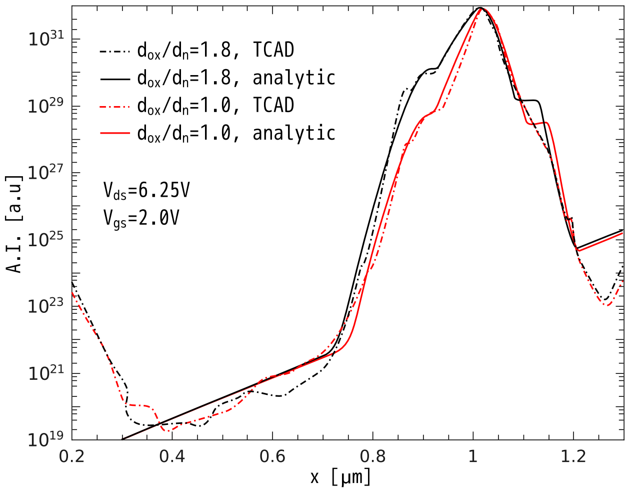


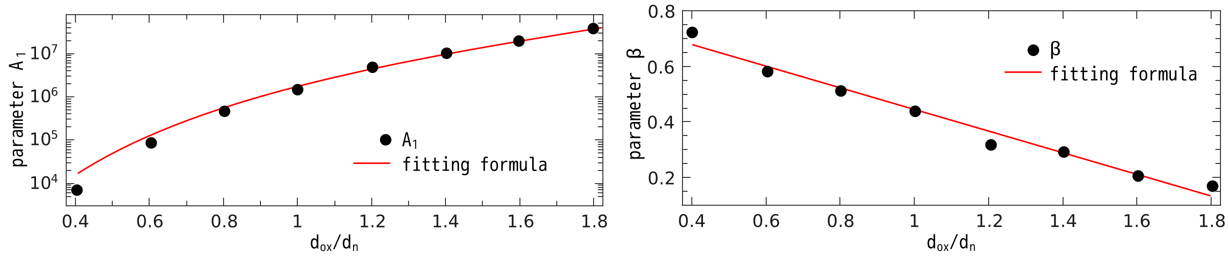
Figure 5.7: Comparison between AIs calculated with the TCAD model and the analytical expression for two different oxide thicknesses.

in Figure 5.8 and represented by fitting formulae

$$\begin{aligned} A_1 &= A_{1,0}\left(\frac{t_{\mathrm{ox}}}{t_{\mathrm{n}}}\right)^b; \quad A_{1,0} = 1.85\times 10^6; \quad b = 4.73 \\ \beta &= \beta_0\frac{t_{\mathrm{ox}}}{t_{\mathrm{n}}} + \beta_1; \qquad \beta_0 = -0.33; \qquad \beta_1 = 0.78. \end{aligned} \tag{5.10}$$

This expression is further used for statistical analysis of ΔI_{dlin} in the case of a fluctuating t_{ox}.

Using acceleration integrals from Figure 5.6b, a series of reference curves ΔI_{dlin} was calculated for different values of t_{ox}, Figure 5.9. One can see that $\Delta I_{\mathrm{dlin}}(t)$ substantially changes with varying t_{ox}. This change – especially pronounced for moderate stress times, i.e. less than 10^3s – tends to saturate at 10^4s. For all gate oxide thicknesses the peak of the acceleration integral is pronounced at the same position and the fragment varying with t_{ox} is the left front of the AI. This varying front defines the difference between $\Delta I_{\mathrm{dlin}}(t)$ at short stress times calculated for the devices with different t_{ox}. For the single-carrier process, the bond-breakage events are triggered as described by the activation exponent, i.e. $N_{\mathrm{it}}(x,t) = N_0(1\exp[-I(x)\nu t])$, where ν is the attempt rate. Therefore, at large stress times all the bonds in the vicinity of the AI peak are predominately broken, independently of the particular oxide thickness. As a result,



(a) (b)

Figure 5.8: Dependences of parameters A_1 (a) and β (b) on the relative oxide thickness $t_{\mathrm{ox}}/t_{\mathrm{n}}$.

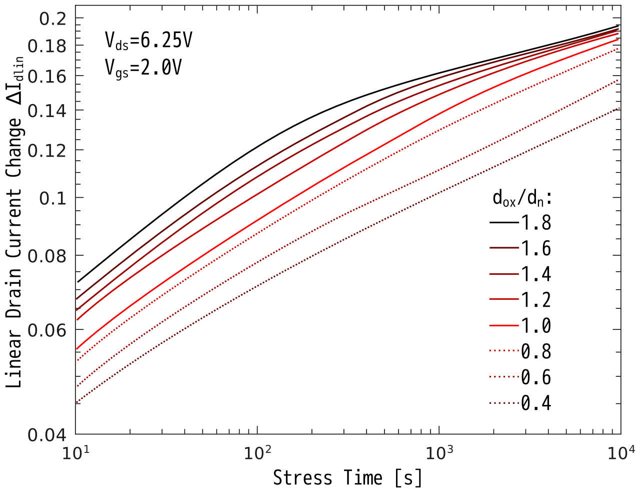


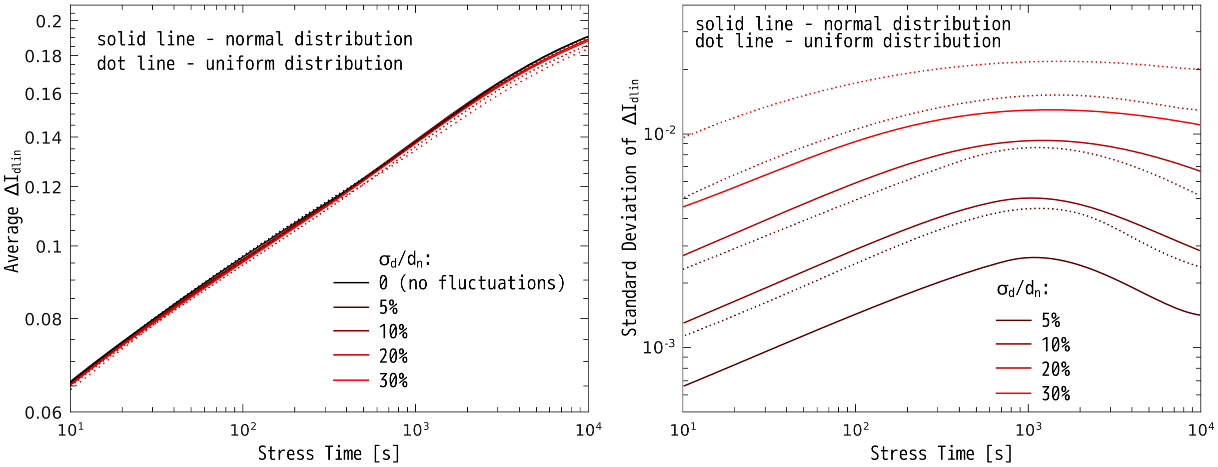
Figure 5.9: Dependences $\Delta I_{\mathrm{dlin}}(t)$ calculated for different values of t_{ox}.

the difference between $\Delta I_{\mathrm{dlin}}(t)$ curves tends to vanish at $t \sim 10^4$s. Since the parameter t_{ox} may fluctuate, one should consider the mean value of ΔI_{dlin} and its standard deviation as functions of time

$$\begin{aligned} \langle \Delta I_{\mathrm{dlin}} \rangle (t) &= \int \Delta I_{\mathrm{dlin}}(t_{\mathrm{ox}}, t) D(t_{\mathrm{ox}}, t_{\mathrm{n}}, \sigma_{\mathrm{d}}) \mathrm{d}t_{\mathrm{ox}} \\ \sigma^2_{\Delta I_{\mathrm{dlin}}}(t) &= \int [\Delta I_{\mathrm{dlin}}(t_{\mathrm{ox}}, t) - \Delta I_{\mathrm{dlin}}(t_{\mathrm{n}}, t)]^2 D(t_{\mathrm{ox}}, t_{\mathrm{n}}, \sigma_{\mathrm{d}}) \mathrm{d}t_{\mathrm{ox}}, \end{aligned} \tag{5.11}$$

where D is the distribution of the oxide thickness and σ_{d} the standard deviation. As an example, these dependences were evaluated for situations when t_{ox} obeys Gaussian and uniform distributions. In the first case the standard deviation of the oxide thickness was $\sigma_{\mathrm{d}} = 5\%$, 10%, 20%, and 30% of t_{n}. In the second case it is assumed that t_{ox} is homogeneously distributed in the interval of $[t_{\mathrm{ox}} - 3\sigma_{\mathrm{d}};\ t_{\mathrm{ox}} + 3\sigma_{\mathrm{d}}]$ and σ_{d} has the same values as in the first case (in the case of a uniform distribution with such a span, the standard deviation, differs from σ_{d}). The results are presented in Figure 5.10. Notwithstanding the fact that $\Delta I_{\mathrm{dlin}}(t)$ is very sensitive to t_{ox} variations (Figure 5.9) the mean value $\langle \Delta I_{\mathrm{dlin}} \rangle (t)$ substantially differs from the nominal one (i.e. calculated for the fixed value $t_{\mathrm{ox}} = t_{\mathrm{n}}$) only in the case of a wide uniform distribution, see Figure 5.10 demonstrating rather low $\Delta I_{\mathrm{dlin}}(t)$ dispersions at all stress times also confirms this trend.

Using the TCAD version of our physics-based model of hot-carrier degradation a series of carrier acceleration integrals for devices with identical architecture but with different oxide thicknesses was calculated. Then, the information describing how the acceleration integral changes with t_{ox} was incorporated into analytical approach to HCD modeling. It was found that only two parameters in the AI analytical formula change with t_{ox}. Their dependences on $t_{\mathrm{ox}}/t_{\mathrm{n}}$ have been represented by fitting expressions covering the entire range of $t_{\mathrm{ox}}/t_{\mathrm{n}}$ variations. This fitting allow us to avoid the time consuming Monte-Carlo simulations while studying HCD in devices with different t_{ox}. Moreover, following the suggested approach it is possible to analyze the impact of oxide thickness variations on the linear drain current change vs. time during hot-carrier stress. The calculations have demonstrated that $\Delta I_{\mathrm{dlin}}(t)$ is rather sensitive to an oxide thickness change at short and moderate stress times. For longer stress times, the difference between $\Delta I_{\mathrm{dlin}}(t)$ obtained for different oxide thicknesses tends to vanish. This is because at



(a) (b)

Figure 5.10: (a) The mean value of $I_{\mathrm{dlin}}(t)$ vs. t calculated for different σ_{d}. (b) The standard deviation of $\Delta I_{\mathrm{dlin}}(t)$ calculated for different σ_{d}.

long times all Si – H bonds in the vicinity of the acceleration integral peak are predominately broken independently of t_{ox}. As an example, the mean value and the standard deviation of $\Delta I_{\mathrm{dlin}}(t)$ for different t_{ox} distributions were also calculated.

6 Summary and Outlook

A comprehensive physics-based model for hot-carrier degradation must consider three different aspects: carrier transport, microscopic description of the defect build-up, and the impact of the generated defects on the device performance. While exchanging the information between these blocks one should perform a thorough check at each stage.

A direct analysis of the interface state density profile has been carried out, assessed within the developed model and that extracted from charge-pumping measurements. The approach for the modeling of the interface state density profile relies on the thorough evaluation of the carrier energy distribution function by means of a full-band Monte-Carlo device simulator. The interplay between "hot" and "colder" carriers or – in other words – the interplay between the single-particle (SP) and multiple-particle (MP) component of the $\mathrm{Si-H}$ bond-breakage is considered. The importance of the multiple-particle component, even in the case of long-channel devices and relatively high stress voltages, was demonstrated.

The carrier acceleration integrals for the SP and MP processes are characterized by rather different profiles. The SP-component is strongly localized near the drain end of the gate electrode while the MP-component is more or less homogeneously distributed. The latter means that under high stress conditions the MP-mechanism is still triggered. It must be linked with the low-energy part of the distribution function, which is populated under each stress conditions.

Note that in order to calculate the SP-related acceleration integral the parameter set of Bravaix *et al.* [30, 31] is used while for the MP-component in the Keldysh-like reaction cross section the exponent of 0.1 (i.e. process is only slightly sensitive to the carrier energy variations) is used. Such a small value may be attributed to the dispersion of the parameters of the $\mathrm{Si-H}$ bond energetics repeatedly reported in the literature [24, 40]. This question deserves more attention and has to be clarified in the future. At the moment the transport module of the approach is represented by the stochastic solver MONJU (Figure 1.12), which evaluates the carrier energy distribution function for a particular device topology. Due to a complicated transport module realization and the necessity of device conversion, additional computational noise can be seen at distribution function DF simulation. Namely, consideration of different device oxide thicknesses (real and flatted) results in a shift of the AI peak as far for electrons as for holes (see Figure 6.1). To be precise, if during simulations the oxide thickness is assumed uniform ($t_{\mathrm{ox}}(x)=t_{\mathrm{ox}}$), both AI peaks are shifted which implicitly confirms that the device architecture peculiarities must be taken into account as accurately as possible.

The thorough calculation of the DF is a time-consuming task because the key information is provided by high-energy tails of the DF which are lowly-populated. As a result, the set of distribution functions (at each position along the $\mathrm{Si/SiO_2}$ interface) for certain stress conditions is calculated only once and then used throughout the entire degradation time period. In reality,

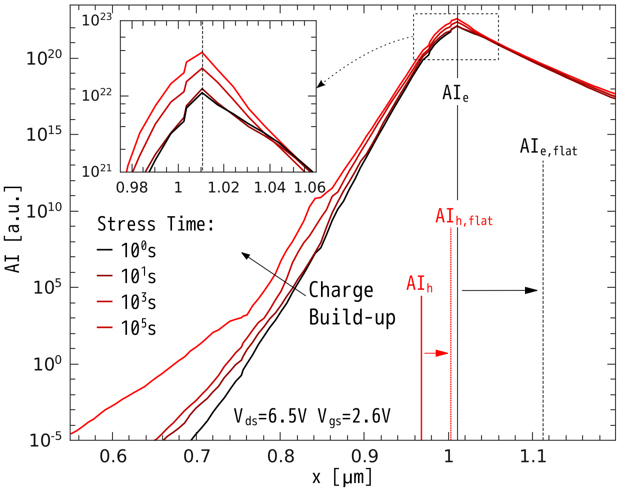


Figure 6.1: The evolution of electron AI with stress time for V_{ds}=6.5V and V_{gs} =2.6V. The position of the $N_{\mathrm{it}}(x)$ peaks does not change with the stress time. The assumption of a constant oxide thickness results in a significant shift of the AI peak for holes and electrons ($\mathrm{AI}_{\mathrm{e(h)}} \rightarrow \mathrm{AI}_{\mathrm{e(h),flat}}$).

the DF is also distorted during the stress. For instance, the interface states – if charged – act as additional scattering centers, thereby changing the shape of the DF. Thus the DF should be recalculated at each time step. To additionally check this fact the acceleration integrals are self-consistently calculated while considering $N_{\mathrm{it}}(x)$. Figure 6.1 shows the AI vs. x and confirms the described tendency. If the position of the AI peaks does not change with time (which is in good agreement with the present model predictions where the position of the carrier acceleration integral maximum coincides with the $N_{\mathrm{it}}(x)$ peak position), the AI distribution at the middle of the channel significantly increases. However, such a consistent consideration of DF dramatically impacts the computational time and its realization within the framework with the Monte-Carlo device simulator as a transport module is practically impossible.

A potential remedy of this problem uses the flexible BTE solver ViennaSHE. In contrast to a stochastic BTE solver, which requires long computation times, a deterministic approach usually needs a high amount of memory. Certain efforts to overcome this drawback have already been performed at our Institute and realized in the device simulator ViennaSHE.

The fast BTE solver ViennaSHE will in future be used as the kernel of a comprehensive HCD model. This platform will allow for a consistent consideration of microscopic models for defect generation and carrier transport. For instance the impact of the charged traps on the carrier DF will be self-consistently incorporated. Moreover, due to the low computational costs, 3D device simulations will be possible. This is important not only for the modeling of transistors with 3D architecture but also to capture self-heating effects, which play a significant role e.g. in the context of high-voltage devices.

Another refinement is associated with the interplay between interface, border and bulk traps. The calibration of the model is expected to allow us to resolve the spatial profiles of border/bulk traps (not only vs. x but also along the coordinate inside the oxide film) as well as their energy distributions. For this purpose it is suggested to focus on HV devices where, as expected, the bulk oxide charge plays a significant role. Note that either in the case of high-voltage transistors or while interpreting results of charge-pumping measurements all types of states – i.e. interface, border and bulk – are to be considered consistently.

To conclude, the following factors must be covered for a proper modeling of hot-carrier degradation phenomena:

- Track the correlation between N_{it} and $N_{\mathrm{bt}}/N_{\mathrm{ot}}$.
- Deduce the impact of the gate oxide thickness on trapping in the oxide bulk. The profiling of $N_{\mathrm{bt}}/N_{\mathrm{ot}}$ means not only their dependences on the lateral coordinate but also on the coordinate inside the oxide film.
- Analyze of mechanisms for the generation of new bulk traps during the HC stress.
- Link the $N_{\mathrm{it}}/N_{\mathrm{bt}}/N_{\mathrm{ot}}$ generation and stress-induced induced leakage current.

In addition to the profiling of N_{it} and N_{ot}, the shape of trap density-of states should be also obtained. This information will help us resolve the question whether the SP- and MP-mechanisms really produce two types of states with different DOS. That is, the consistent consideration of SP- and MP-components or their individual treatment should be chosen.

Bibliography

[1] *International Technology Roadmap for Semiconductors (ITRS)*, 2011.

[2] A. Acovic, G. L. Rosa, and Y. Sun, "A review of hot carrier degradation mechanism in MOSFETs," *Microel. Reliab.*, vol. 36, no. 7/8, pp. 845–869, 1996.

[3] T. Ning, P. Cook, R. Dennard, C. Osburn, S. Schuster, and H. Yu, "1 μm most VLSI technology – part IV: hot-electron design constraints," *IEEE Trans. Electron Dev.*, vol. 26, pp. 346–353, 1979.

[4] E. Takeda, "Hot-carrier effects in submicrometer MOS VLSIs," *IEEE Proc.*, vol. 131, no. 5, pp. 153–162, 1984.

[5] C. Hu, S. Tam, F. Hsu, P.-K. Ko, T.-Y. Chan, and K. Terrill, "Hot-electron-induced MOSFET degradation model, monitor and improvement," *IEEE Trans. Electron Dev.*, vol. 48, no. 4, pp. 375–385, 1985.

[6] P. Heremans, P. Bellens, G. Groeseneken, and H. Maes, "Consistent model for the hot carrier degradation in n-channel and p-channel MOSFETs," *IEEE Trans. Electron Dev.*, vol. 35, no. 12, pp. 2194–2209, 1988.

[7] C. Hu, "Lucky electron model for channel hot electron emission," in *Proc. International Electron Devices Meeting (IEDM)*, pp. 22–25, 1979.

[8] S. Rauch, F. Guarin, and G. La Rosa, "Impact of e-e scattering to the hot carrier degradation of deep submicron nMOSFETs," *IEEE Electron Dev. Lett.*, vol. 19, no. 12, pp. 463–465, 1998.

[9] E. Takeda and N. Suzuki, "An empirical model for device degradation due to hot-carrier injection," *IEEE Electron Dev. Lett.*, vol. 4, no. 5, pp. 111–113, 1983.

[10] J.-S. Goo, Y.-G. Kim, H. Lee, H.-Y. Kwon, and H. Shin, "An analytical model for hot-carrier-induced degradation of deep-submicron n-channel LDD MOSFETs," *Solid-State Electron.*, vol. 38, no. 6, pp. 1191–1196, 1995.

[11] R. Dreesen, K. Croes, J. Manca, W. D. Ceunick, L. D. Schepper, A. Pergoot, and G. Groeseneken, "Modeling hot-carrier degradation of LDD nMOSFETs by using a high resolution measurement technique," *Microel. Reliab.*, vol. 39, pp. 785–790, 1999.

[12] R. Dreesen, K. Croes, J. Manca, W. D. Ceunick, L. D. Schepper, A. Pergoot, and G. Groeseneken, "A new degradation model and lifetime extrapolation technique for lightly doped drain nMOSFETs under hot-carrier degradation," *Microel. Reliab.*, vol. 41, pp. 437–443, 2001.

[13] R. Woltjer and G. Paulzen, "Universal description of hot-carrier-induced interface states in nMOSFETs," in *Proc. International Electron Devices Meeting (IEDM)*, pp. 535–538, 1992.

[14] R. Woltjer, G. Paulzen, H. Pomp, H. Lifka, and P. Woerlee, "Three hot-carrier degradation mechanisms," *IEEE Trans. Electron Dev.*, vol. 42, no. 1, pp. 109–115, 1995.

[15] K. Mistry and B. Doyle, "A model for AC hot-carrier degradation in n-channel MOSFETs," *IEEE Electron Dev. Lett.*, vol. 12, no. 9, pp. 492–494, 1991.

[16] K. Mistry and B. Doyle, "AC versus DC hot-carrier degradation in n-channel MOSFETs," *IEEE Trans. Electron Dev.*, vol. 40, no. 1, pp. 96–104, 1993.

[17] S. Rauch, G. La Rosa, and F. Guarin, "Role of e-e scattering in the enhancement of channel hot carrier degradation of deep-submicron nMOSFETs at high V_{gs} conditions," *IEEE Trans Dev. Material. Reliab.*, vol. 1, no. 2, pp. 113–119, 2001.

[18] C. Guerin, V. Huard, and A. Bravaix, "The energy-driven hot-carrier degradation modes of nMOSFETs," *IEEE Trans. Dev. Material. Reliab.*, vol. 7, no. 2, pp. 225–235, 2007.

[19] P. Moens and G. van den Bosch, "Characterization of total self-operating area of lateral DMOS transistors," *IEEE Trans. Electron Dev.*, vol. 6, no. 3, pp. 349–357, 2006.

[20] P. Moens, G. van den Bosch, and G. Groeseneken, "Competing hot carrier degradation mechanisms in lateral n-type DMOS transistors," in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 214–221, 2003.

[21] P. Moens and M. Tack, "Hole trapping and de-trapping effects in LDMOS devices under dynamic stress," in *Proc. International Electron Devices Meeting (IEDM)*, 2006.

[22] P. Moens, F. Bauwens, M. Nelson, and M. Tack, "Electron trapping and interface trap generation in drain extended pMOS transistors," in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 93–96, 2005.

[23] W. McMahon, A. Haggaag, and K. Hess, "Reliability scaling issues for nanoscale devices," *IEEE Trans. Nanotech.*, vol. 2, no. 1, pp. 33–38, 2003.

[24] K. Hess, A. Haggag, W. McMahon, K. Cheng, J. Lee, and J. Lyding, "The physics of determining chip reliability," *Circuits and Devices Mag.*, pp. 33–38, 2001.

[25] O. Penzin, A. Haggag, W. McMahon, E. Lyumkis, and K. Hess, "MOSFET degradation kinetics and its simulation," *IEEE Trans. Electron Dev.*, vol. 50, no. 6, pp. 1445–1450, 2003.

[26] H. Kufluoglu and M. Alam, "A computational model of NBTI and hot carrier injection time-exponents for MOSFET reliability," *Journ. Comput. Electron.*, vol. 3, pp. 165–169, 2004.

[27] H. Kufluoglu, *MOSFET degradation due to negative bias temperature instability (NBTI) and hot carrier degradation (HCI) and its applications for reliability-aware VLSI design.* PhD thesis, Purdue University, West Lafayette, Indiana, USA, 2007.

[28] S. Rauch and G. L. Rosa, "The energy driven paradigm of nMOSFET hot carrier effects," in *Proc. International Reliability Physics Symposium (IRPS)*, 2005.

[29] S. Rauch and G. L. Rosa, "CMOS hot carrier: From physics to end of life projections, and qualification," in *Proc. International Reliability Physics Symposium (IRPS), tutorial*, 2010.

[30] A. Bravaix, C. Guerin, V. Huard, D. Roy, J. Roux, and E. Vincent, "Hot-carrier acceleration factors for low power management in DC-AC stressed 40nm nMOS node at high temperature," in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 531–546, 2009.

[31] C. Guerin, V. Huard, and A. Bravaix, "General framework about defect creation at the Si/SiO$_2$ interface," *Journ. Appl. Phys.*, vol. 105, no. 114513, 2009.

[32] W. McMahon, K. Matsuda, J. Lee, K. Hess, and J. Lyding, "The effects of a multiple carrier model of interface states generation of lifetime extraction for MOSFETs," in *Proc. Int. Conf. Mod. Sim. Micro*, vol. 1, pp. 576–579, 2002.

[33] K. Hess, I. C. Kizilyalli, and J. W. Lyding, "Giant isotope effect in hot electron degradation of metal oxide silicon devices," *IEEE Trans Electron Dev.*, vol. 45, no. 2, pp. 406–416, 1998.

[34] T. Grasser, W. Gös, and B. Kaczer, "Critical modeling issues in negative bias temperature instability," *ECS Transactions*, vol. 19, no. 2, pp. 265–287, 2009.

[35] W. Chang, B. Davari, M. Wordeman, Y. Taur, C. Hsu, and M. Rodriguez, "A high performance 0.25μm CMOS technology. I. design and characterization," *IEEE Trans. Electron Dev.*, vol. 39, p. 959, 1992.

[36] D. Bursky, "ASIC family crams up to 1.2mum usable gates/chip," *Electronic Design*, vol. 41, pp. 111–116, 1993.

[37] D. Frank, R. Dennard, E. Nowak, P. Solomon, M. Stettler, and M. Bohr, "Scaling challenges and device design requirements for high performance sub-50 nm gate length planar CMOS transistors," in *Proc. VLSI Symposium Tech. Digest*, pp. 174–175, 2000.

[38] L. Hong, *Characterization of hot carrier reliability in deep submicrometer MOSFETs*. PhD thesis, National University of Singapore, 2005.

[39] F.-C. Hsu and K.-Y. Chu, "Evaluation of LDD MOSFET's based on hot-electron-induced degradation," *IEEE Electron Dev. Lett.*, vol. 5, no. 5, pp. 162–165, 1984.

[40] A. Haggag, W. McMahon, K. Hess, K. Cheng, J. Lee, and J. Lyding, "High-performance chip reliability from short-time-tests. statistical models for optical interconnect and HCI/TDDB/NBTI deep-submicron transistor failures," in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 271–279, 2001.

[41] A. Bravaix and V. Huard, "Hot-carrier degradation issues in advanced CMOS nodes," in *Proc. European Symposium on Reliability of Electron Devices Failure Physics and Analysis (ESREF), tutorial*, 2010.

[42] T. Mizuno, A. Toriumi, M. Iwase, M. Takanashi, H. Niiyama, M. Fukmoto, and M. Yoshimi, "Hot-carrier effects in 0.1 μm gate length CMOS devices," in *Proc. International Electron Devices Meeting (IEDM)*, pp. 695–698, 1992.

[43] J. Bude, "Gate-current by impact ionization feedback in submicron MOSFET technologies," in *Proc. VLSI Symposium Tech. Digest*, pp. 101–102, 1995.

[44] F. Venturi, E. Sangiorgi, and B. Ricco, "The impact of voltage scaling on electron heating and device performance of submicrometer MOSFET's," *IEEE Trans. Electron Dev.*, vol. 38, no. 8, pp. 1895–1904, 1991.

[45] J. Chung, M. Jeng, J. Moon, P. Ko, and C. Hu, "Low-voltage hot-electron currents and degradation in deep-submicrometer MOSFETs," *IEEE Trans. Electron Dev.*, vol. 37, pp. 1651–1657, 1990.

[46] M. Fischetti and S. Laux, "Monte-Carlo study of sub-band-gap impact ionization in small silicon field-effect transistors," in *Proc. International Electron Devices Meeting (IEDM)*, pp. 305–308, 1995.

[47] D. Brisbin, P. Lindorfer, and P. Chaparala, "Substrate current independent hot carrier degradation in nLDMOS devices," in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 329–333, 2006.

[48] M. Annese, S. Carniello, and S.Manzini, "Design and optimization of a hot-carrier resistant high-voltage nMOS transistor," *IEEE Trans. Electron Dev.*, vol. 52, no. 7, pp. 1634–1639, 2005.

[49] P. Santos, H. Quaresma, A. Silva, and M. Lanca, "High-voltage nMOS design in fully implanted twin-well CMOS," *Microel. Jour.*, vol. 35, no. 9, pp. 723–730, 2004.

[50] W. Qin, W. Chim, D. H. Chan, and C. Lou, "Modelling the degradation in the subthreshold characteristics of submicrometre LDD pMOSFETs under hot-carrier stressing," *Semicond. Sci. Technol.*, vol. 13, no. 5, pp. 453–459, 1998.

[51] S. Manzini and A. Gallerano, "Avalanche injection of hot holes in the gate oxide of LDMOS," *Solid-State Electron.*, vol. 44, no. 1, pp. 1325–1330, 2000.

[52] V. Reddy, "An introduction to CMOS semiconductor reliability," in *Proc. International Reliability Physics Symposium (IRPS), tutorial*, 2004.

[53] P. Moens, G. van den Bosch, C. D. Keukeleire, R. Degraeve, M. Tack, and G. Groseneken, "Hot hole degradation effects in lateral nDMOS transistors," *IEEE Trans. Electron Dev.*, vol. 51, no. 10, pp. 1704–1710, 2004.

[54] D. Brisbin, A. Strachan, and P. Chaparala, "3-d and 2-d hot carrier layout optimization of n-LDMOS transistor arrays," in *Proc. International Integrated Reliability Workshop (IIRW)*, pp. 120–124, 2002.

[55] S. Manzini and C. Contiero, "Hot-electron-induced degradation in high-voltage submicron DMOS transistors," in *Proc. International Symposium on Power semiconductor Devices and ICs*, pp. 65–68, 1996.

[56] Q. Wang, L. Sun, and A. Yap, "Investigation of hot carrier degradation in asymmetric nDEMOS transistors," *Microel. Reliab.*, vol. 48, pp. 508–513, 2008.

[57] Q. Wang, L. Sun, Z. Zhang, A. Yap, H. Li, and S. Liu, "Locating hot carrier degradation in asymmetric nDEMOS transistors by gated diode technique," *Journ. Non-Crystalline Solids*, vol. 354, pp. 1871–1875, 2008.

[58] K.-M. Wu, J. Chen, Y. Su, J. Lee, K. Lin, J. Shih, and S. Hsu, "Effects of gate bias on hot-carrier reliability in drain extended metal-oxide-semiconductor transistors," *Appl. Phys. Lett.*, vol. 89, no. 183522, 2006.

[59] J. Lee, J. Chen, K.-M. Wu, C. Liu, and S. Hsu, "Effect of hot-carrier-induced interface states distribution on linear drain current degradation in 0.35 μm n-type lateral diffused metal-oxide-semiconductor transistors," *Appl. Phys. Lett.*, vol. 92, no. 103510, 2008.

[60] J. Chen, S.-Y. Chen, K.-M. Wu, and C. Liu, "Investigation of hot-carrier-induced degradation mechanisms in p-type high-voltage drain extended metal-oxide-semiconductor transistors," *Jpn. Journ. Appl. Phys.*, vol. 48, no. 04C039, 2009.

[61] J. Chen, S.-Y. Chen, K.-M. Wu, and C. Liu, "Channel length dependence of hot-carrier-induced degradation in n-type drain extended metal-oxide-semiconductor transistors," *Appl. Phys. Lett.*, vol. 93, no. 223504, 2008.

[62] J. Chen, K.-S. Tian, S.-Y. Chen, K.-M. Wu, and C. Liu, "Mechanism and modeling of on-resistance degradation in n-type lateral diffused metal-oxide-semiconductor transistors," *Jpn. Journ. Appl. Phys.*, vol. 48, no. 04C040, 2009.

[63] M. Pagey, "Characterization and modeling of hot-carrier degradation in sub-micron nMOSFETs," Master's thesis, Vanderbilt University, Aug. 2002.

[64] *International Technology Roadmap for Semiconductors (ITRS)*, 2009.

[65] M. Ancona, N. Saks, and D. McCarthy, "Lateral disrtribution of hot-carrier-induced interface traps in MOSFETs," *IEEE Trans Electron Dev.*, vol. 35, no. 12, pp. 221–2228, 1988.

[66] S. Mahapatra, C. Parikh, V. Rao, C. Viswanathan, and J. Vasi, "Device scaling effects on hot-carrier induced interface and oxide-trapped charge distributions in MOSFETs," *IEEE Trans. Electron Dev.*, vol. 47, no. 4, pp. 789–796, 2000.

[67] S. Cristoloveanu, H. Haddara, and N. Revil, "Defect localization induced by hot carrier injection in short-channel MOSFETs: concept, modeling and characterization," *Microel. Reliab.*, vol. 33, no. 9, pp. 1365–1385, 1993.

[68] F.-C. Hsu and K.-Y. Chu, "Temperature dependence of hot-electron induced degradation in MOSFETs," *IEEE Electron Dev. Lett.*, vol. 5, no. 5, pp. 148–150, 1984.

[69] P. Heremans, G. V. den Bosch, R. Bellens, G. Groseneken, and H. Maes, "Temperature dependence of the channel hot-carrier degradation of n-channel MOSFETs," *IEEE Trans. Electron Dev.*, vol. 37, no. 4, pp. 980–992, 1990.

[70] M. Song, K. MacWilliams, and C. Woo, “Comparison of nMOS and pMOS hot carrier effects from 300 to 77 K,” *IEEE Trans. Electron Dev.*, vol. 44, no. 2, pp. 268–276, 1997.

[71] A. Bravaix, D. Goguenheim, N. Revil, E. Vincent, M. Varrot, and P. Mortini, “Analysis of high temperatures effects on performance and hot-carrier degradation in DC/AC stressed 0.35 μm n-MOSFETs,” *Microel. Reliab.*, vol. 39, no. 1, pp. 35–44, 1999.

[72] P. Moens, J. Mertens, F. Bauwens, P. Joris, W. D. Ceuninck, and M. Tack, “A comprehensive model for hot carrier degradation in LDMOS transistors,” in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 492–497, 2007.

[73] H. Enichlmair, S. Carniello, J. Park, and R. Minixhofer, “Analysis of hot carrier effects in a 0.35 μm high voltage n-channel LDMOS,” *Microel. Reliab.*, vol. 47, no. 9-11, pp. 1439–1443, 2007.

[74] K. Lee, C. Kang, O. S. Yoo, R. Choi, B. Lee, J. Lee, H.-D. Lee, and Y.-H. Jeong, “PBTI-associated high-temperature hot carrier degradation of nMOSFETs with metal-gate/high-k dielectrics,” *IEEE Electron Dev. Lett.*, vol. 29, pp. 389–391, Apr. 2008.

[75] M. Jo, S. Kim, C. Cho, M. Chang, and H. Hwang, “Gate voltage dependence on hot carrier degradation at an elevated temperature in a device with ultrathin silicon oxynitride,” *Appl. Phys. Lett.*, vol. 94, no. 5, pp. 053505–1–053505–3, 2009.

[76] E. Amat, T. Kauerauf, R. Degraeve, R. Rodriguez, M. Nafria, X. Aymerich, and G. Groeseneken, “Channel hot-carrier degradtion in pMOS and nMOS short channel transistors with high-k dielectric stack,” *Microel. Engineering*, vol. 87, no. 1, pp. 47–50, 2010.

[77] A. Ludikhuize, “Kirk effect limitations of HV ICs,” in *Proc. International Symposium on Power Semiconductor devices and ICs*, pp. 249–252, 1994.

[78] A. Ludikhuize, M. Slotboom, A. Nezar, N. Nowlin, and R. Brock, “Analysis of hot-carrier-induced degradation and snapback in submicron 50V lateral MOS transistors,” in *Proc. International Symposium on Power Semiconductor devices and ICs*, pp. 53–56, 1997.

[79] K.-M. Wu, *Development and hot-carrier reliability study of integrated high-voltage MOSFET transistor.* PhD thesis, National Cheng Kung University Tainan, 2007.

[80] H. Wu, W. Sun, Y. Yi, H. Li, and L. Shi, “The degradation mechanisms in high voltage pLEDMOS transistor with thick gate oxide,” *Microel. Reliab.*, vol. 48, no. 11-12, pp. 1804–1808, 2008.

[81] C. Gao, J. Wang, L. Wang, A. Yap, and H. Li, “Two-stage hot-carrier degradation behavior of 0.18 μm 18V n-type DEMOS and its recovery effect,” *Microel. Reliab.*, vol. 49, no. 1, pp. 8–12, 2009.

[82] B. Doyle, M. Bourcerie, J.-C. Marchetaux, and A. Boudou, “Relaxation effects in nMOS transistors after hot-carrier stressing,” *IEEE Electron Dev. Lett.*, vol. 8, no. 5, pp. 234–236, 1987.

[83] P. Cuevas, “A simple explanation for the apparent relaxation effect associated with hot-carrier phenomenon in MOSFETs,” *IEEE Electron Dev. Lett.*, vol. 9, no. 12, pp. 627–629, 1988.

[84] C. Young, J. Yang, K. Metthews, S. Suthram, M. Hussain, G. Bersuker, C. Smith, R. Harris, R. Choi, B. Lee, and H.-H. Tseng, “Hot carrier degradation in HfSiON/TiN fin shaped field effect transistor with different substrate orientation,” *J. Vac. Sci. Technol. B*, vol. 27, no. 1, pp. 468–471, 2009.

[85] S.-Y. Chen, C.-H. Tu, J.-C. Lin, M.-C. Wang, P.-W. Kao, M.-H. Ling, S.-H. Wu, Z.-W. Jhou, J. Ko, and H.-S. Haung, “Investigation of DC hot-carrier degradation at elevated temperatures fro p-channel metal-oxide-semiconductor field-effect transistors of 0.13 μm technology,” *Jpn. Journ. Appl. Phys.*, vol. 47, pp. 1527–1531, 2008.

[86] H. Enichlmair, J. Park, S. Carniello, B. Loeffer, R. Minixhofer, and M. Levy, “Hot carrier straess degradation modes in p-type high voltage LDMOS transistors,” in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 426–430, 2009.

[87] C. Salm, A. Hof, F. Kuper, and J. Schmitz, “Reduced temperature dependence of hot carrier degradation in deuterated nMOSFETs,” *Microel. Reliab.*, vol. 46, no. 9-11, pp. 1617–1622, 2006.

[88] Z. Chen, P. Ong, A. Mylin, V. Singh, and S. Cheltur, “Direct evidence of multiple vibrational excitation for the Si-H/D bond breaking in metal-oxide-semiconductor transistors,” *Appl. Phys. Lett.*, vol. 81, no. 17, pp. 3278–3280, 2002.

[89] E. Li, E. Rosenbaum, J. Tao, G.-F. Yeap, M. Lin, and P. Fang, “Hot-carrier effects in nMOSFETs in 0.1 μm CMOS technology,” in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 253–258, 1999.

[90] C. Lin, S. Biesemans, L. Han, K. Houlihan, T. Schiml, K. Schruefer, C. Wann, and R. Markhopf, “Hot carrier reliability for 0.13 μm CMOS technology with dual gate oxide thickness,” in *Proc. International Electron Devices Meeting (IEDM)*, pp. 135–138, 2000.

[91] R. Woltjer, A. Hamada, and E. Takeda, “pMOSFET hot carrier damage: oxide charge and interface states,” *Semicond Sci. Technol.*, vol. 7, pp. pp. B581–B584, 1992.

[92] A. Bravaix, D. Goguenheim, N. Revil, and E. Vincent, “Hole injection enhanced hot-carrier degradation in pMOSFETs used for systems on chip applications with 6.5-2nm thick gate oxides,” *Microel. Reliab.*, vol. 44, no. 1, pp. 65–77, 2004.

[93] T. Grasser, H. Kosina, and S. Selberherr, “Influence of the distribution function shape and the band structure on impact ionization modeling,” *Journ. Appl. Phys.*, vol. 90, no. 12, pp. 6165–6171, 2001.

[94] A. Gehring, T. Grasser, H. Kosina, and S. Selberherr, “Simulation of hot-electron oxide tunneling current based on a non-Maxwellian electron energy distribution function,” *Journ. Appl. Phys.*, vol. 92, no. 10, pp. 6019–6027, 2002.

[95] T. Grasser, H. Kosina, and S. Selberherr, “Hot carrier effects within macroscopic transport models,” *International Journal of High Speed Electronics and Systems*, vol. 13, no. 3, pp. 873–901, 2003.

[96] A. Zaka, Q. Rafhay, M. Iellina, P. Palestri, R. Clerc, D. Rideau, D. Garetto, J. Singer, G. Pananakakis, C. Tavernier, and H. Jaouen, “On the accuracy of current TCAD hot carrier injection models in nanoscale devices,” *Solid-State Electron.*, vol. 54, pp. 1669-1674, 2010.

[97] M. Vasicek, J. Cervenka, D. Esseni, P. Palestri, T. Grasser, “Applicability of macroscopic transport models to decananometer MOSFETs,” *IEEE Trans. Electron Dev.*, vol. 59, pp. 639 - 646, 2012.

[98] B. Persson and P. Avouris, “Local bond breaking via STM-induced excitations: the role of temperature,” *Surface Science*, vol. 390, pp. 45–54, 1997.

[99] K. Hess, L. Register, B. Tuttle, J. Lyding, and I. Kizilyalli, “Impact of nanostructure research on conventional solid-state electronics: the giant isotope effect in hydrogen desorption and CMOS lifetime,” *Physica E*, vol. 3, pp. 1–7, 1998.

[100] D. Wolters and A. Z.-V. Duyhoven, “Trapping of hot electrons,” *Appl. Surf. Sci*, vol. 39, pp. 565–577, 1989.

[101] R. Woltjer and G. Paulzen, “Modeling of oxide-charge generation during hot-carrier degradation of pMOSFETs,” *IEEE Trans. Electron Dev.*, vol. 41, no. 9, pp. 1639–1645, 1994.

[102] S. Mahapatra, D. Parikh, V. Rao, C. Viswanathan, and J. Vasi, “A comprehensive study of hot-carrier induced interface and oxide trap distribution in MOSFETs using a novel charge pumping technique,” *IEEE Trans. Electron Dev.*, vol. 47, no. 1, pp. 171–177, 2000.

[103] S. Samanta, N. Patel, K. ManjulaRani, and K. Jang, “Stress voltage dependence HCI induced traps distribution in 60V nLDMOS,” *IIRW Final Report*, pp. 120–123, 2009.

[104] C. Cheng, K. Tu, T. Wang, T. Hsieh, J. Tzeng, Y. Yong, R. Liou, and S. Hsu, “Investigation of hot carrier degradation modes in LDMOS by using a novel three-region charge pumping technique,” in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 334–337, 2006.

[105] D. Fleetwood, “Fast and slow border traps in MOS devices,” *IEEE Trans. Nuclear Sci.*, vol. 43, no. 3, pp. 779–786, 1996.

[106] D. Fleetwood, H. Xiong, Z.-Y. Lu, J. Felix, R. Schrimpf, and S. Pantelidis, “Unified model of hole trapping, 1/f noise, and thermally activated current in MOS devices,” *IEEE Trans. Nuclear Sci.*, vol. 49, no. 6, pp. 2674–2683, 2002.

[107] A. Bravaix, C. Trapes, N. Revil, and E. Vincent, “Carrier injection efficiency for the reliability study of 3.5-1.2 nm thick gate-oxide CMOS technologies,” *Microel. Reliab.*, vol. 43, no. 8, pp. 1241–1246, 2003.

[108] A. Schwerin, W. Hänsch, and W. Weber, "The relationsjip between the oxide charge and device degradation: a comparative study of n- and p-channel MOSFETs," *IEEE Trans. Electron Dev.*, vol. 34, no. 12, pp. 2493–2500, 1987.

[109] D. DiMaria and J. Stasiak, "Trap creation in silicon dioxide produced by hot electrons," *Journ. Appl. Phys.*, vol. 65, no. 6, pp. 2342–2356, 1989.

[110] D. DiMaria, "Defect generation under substrate-hot-electron injection into ultrathin silicon dioxide layers," *Journ. Appl. Phys.*, vol. 86, no. 4, pp. 2100–2109, 1999.

[111] H. Momose, S.-I. Nakamura, T. Ohguro, T. Yoshitomi, E. Morifuji, T. Morimoto, Y. Katsumata, and H. Iwai, "A study of hot-carrier degradation in n- and p-MOSFETs with ultra-thin gate oxides in the direct-tunneling regime," *IEDM Technical Digest*, pp. 453–456, 1997.

[112] I. Polishchuk, Y.-C. Yeo, Q. Lu, T.-J. King, and C. Hu, "Hot-carrier reliability of p-MOSFET with ultra-thin silicon nitride gate dielectric," in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 425–430, 2001.

[113] J. Sim, B. Lee, R. Choi, K. Matthews, D. Kwong, L. Larson, P. Tsui, and G. Bersuker, "Hot carrier reliability of HfSiON nMOSFETs with poly and TiN metal gate," *Device Research Conference*, vol. 1, pp. 99–100, 2004.

[114] J. Sim, B. Lee, R. Choi, S.-C. Song, and G. Bersuker, "Hot carrier degradation of HfSiON gate dielectrics with TiN electrode," *IEEE Trans. Dev. Material Realiab.*, vol. 5m, no. 2, pp. 117–182, 2005.

[115] N. Chowdhurz, P. Srinivasan, and D. Misra, "Trapping in deep defects under substrate hot electron stress in TiN/Hf-silicate based gate stacks," *Solid-State Electron.*, vol. 51, pp. 102–110, 2007.

[116] A. Sarwar, M. Siddiqui, R. Siddiqui, and Q. Khosru, "Effects of interface traps and oxide traps on gate capacitance of MOS devices with ultrathin (EOT $\sim$ 1 nm) high-k stacked gate devices," *TENCON*, pp. 1–5, 2009.

[117] S.-G. Hu, Y. Hao, X.-H. Ma, Y.-R. Cao, C. Chen, and X.-F. Wu, "Hot-carrier stress effects on GIDL and SILC in 90 nm LDD-MOSFET with ulgtra-thin gate oxide," *Chin. Phys. Lett.*, vol. 26, no. 1, 2009.

[118] E. Nicollian, C. Berglund, P. Schmidt, and J. Andrews, "Electrochemical charging of thermal SiO_2 films by injected electron currents," *Journ. Appl. Phys.*, vol. 42, no. 12, pp. 5654–5664, 1971.

[119] T. Ning, "Thermal reemission of trapped electrons in SiO_2," *Journ. Appl. Phys.*, vol. 49, no. 12, pp. 5997–6003, 1978.

[120] D. Young, E. Irene, D. DiMaria, R. D. Keersmaecker, and H. Massoud, "Electron trapping in SiO_2 at 295 and 77 K," *Journ. Appl. Phys.*, vol. 50, no. 10, pp. 6366–6372, 1979.

[121] D. Wolters and J. van der Schoot, "Kinetics of charge trapping in dielectrics," *Journ. Appl. Phys.*, vol. 58, no. 2, pp. 831–837, 1985.

[122] J. Maserjian and N. Zamani, "Observation of positively charged state generation near the Si/SiO$_2$ interface during Fowler-Nordheim tunneling," *Journ. Vacuum Sci. Technol.*, vol. 20, no. 3, pp. 743–746, 1982.

[123] R. Rofam and C. Hu, "Stress-induced oxide leakage," *IEEE Electron Dev. Lett.*, vol. 12, no. 11, pp. 632–634, 1991.

[124] T. Wang, N.-K. Zous, J.-L. Lai, and C. Huang, "Hot hole stress induced leakage current (SILC) transient in tunnel oxides," *IEEE Electron Dev. Lett.*, vol. 19, no. 11, pp. 411–413, 1998.

[125] D. Dumin and J. Maddux, "Correlation of stress-induced leakage current in thin oxides with trap creation inside the oxides," *IEEE Trans. Electron Dev.*, vol. 90, no. 5, pp. 986–992, 1993.

[126] S. Mahapatra, D. Saha, and P. Kumar, "On the generation and recovery of interface traps in MOSFETs subjected to NBTI, FN, and HCI stress," *IEEE Trans. Electron Dev.*, vol. 53, no. 7, pp. 1583–1592, 2006.

[127] S. Tsujikawa, "SILC and NBTI in pMOSFETs with ultrathin SiON gate dielecrtics," *IEEE Trans. Electron Dev.*, vol. 54, no. 3, pp. 524–530, 2007.

[128] D. Schroder, "Negative bias temperature instability: what do we understand?," *Microel. Reliab.*, vol. 47, no. 6, pp. 841–852, 2007.

[129] Y.-R. Cao, Y. Hao, X.-H. Ma, L. Yu, and S.-G. Hu, "SILC during NBTI stress in pMOSFETs with ultra-thin SiON gate dielectric," *Chin. Phys. Lett.*, vol. 25, no. 4, pp. 1427–1430, 2008.

[130] A. Schenk and U. Krumbein, "Coupled defect-level recombination: Theory and application to anomalous diode characteristics," *Journ. Appl. Phys.*, vol. 78, no. 5, pp. 3185–3191, 1995.

[131] P. Bloechl, "First-principle calculations of defects in oxygen-deficient silica exposed to hydrogen," *Phys. Rev. B*, vol. 62, no. 10, pp. 6159–6179, 2000.

[132] M. Houssa, A. Stesmans, and M. Heyns, "Model for the trap-assistant tunneling current through very thin SiO$_2$/ZrO$_2$ gate dielectric stacks," *Semicond. Sci. Technol.*, vol. 16, pp. 427–432, 2001.

[133] L. Larcher, "Statistical simulation of leakage currents in MOS and flash memory devices with a new multiphonon trap-assisted tunneling model," *IEEE Trans. Electon Dev.*, vol. 50, no. 2, pp. 1246–1253, 2003.

[134] O. Blank, H. Reisinger, R. Stengl, M. Gutsche, F. Weist, V. Capodieci, J. Schulze, and I. Eisele, "A model for multistep trap-assisted tunneling in thin high-k dielectrics," *Journ. Appl. Phys.*, vol. 97, no. 4, 2005.

[135] G. Bersuker, D. Heh, C. Young, H. Park, P. Khanal, L. Larcher, A. Padovani, P. Lenahan, J. Ryan, B. Lee, H. Tseng, and R. Jammy, “Breakdown in the metal/high-k gate stack: identifying the “weak link” in the multilayer dielectric,” *IEDM Technical Digest*, pp. 791–794, 2008.

[136] T. Grasser, H. Reisinger, P.-J. Wagner, D. Kaczer, F. Schanowsky, and W. Gös, “The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability,” in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 16–25, 2010.

[137] T. Aichinger, M. Nelhiebel, and T. Grasser, “On the temperature dependence of NBTI recovery,” *Microel. Reliab.*, vol. 48, pp. 1178–1184, 2008.

[138] T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P.-J. Wagner, F. Schanowsky, J. Franco, P. Roussel, and M. Nelhiebel, “Recent advances in understanding the bias temperature instability,” in *Proc. International Electron Devices Meeting (IEDM)*, pp. 82–85, 2010.

[139] R. Walkup, D. Newns, and P. Avouris, “Role of mutliple inelastic transistions in atom transfet with the scanning tunneling microscope,” *Phys. Rev. B*, vol. 48, no. 3, pp. 1858–1861, 1993.

[140] J. Lyding, K. Hess, G. Abeln, D. Thompson, J. Moore, M. Hersam, E. Foley, J. Lee, S. Hwang, H. Choi, P. Avouris, and I. Kizialli, “Ultrahigh vacuum-scanning tunneling microscopy nanofabrication and hydrogen/deuterium desorption from silicon surfaces: implications for complementary metal oxide semiconductor technology,” *Appl. Surf. Sci.*, vol. 13-132, pp. 221–230, 1998.

[141] K. Stokbro, C. Thirstrup, M. Sakurai, U. Quaade, B. Y.-K. Hu, F. Perez-Murano, and F. Grey, “STM-induced hydrogen desorption via a hole resonance,” *Phys. Rev. Lett.*, vol. 80, pp. 2618–2621, 1998.

[142] M. Budde, G. Lüpke, E. Chen, X. Zhang, N. H. Tolk, L. C. Feldman, E. Tarhan, A. K. Ramdas, and M. Stavola, “Lifetimes of hydrogen and deuterium related vibrational modes in silicon,” *Phys. Rev. Lett.*, vol. 87, no. 4, pp. 1455–1461, 2001.

[143] J. Sune and Y. Wu, “Hydrogen-release mechanisms in the breakdown of thin SiO_2 films,” *Phys. Rev. Lett.*, vol. 92, no. 8, pp. 087601 (1–4), 2004.

[144] J. Sune and Y. Wu, “Mechanisms of hydrogen release in the breakdown of SiO_2-based oxides,” in *Proc. International Electron Devices Meeting (IEDM)*, pp. 388–391, 2005.

[145] R. Biswas, Y.-P. Li, and B. C. Pan, “Enhanced stability of deuterium in silicon,” *Appl. Phys. Lett.*, vol. 72, no. 26, pp. 3500–3503, 1998.

[146] G. Ribes, S. Bruyere, M. Denais, F. Monsieur, V. Huard, D. Roy, and G. Ghibaudo, “Multi-vibrational hydrogen release: physical origin of T_{bd}, Q_{bd} power-law voltage dependence of oxide breakdown in ultra-thin gate oxides,” *Microel. Reliab.*, vol. 45, pp. 1842–1854, 2005.

[147] B. Tuttle and C. V. de Walle, “Structure, energetics, and vibrational properties of Si-H bond dissociation in silicon,” *Phys. Rev. B*, vol. 59, no. 20, pp. 12884–12889, 1999.

[148] K. Hess, A. Haggag, W. McMahon, B. Fischer, K. Cheng, J. Lee, and L. Lyding, "Simulation of $\mathrm{Si-SiO_2}$ defect generation in CMOS chips: From atomistic structure to chip failure rates," in *Proc. International Electron Devices Meeting (IEDM)*, pp. 93–96, 2000.

[149] H. Kufluoglu and M. Alam, "A geometrical unification of the theories of NBTI and HCI time exponents and its implications for ultra-scaled planar and surround-gate MOSFETs," in *Proc. International Electron Devices Meeting (IEDM)*, pp. 113–116, 2004.

[150] T.Grasser, W. Gös, and B. Kaczer, "Dispersive transport and negative bias temperature instability: boundary conditions, initial conditions, and transport models," *IEEE Trans. Dev. Material. Reliab.*, vol. 8, no. 1, pp. 79–97, 2008.

[151] T. Grasser, H. Reisinger, W. Goes, T. Aichinger, P. Hehenberger, P.-J. Wagner, M. Nelhiebel, J. Franco, and B. Kaczer, "Switching oxide traps as the missing link between negative bias temperature instability and random telegraph noise," in *Proc. International Electron Devices Meeting (IEDM)*, 2009.

[152] S. Rauch and G. L. R. and, "The energy-driven paradigm of nMOSFET hot-carrier effects," *IEEE Trans. Dev. Material. Reliab.*, vol. 5, no. 4, pp. 701–705, 2005.

[153] C. Gardner, "The classical and quantum hydrodynamic models.," *in Proc. Intl. Workshop on Computational Electronics, (University of Leeds)*, pp. 25–36, 1993.

[154] C. Jungemann and B. Meinerzhagen, *Hierarchical Device Simulation.* Springer Verlag Wien/New York, 2003.

[155] S. Sze and K. Ng, *Physics of semiconductor devices.* Wiley-Interscience publication, Wiley-Interscience, 2007.

[156] M.-T. Vasicek, *Advanced macroscopic transport models.* PhD thesis, Vienna Technical University, Vienna, Austria, 2009.

[157] *Sentaurus device Monte Carlo user guide*, SYNOPSYS, inc., 2009.

[158] S.-M. Hong and C. Jungemann, "A fully coupled scheme for a Boltzmann-Poisson equation solver based on a spherical harmonics expansion," *Journ. Comput. Electron.*, vol. 8, no. 3-4, pp. 225–241, 2009.

[159] K. Rupp, *Numerical solution of the Boltzmann transport equation*, PhD thesis, Vienna Technical University, Vienna, Austria, 2012.

[160] K. Rupp, C. Jungemann, M. Bina, A. Jngel, T. Grasser, "Bipolar spherical harmonics expansions of the Boltzmann transport equation," in *Proc. International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*,pp. 19–22 2012.

[161] A. Gnudi, D. Ventura, G. Baccarani, and F. Oden, "Two-dimensional MOSFET simulations by means of of a multidimensional spherical harmonics expansion of the Boltzmann transport equation," *Solid-State Electron.*, vol. 36, no. 4, pp. 575–581, 1993.

[162] S.-M. Hong, C. Jungemann, and M. Bollhofer, "A deterministic Bolzmann equation solver for two-dimensional semiconductor devices," in *Proc. International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, pp. 293–296, 2008.

[163] Institute for Microelectronic, TU Wien, *MiniMOS-NT Device and Circuit Simulator.*

[164] W. Emden, W. Krautschneider, G. Tempel, R. Hagenbeck, and F. Beug, “A modified constant field charge pumping method for sensitive profiling of near-junction charges,” in *Proc. European Solid-State Device Research Conference (ESSDERC)*, pp. 279–282, 2007.

[165] H.-S. Wong, M. White, J. Krutsick, and R. Booth, “Modeling of transconductance degradation and extraction of threshold voltage in thin oxide MOSFET’s,” *Solid-State Electron.*, vol. 30, no. 9, pp. 953–958, 1987.

[166] A. G. Prakash, S. Ke, and K. Siddappa, “High-energy radiation effects on subthreshold characteristics, transconductance and mobility of n-channel MOSFETs,” *Semicond. Sci. Technol.*, vol. 18, no. 12, pp. 1037–1042, 2003.

[167] J. Wang-Ratkovic, R. Lacoe, K. Williams, M. Song, S. Brown, and G. Yabiku, “New understanding of LDD CMOS hot-carrier degradation and device lifetime at cryogenic temperatures,” in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 312–314, 2003.

[168] *DESSIS ISE TCAD Manual Release 9.0*, 2002.

[169] T. Grasser, C. Jungemann, H. Kosina, B. Meinerzhagen, and S. Selberherr, “Advanced transport models for sub-micrometer devices,” in *Proc. Simulation of Semiconductor Processes and Devices (SISPAD)*, pp. 1–8, 2004.

[170] A. Gehring, T. Grasser, H. Kosina, and S. Selberherr, “Simulation of hot-electron oxide tunneling current based on a non-Maxwellian electron energy distribution function,” *Journ. Appl. Phys.*, vol. 92, no. 10, pp. 6019–6027, 2002.

[171] C. Fiegna, F. Venturi, M. Melanotte, E. Sangiorgi, and B. Ricco, “Simple and efficient modeling of EPROM writing,” *IEEE Trans. Electron Dev.*, vol. 38, no. 3, pp. 603–610, 1991.

[172] G. Groeseneken, H. Maes, N. Beltran, and R.F. De Keersmaecker, “Reliable approach to charge-pumping measurements in MOS transistors,” *IEEE Trans. Electron Dev.*, vol. 31, pp. 42-53, 1984.

[173] P. Heremans, J. Witters, G. Groeseneken, and H. Maes, “Analysis of the charge pumping technique and its applicationfor the evaluation of the MOSFET degradation,” *IEEE Trans. Electron Dev.*, vol. 36, p. 1318, 1989.

[174] S. Okhonin, T. Hessler, and M. Dutoit, “Comparison of gate-induced drain leakage and charge pumping measurements for determining lateral interface trap profiles in electrically stressed MOSFETs,” *IEEE Trans. Electron Dev.*, vol. 43, no. 4, pp. 605–612, 1996.

[175] M. Tsuchiaki, H. Hara, T. Morimoto, and H. Iwai, “A new charge pumping method for determining the spatial distribution of hot-carrier-induced fixed charge in p-MOSFET’s,” *IEEE Trans Electron Dev.*, vol. 40, no. 10, pp. 1768–1799, 1993.

[176] S. Chung and J.-J. Yang, “A new approach for characterizing structure-dependent hot-carrier effects in drain-engineered MOSFET’s,” *IEEE Trans. Electron Dev.*, vol. 46, pp. 1371–1377, July 1999.

[177] W. K. Chim, S. E. Leang, and D. S. H. Chan, “Extraction of metal-oxide semiconductor field-effect-transistor interface state and trapped charge spatial distributions using a physics-based algorithm,” *Journ. Appl. Phys.*, vol. 81, no. 4, pp. 1992–1997, 1997.

[178] R. G.-H. Lee, J.-S. Su, and S. S. Chung, “A new method for characterizing the spatial distributions of interface states and oxide-trapped charges in LDD n-MOSFETs,” *IEEE Trans. Electron Dev.*, vol. 43, no. 1, pp. 81–89, 1996.

[179] H.-H. Li, Y.-L. Chu, and C.-Y. Wu, “A novel charge-pumping method for extracting the lateral distributions of interface-trap and effective oxide-trapped charge densities in MOSFET devices,” *IEEE Trans. Electron Dev.*, vol. 44, pp. 782–791, may 1997.

[180] A. Elliot and and, “The use of charge pumping currents to measure surface state densities in MOS transistors,” *Solid-State Electron.*, vol. 19, pp. 241–247, 1976.

[181] B. Djezzar, S. Oussalah, and A. Smatti, “A new oxide-trap based on charge-pumping (OTCP) extraction method for irradiated MOSFET devices: part I (high frequencies),” *IEEE Trans. Nuclear Science*, vol. 51, pp. 1724–1731, aug. 2004.

[182] H.-H. Li, Y.-L. Chu, and C.-Y. Wu, “A new simplified charge pumping current model and its model parameter extraction,” *IEEE Trans. Electron Dev.*, vol. 43, pp. 1857–1863, nov 1996.

[183] W. Chen, A. Q. Balasinski, and T. P. Ma, “Lateral profiling of oxide charge and interface traps near MOSFET junctions,” *IEEE Trans. Electron Devices*, vol. 40, no. 1, pp. 187–196, 1993.

[184] N. Pesonen, W. Kahn, R. Allen, M. Cresswell, and M. Zaghloul, “Application of conformal mapping approximation techniques: parallel conductors of finite dimensions,” *IEEE Trans. Instrum. Meas.*, vol. 53, pp. 812–821, june 2004.

[185] A. N. Tikhonov and A. A. Samarskii, *Equations of Mathematical Physics.* Dover Publications, 2011.

[186] V. Huard, M. Denais, and C. Parthasarathy, “NBTI degradation: from physical mechanisms to modeling,” *Microel. Reliab.*, vol. 46, no. 1, pp. 1–23, 2006.

[187] T. Grasser, T. Achinger, G. Pobegen, H. Reisinger, P.-J. Wagner, J. Franco, M. Nelhiebel, and B. Kaczer, “The ‘permanent’ component of NBTI: composition and annealing,” in *Conference Proceedings of International Reliability Physics Symposium (IRPS 2011)*, 2011.

[188] M. Houssa, M. Tuominen, M. Naili, V. Afanas’ev, A. Stesmans, S. Haukka, and M. M. Heyns, “Trap-assisted tunneling in high permittivity gate dielectric stacks,” *Journ. Appl. Phys.*, vol. 87, no. 12, pp. 8615–8620, 2000.

[189] T. Grasser, “Stochastic charge trapping in oxides: from random telegraph noise to bias temperature instabilities,” *Microel. Reliab.*, vol. 52, no. 1, pp. 39–70, 2011.

[190] C. Chen and T. Ma, "Direct lateral profiling of both interface traps and oxide charge in thin gate MOSFET devices," *IEEE Trans. Electron Dev.*, vol. 45, no. 2, pp. 512–520, 1998.

[191] Y.-L. Chu, D.-W. Lin, and C.-Y. Wu, "A new charge-pumping technique for profiling the interface-states and oxide-trapped charges in MOSFETs," *IEEE Trans. Electron Dev.*, vol. 47, no. 2, pp. 348–353, 2000.

[192] H. Reisinger, T. Grasser, C. Schlunder, and W. Gustin, "The statistical analysis of individual defects constituting NBTI and its implications for modeling DC- and AC-stress," in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 7–15, 2010.

[193] B. Kaczer, T. Grasser, P. Roussel, J. Franco, R. Degraeve, L. Ragnarsson, E. Simoen, H. Reisinger, *et al.*, "Origin of NBTI variability in deeply scaled pFETs," in *Proc. International Reliability Physics Symposium (IRPS)*, pp. 26–32, 2010.

[194] V. Huard, M. Denais, F. Perrier, N. Revil, C. Parthasarathy, A. Bravaix, and E. Vincent, "A thorough investigation of MOSFETs NBTI degradation," *Microel. Reliab.*, vol. 45, pp. 83–98, 1 2005.

[195] B. Bindu, W. Goes, B. Kaczer, and T. Grasser, "Analytical solution of the switching trap model for negative bias temperature stress," in *Proc. International Integrated Reliability Workshop (IIRW)*, pp. 93–96, 2009.

[196] N. Stojadinovic, M. Pejovic, S. Golubovic, G. Ristic, V. Davidovic, and S. Dimitriev, "Effect of radiation-induced oxide-trapped charge on mobility in p-channel MOSFETs," *Electronics Letters*, vol. 31, no. 6, pp. 497–498, 1995.

[197] M. Abramowitz and I. Stegun, *Handbook on mathematical functions.* New York: Dover, 1972.

Own Publications

[1] I. Starkov, H. Enichlmair, S. Tyaginov, and T. Grasser: "Charge-Pumping Extraction Techniques for Hot-Carrier Induced Interface and Oxide Trap Spatial Distributions in MOSFETs"; in *Proceedings of the* 19^{th} *International Symposium on the Physical and Failure Analysis of Integrated Circuits*, 2012, 6 pages; Talk at the IPFA, Singapore; 2010-07-02 – 2010-07-06.

[2] S.E. Tyaginov, I.A. Starkov, O. Triebl, M. Karner, Ch. Kernstock, C. Jungemann, H. Enichlmair, J.M. Park, and T. Grasser: "Impact of Gate Oxide Thickness Variations on Hot-Carrier Degradation"; in *Proceedings of the* 19^{th} *International Symposium on the Physical and Failure Analysis of Integrated Circuits*, 2012, 5 pages; Poster at the IPFA, Singapore; 2010-07-02 - 2010-07-06.

[3] I. Starkov, H. Enichlmair, T. Grasser: "Local Oxide Capacitance as a Crucial Parameter for Characterization of Hot-Carrier Degradation in High-Voltage n-MOSFET"; in *Abstract Booklet of the* 17^{th} *Workshop on Dielectrics in Microelectronics*, 2012, p. 40. Talk at the WODIM, Dresden, Germany; 2012-06-25 – 2012-06-27.

[4] I. Starkov, H. Enichlmair, S. Tyaginov, T. Grasser: "Analysis of the Threshold Voltage Turn-Around Effect in High-Voltage n-MOSFETs Due to Hot-Carrier Stress"; in *Conference Proceedings of International Reliability Physics Symposium*, 2012, 6 pages; Poster at the IRPS, California, USA; 2012-04-17 – 2012-04-19.

[5] I. Starkov, A. Starkov, S. Tyaginov, H. Enichlmair, H. Ceric, and T. Grasser: "An Analytical Model for MOSFET Local Oxide Capacitance"; in *Proceedings of the International Semiconductor Device Research Symposium*, 2011, 2 pages; Poster at the ISDRS, Washington DC, USA; 2011-12-07 – 2011-12-09.

[6] S. Tyaginov, I. Starkov, H. Enichlmair, C. Jungemann, J.M. Park, E. Seebacher, R. Orio, H. Ceric, T. Grasser: "An Analytical Approach for Physical Modeling of Hot-Carrier Induced Degradation"; in *Proceedings of the* 22^{nd} *European Symposium on the Reliability of Electron Devices, Failure Physics and Analysis*, 2011, pp. 1525-1529; Talk at the ESREF, Bordeaux, France; 2011-10-03 – 2011-10-07.

[7] I. Starkov, S. Tyaginov, H. Enichlmair, J. M. Park, H. Ceric, and T. Grasser: "Accurate Extraction of MOSFET Unstressed Interface State Spatial Distribution from Charge Pumping Measurements"; *Solid State Phenomena*, vol. 178-179, pp. 267-272, 2011.

[8] S. Tyaginov, I. Starkov, H. Enichlmair, C. Jungemann, J.M. Park, E. Seebacher, R. Orio, H. Ceric, T. Grasser: "An Analytical Approach for Physical Modeling of Hot-Carrier Induced Degradation"; *Microelectronics Reliability*, vol. 51, pp. 1525-1529, 2011.

[9] S. Tyaginov, I. Starkov, C. Jungemann, H. Enichlmair, J.M. Park, T. Grasser: “Impact of the Carrier Distribution Function on Hot-Carrier Degradation Modeling”; in *Proceedings of the* 41st *European Solid-State Device Research Conference*, 2011, pp. 151-154; Talk at the ESSDERC, Helsinki, Finland; 2011-09-12 – 2011-09-16.

[10] S. Tyaginov, I. Starkov, H. Enichlmair, J.M. Park, C. Jungemann, T. Grasser: “Physics-Based Hot-Carrier Degradation Modeling”; in *Meet. Abstr. - Electrochem. Soc. 2011*, 2011, 1 page; Talk at the ECS Meeting, Montreal, Canada; 2011-05-01 – 2011-05-06.

[11] I. Starkov, S. Tyaginov, H. Ceric, T. Grasser: “Analysis of Worst-Case Hot-Carrier Conditions for n-type MOSFET”; in *Proceedings of the* 7th *Conference on PhD Research in Microelectronics and Electronics*, 2011, pp. 197-200; Talk at the PRIME, Madonna di Campiglio, Italy; 2011-03-07 – 2011-07-07.

[12] S. Tyaginov, I. Starkov, H. Enichlmair, J.M. Park, C. Jungemann, T. Grasser: “Physics-Based Hot-Carrier Degradation Models”; in *Silicon Nitride, Silicon Dioxide, and Emerging Dielectrics*, ECS Transactions, 2011, pp. 321-352.

[13] I.Starkov and H. Ceric: “Impact of Interface State Density on MOSFET Local Oxide Capacitance Degradation During Hot-Carrier Stress”; in *Extended Abstracts of 2011 International Conference on Solid State Devices and Materials*, 2011, pp. 90-91; Poster at the SSDM, Nagoya, Japan; 2011-28-10 – 2011-30-10.

[14] I. Starkov, S. Tyaginov, H. Enichlmair, J. M. Park, H. Ceric, and T. Grasser: “Accurate Extraction of MOSFET Interface State Spatial Distribution from Charge Pumping Measurements”; in *GADEST 2011: Abstract Booklet*, 2011, pp. 105-106; Talk at the GADEST, Loipersdorf, Austria; 2011-25-09 – 2011-30-09.

[15] I. Starkov, H. Ceric, S. Tyaginov, T. Grasser, H. Enichlmair, J.M. Park, C. Jungemann: “Analysis of Worst-Case Hot-Carrier Degradation Conditions in the Case of n- and p-channel High-Voltage MOSFETs”; in *Proceedings of the* 15th *International Conference on Simulation of Semiconductor Processes and Devices*, 2011, 4 pages; Talk at the SISPAD, Osaka, Japan; 2011-09-08 – 2011-09-10.

[16] S. Tyaginov, I. Starkov, O. Triebl, H. Ceric, T. Grasser, H. Enichlmair, J.M. Park, C. Jungemann: ”Secondary Generated Holes as a Crucial Component for Modeling of HC Degradation in High-Voltage n-MOSFET”; in *Proceedings of the* 15th *International Conference on Simulation of Semiconductor Processes and Devices*, 2011, 4 pages; Talk at the SISPAD, Osaka, Japan; 2011-09-08 – 2011-09-10.

[17] I. Starkov, S. Tyaginov, H. Enichlmair, J. Cervenka, C. Jungemann, S. Carniello, J. M. Park, H. Ceric, and T. Grasser: “Hot-carrier degradation caused interface state profile - Simulation versus experiment”; *J. Vac. Sci. Technol. B*, vol. 29, pp. 01AB09-1–01AB09-8, 2011.

[18] S.E. Tyaginov, I.A. Starkov, O. Triebl, J. Cervenka, C. Jungemann, S. Carniello, J.M. Park, H. Enichlmair, M. Karner, Ch. Kernstock , E. Seebacher, R. Minixhofer, H. Ceric, T. Grasser: “Hot-Carrier Degradation Modeling Using Full-Band Monte-Carlo Simulations”; in: *Proceedings of the* 17th *IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits*, 2010, pp. 341-345; Talk at the IPFA, Singapore; 2010-07-05 – 2010-07-09.

[19] I.A. Starkov, S.E. Tyaginov, O. Triebl, J. Cervenka, C. Jungemann, S. Carniello, J.M. Park, H. Enichlmair, M. Karner, Ch. Kernstock, E. Seebacher, R. Minixhofer, H. Ceric, T. Grasser: "Analysis of Worst-Case Hot-Carrier Conditions for High Voltage Transistors Based on Full-Band Monte-Carlo Simulations"; in *Proceedings of the* 17*th* *IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits*, 2010, pp. 139-144; Poster at the IPFA, Singapore; 2010-07-05 – 2010-07-09.

[20] I.Starkov, S. Tyaginov, H. Enichlmair, O. Triebl, J. Cervenka, C. Jungemann, S. Carniello, J.M. Park, H. Ceric, T. Grasser: "HC degradation model: interface state profile – simulations vs. experiment"; in *Book of Abstracts WoDiM 2010*, 2010, p. 128; Poster at the WoDiM, Bratislava, Slovakia; 2010-06-28 – 2010-06-30.

[21] S.E. Tyaginov, I.A. Starkov, O. Triebl, J. Cervenka, C. Jungemann, S. Carniello, J.M. Park, H. Enichlmair, M. Karner, Ch. Kernstock, E.Seebacher, R. Minixhofer, H. Ceric, T. Grasser: "Interface states charges as a vital component for HC degradation modeling"; in *Proceedings of the* 21*st* *European Symposium on the Reliability of Electron Devices, Failure Physics and Analysis*, 2010, pp. 1267-1272; Talk at the ESREF, Gaeta, Italy; 2010-10-11 – 2010-10-15.

[22] S.E. Tyaginov, I.A. Starkov, O. Triebl, J. Cervenka, C. Jungemann, S. Carniello, J.M. Park, H. Enichlmair, M. Karner, Ch. Kernstock, E. Seebacher, R. Minixhofer, H. Ceric, T. Grasser: "Interface Traps Density-Of-States as a Vital Component for Hot-Carrier Degradation Modeling"; *Microelectronics Reliability*, vol. 50, pp. 1267-1272, 2010.

[23] I. Starkov, S. Tyaginov, T. Grasser: "Greens Function Asymptotic in Two-Layered Periodic Medium"; in *Proceedings of the International Symposium NANOSTRUCTURES: Physics and Technology*, 2009, pp. 111 - 112; Talk at the ISNPT, Minsk; 2009-06-22 – 2009-06-26.

[24] S. Tyaginov, V. Sverdlov, I. Starkov, W. Goes, T. Grasser: "Impact of O-Si-O Bond Angle Fluctuations on the Si-O Bond-Breakage Rate"; *Microelectronics Reliability*, vol. 49, pp. 998 - 1002, 2009.

[25] S. Tyaginov, V. Sverdlov, I. Starkov, W. Goes, T. Grasser: "Impact of O-Si-O Bond Angle Fluctuations on the Si-O Bond-Breakage Rate"; in *Proceedings of the* 20*th* *European Symposium on the Reliability of Electron Devices, Failure Physics and Analysis*, 2009, pp. 998-1002; Talk at the ESREF, Bordeaux, France; 2009-10-05 – 2009-10-09.